Measuring the Power of Learning.™

Conceptualizing Accessibility for English Language Proficiency Assessments

ETS RR-16-07

Danielle Guzman-Orth • Cara Laitusis • Martha Thurlow • Laurene Christensen

NCEO National Center on Educational Outcomes

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Conceptualizing Accessibility for English Language Proficiency Assessments

Danielle Guzman-Orth,[1] Cara Laitusis,[1] Martha Thurlow,[2] & Laurene Christensen[2]

1 Educational Testing Service, Princeton, NJ
2 National Center on Educational Outcomes, University of Minnesota, Minneapolis, MN

This paper is the second in a series from Educational Testing Service (ETS) that conceptualizes next-generation English language proficiency (ELP) assessment systems for K-12 English learners (ELs) in the United States. The first paper articulated a high-level conceptualization of next-generation ELP assessment systems (Hauck, Wolf, & Mislevy, 2016), the third paper addressed issues related to summative ELP assessments that emerged from the presentations and discussions at the English Language Proficiency Assessment Research working meeting (Wolf, Guzman-Orth, & Hauck, 2016), and the fourth paper focused on a key concern within such systems—the initial identification and classification of ELs (Lopez, Pooler, & Linquanti, 2016). The goal of this paper is to address accessibility issues in the context of ELP assessments and to discuss critical considerations to improve the accessibility of ELP assessments for ELs and ELs with disabilities. Although accessibility for ELs and ELs with disabilities who are taking content assessments is also important, a discussion about content assessments is beyond the scope of the paper at this time. In this paper, we discuss challenges and areas of possible directions to pursue for ongoing and future ELP assessment development, policy implications, and research considerations to improve the ELP testing experience for all users.

Keywords English language proficiency; English learners; English learners with disabilities; accessibility

doi:10.1002/ets2.12093

Assessing English language proficiency (ELP) for accountability purposes is a complicated task given the existing heterogeneity within the target test-taking population. Some of this heterogeneity is due to students with varying levels of English and home language proficiency. Differences in students' educational experiences, including language (e.g., heritage language speakers) and formal educational opportunities (i.e., students with interrupted formal education experiences, refugees, and migrant students) add to the heterogeneity. Contributing to this complexity is the practical need to assess students enrolled in kindergarten through 12th grade who may be English learners (ELs; at the screening/identification stage), students who are ELs, and students who are ELs with disabilities.

Overall, the EL population continues to increase in number and diversity in the United States. Nearly every state in the United States has felt the impact of the increasing numbers of ELs. In 2010, nearly 10% of students attending elementary and secondary schools in the United States were ELs (U.S. Department of Education, 2014a). Still, the range in percentages across states varies considerably, from 1% to 22% of the state student population (see Figure 1; data from U.S. Department of Education, 2014b). Data collected in accordance with the Individuals With Disabilities Education Act ([IDEA], 2004) in the same year indicated that about 9% of all students with individualized education programs (IEPs) were ELs with disabilities served through IEPs. State rates ranged from 0% to 31%. All of these students should be participating in states' and consortia's ELP assessments.

The current educational reform to adopt more rigorous college and career readiness standards such as the Common Core State Standards and Next-Generation Science Standards (which are designed to bring comparability of instructional standards and expectations across states) also bring certain challenges because schools serve a heterogeneous student population. This reform, coupled with the current practice of inclusive classrooms and inclusive assessment policies (e.g., Title III of the Elementary and Secondary Education Act, Individuals With Disabilities Education Act of 2004), challenges states and multistate consortia to devise policies and procedures to ensure equity and opportunity for ELs, including ELs with disabilities, so they can access the content being assessed to show what they know and are able to do.

With support from the U.S. Department of Education Race-to-the-Top Program, consortia of states formed to develop innovative technology-based assessments of the common standards—the Partnership for Assessment of Readiness for

Corresponding author: D. Guzman-Orth, E-mail: dguzman-orth@ets.org

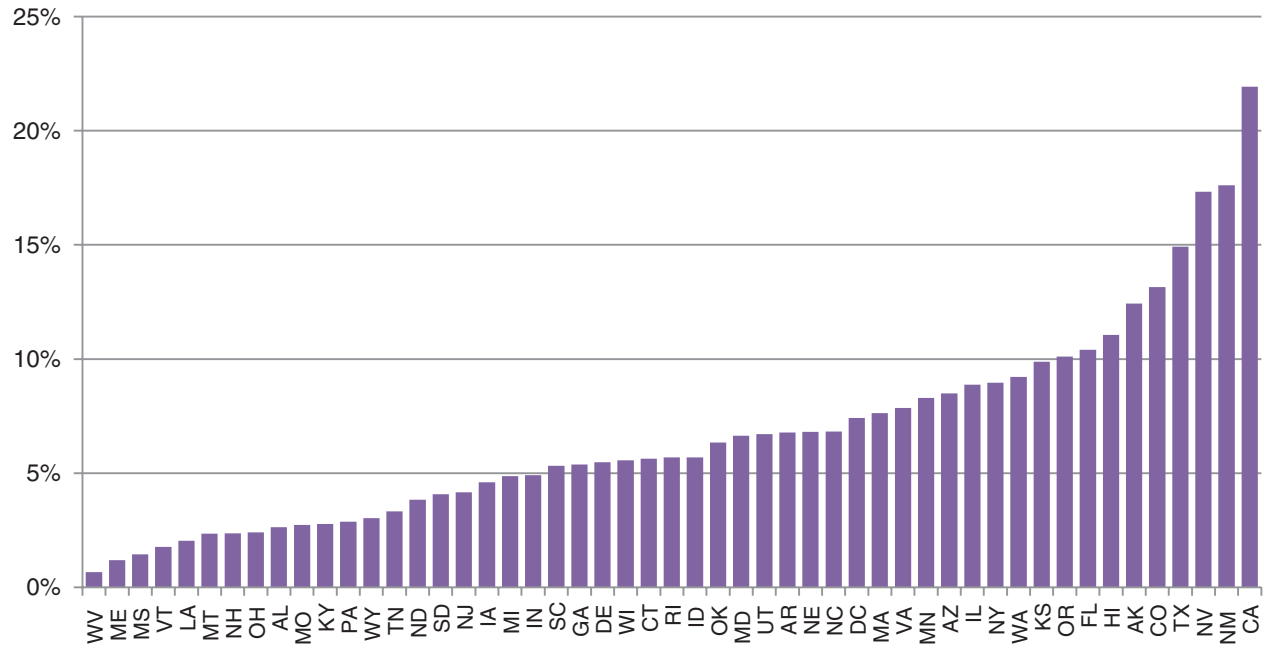

Figure 1 Percentage of English language students in states (all limited English proficient students divided by total enrollment) in 2011–2012. Data source: U.S. Department of Education (2014b).

College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (Smarter Balanced). At the same time, two other consortia of states received funding from the U.S. Department of Education Office of Special Education Programs to develop technology-based alternate assessments based on alternate achievement standards (AA-AAS) for students with significant cognitive disabilities—the Dynamic Learning Maps (DLM) and the National Center and State Collaborative (NCSC). All of these consortia were charged with developing assessments that are accessible for the wide range of students who are expected to take them.

Similarly, with support from funds through the U.S. Department of Education's Enhanced Assessment Grants, two consortia for developing ELP assessments, Assessment Services Supporting ELs through Technology Systems (ASSETS) and English Language Proficiency for the 21st Century (ELPA21), are designing measures used for high-stakes accountability and annual monitoring of ELs' skills in listening, speaking, reading, and writing in English. These next-generation assessments, like the others, must address the accessibility needs of the diverse population of students who will participate in them. At the time of this writing, mandates issued from the U.S. Department of Education made explicit the federal requirement to include ELs with disabilities in the general ELP assessment or an alternate ELP assessment, without exceptions (U.S. Department of Education, 2014a).

These federal mandates, coupled with the high-stakes uses of next-generation ELP assessments (including EL identification and classification, instructional placement, and reclassification or exiting from EL designation), make it a critical priority to explore accessibility practices in relation to the assessment of ELs and ELs with disabilities.

Attention to accessibility at the early stages in test development and throughout the development process can help minimize the effects of construct-irrelevant variance (e.g., confounding ELP score interpretation due to test administration or accommodation difficulties), resulting in a more valid and reliable assessment experience. The following sections address accessibility considerations and challenges for ELs and ELs with disabilities, and the paper closes with suggestions for ongoing and future test development, corresponding policy considerations, and possible areas of research for next-generation ELP assessments.

Current Test Accessibility and Accommodations Practices

There are several pathways to accomplish the goal of accessibility for ELP assessments. Traditionally, with discrete ELP domains (listening, speaking, reading, and writing), test takers with disabilities may have difficulty participating in the assessment of the domain that may be impacted by their disability (Christensen, Albus, Liu, Thurlow, & Kincaid, 2013).

For example, test takers with a learning disability (e.g., a reading disability) may have difficulty demonstrating their knowledge and skills on the reading section of an assessment because the measurement of one of the target skills (decoding, vocabulary knowledge, or comprehending grade-level text) could be confounded by the students' disability (i.e., decoding). In some states, such students may be asked to attempt a certain number of test items before they can move on to testing in another domain. Another option would be to incorporate the student's IEP-approved accommodation. For example, for test takers with visual impairments (e.g., blindness), items that require sight (visuals, text, etc.) would need an alternate representation. In some cases, this alternate representation could be braille (or text-to-speech, manipulatives, and tactile representation), but the representation would depend on the construct and the students' braille literacy skills. Each of these approaches and others that have been suggested (e.g., Thurlow, Liu, Lazarus, & Moen, 2005) have the potential to confound scoring and the interpretation of results.

To ensure ELs (including ELs with disabilities) have opportunity to access test material, several practices are commonly used, including universal design considerations and general test development guidelines for assessments. Developed by Thompson, Johnstone, and Thurlow (2002, pp. 6–20), universal design features for assessments include (a) acknowledgement of the inclusive assessment populations to which the assessment will be administered; (b) precisely designed constructs so each item measures what it is designed to measure; (c) accessible, nonbiased items; (d) items that are amenable to accommodations without compromising the validity of the score results; (e) simple, clear, and intuitive instructions and procedures; (f) maximum readability and comprehensibility so reading skill does not become a source of construct-irrelevant variance for nonreading items; and (g) maximum legibility so features such as font, style, spacing, contrast, and white space do not become sources of construct-irrelevant variance.

In addition to the universal design principles for assessment, *Guidelines for the Assessment of English Learners* (Pitoniak et al., 2009) include recommendations to consider for implementation in the various stages of the assessment process, including design, implementation, scoring, and score use, to maintain the validity of the target assessment for the intended population. Guiding sections include (a) assessment plans, (b) development of test items and scoring criteria, (c) external reviews of test materials, (d) task evaluation through tryouts, (e) scoring of constructed-response items, (f) testing accommodations, and (g) use of statistics to evaluate the assessment and scoring.

Although these best practices and universal design considerations promote workable paths toward resolving accessibility constraints, further research suggests that incorporating these elements into test design does not always mitigate accessibility challenges test takers may experience (e.g., Liu & Anderson, 2008). Additionally, the emphasis on accessibility and disability-based accommodations for ELP assessments is relatively new. ELP assessments are designed to measure students' language proficiency in English. In this context, disability accommodations may be more appropriate than linguistic accommodations, but accommodations policies for ELP tests are still being developed at this time and are thus beyond the scope of this paper. Analysis of states' accommodations policies in 2011 showed that 37 states indicated that an EL must have an IEP or a 504 plan to receive an accommodation on their ELP assessments (Christensen et al., 2013). Five states did not require an IEP or a 504 plan, and seven states provided no information about accommodations for ELs.

In the three states that had online ELP assessments in 2010–2011, accommodations policies for ELP assessments varied considerably. For example, large print was allowed only for the reading and writing domains in Massachusetts and in all domains in Texas, while Oregon did not provide information about the availability of the large print accommodation. Most often, these states did not provide information on whether specific accommodations could be used. For example, both Oregon and Texas provided no information about reading test directions to the student. Massachusetts indicated that reading test directions was available for the writing domain only; if the accommodation were used for other domains, it would impact both the domain-specific scores and the overall ELP score. These differences are important to consider for future accessibility and accommodations development because both ASSETS and ELPA21 will be technology-based assessments, and their accommodations policies are being developed at the present time.

Not only must test developers create an assessment that is appropriate for assessing inclusive populations, but they also must ensure that the assessment will produce a valid measurement of the rigorous standards, often by using more innovative assessment design. The following section of the paper addresses these accessibility considerations and challenges common to computer and paper-based testing, as well as technology-specific challenges for ELs and ELs with disabilities. Considerations for the heterogeneity of the target test-taking population are also explicitly addressed in this section.

Challenges Common to Computer- and Paper-Based Testing

Delivery mode is an important variable to specify early in the development process. Two main categories for test delivery are computer-based assessments and paper-based assessments. While each mode of assessment has unique implementation challenges, shared challenges also exist across delivery modes. These common challenges correspond to the heterogeneity of the test-taking population and will be discussed from this viewpoint in the following section.

Assessing First-Time and Young Test Takers

Assessing ELP for initial identification, classification, and annual monitoring of potential ELs and current ELs can be a process wrought with many challenges. Although high-stakes content assessment can be a commonly understood concept and practice in Grades 3 – 12 in the United States, unique to ELP assessments are the challenges of assessing first-time test takers (e.g., incoming kindergarten students who are 4 or 5 years old) or students who are new to this country and may have not had experience with high-stakes assessments even in their home country. For example, García Bedolla and Rodriguez (2011) described one state's ELP assessment classification and potential for misclassification, especially when the assessment experience for incoming kindergarten students may not be developmentally friendly or appropriate, such as having 2-hour testing windows or asking students to take an entire test in English when English is a new language. Although this negative experience may not be common to all grade levels, the assessment experience should be one that can elicit a valid measurement of a child's ELP without introducing confounding factors such as familiarity with testing, test-taker characteristics (e.g., affect, motivation), or requiring background knowledge to access the test items and show performance.

Assessing English Learners With Disabilities

Another challenge is the potential for the test taker's disability to interact with the target construct. For example, an EL with a reading disability may possess disability characteristics that would interfere with the traditional ELP reading subtest. IDEA (2004) defined 13 disability categories that manifest in various presentation characteristics. Across disability categories, these characteristics have potential to interact with the measurement of ELP, broadly defined.

This section is intended to provide an overview of the challenge that exists for measuring ELP for federal accountability purposes when test takers are ELs with a disability, as the disability has potential to interact with the domain score, which interacts with the overall ELP score that is used for the high-stakes purposes of identification, placement, and reclassification. Further complicating matters is the fact that many ELs with disabilities may have more than one disability.

This variability in the challenge for measuring ELP can be attributed to several causes, most notably being the individuality of the disability presentation, including the age of onset. It is important to note that certain presentation characteristics of some items within the domains of listening, speaking, reading, and writing may also create interference with the construct measurement. Each item type within and across domains should be examined closely for accessibility challenges. Likewise, given the highly individualized nature of disabilities, students should be evaluated individually for assessment participation decisions rather than assuming participation decisions at the disability category level.

The following sections expand on some of these ELP measurement challenges for selecting high-incidence and low-incidence disabilities. This is not intended to be an inclusive documentation. Instead, the purpose is to highlight key ELP measurement challenges for selecting disability categories. It is important to note that the following descriptions should not be interpreted to suggest that if ELs (with no diagnosed disability) have difficulties with one or more sections of their ELP test, ELs have a specific disability. ELP assessments are designed to measure ELP, not disabilities. Using ELP (screener or summative) performance to diagnose an EL with a disability is not appropriate and goes beyond the intended and appropriate uses of ELP assessment data. If ELs are suspected to have a disability, the appropriate referral and evaluation should be conducted with school psychologists (preferably bilingual to assess ELs in their home language, in addition to English) and multidisciplinary child study teams that include an English language development (ELD) specialist to arrive at the appropriate course of action to support students' needs.

Specific Learning Disability

Learning disability is considered a high-incidence disability. Children with a specific learning disability (e.g., dyslexia, dyscalculia, and dysgraphia) encounter assessment challenges related to the disability presentation (e.g., reading, writing,

calculation difficulties, and difficulty managing multiple details) that can be compounded when measuring multiple constructs in a single task (e.g., a speaking task based on listening stimuli or a constructed-response writing task used to assess both reading comprehension and writing skills). For example, when taking the reading domain subtest on an ELP assessment, a student with a disability that affects decoding text (e.g., dyslexia) may benefit from having the writing prompt read aloud. But, if this task is also designed to measure reading comprehension, the students' read aloud accommodation may affect the validity of the reading-writing score (i.e., if reading is defined to include decoding and reading fluency as well as comprehension). Following this example, numerous studies have examined the effect of audio presentation of test content for tests of English language arts and mathematics (e.g., Buzick & Stone, 2014; Higgins & Katz, 2013; Higgins, Russell, & Hoffman, 2005; Laitusis, 2010), but research on accommodation effects for ELP assessment practices for ELs with disabilities is underrepresented in the empirical literature.

Visual Impairment

For students who have visual impairments (i.e., blind or low vision), the challenges of assessing language skills are compounded by the common use of visual images or text that would require braille production or the application of a different accessibility feature to elicit language. Traditional ELP items for certain domains assess the construct by showing a picture of several items and asking the student to read and match words to the objects (a reading domain task) or do more complex tasks such as watching a video and then describing what happened in the video (a speaking domain task). The process of making items accessible includes some challenges that are compounded when the student has poor braille literacy (i.e., decoding skills, e.g., a student with later onset vision loss may not be proficient in the use of braille or tactile figures), so providing braille or text descriptions of visual images is ineffective since the student is just learning how to read them, or (depending on the construct) the descriptions have the potential to cue the correct answer.

Hearing Impairment

An under-researched challenge is the practice and appropriateness of assessing ELP for students who are hard of hearing or deaf. While some constructs (listening) may be eliminated from the assessment, it may be possible to include substitute constructs or allow accommodations to make the items assessing the construct accessible (e.g., substituting a human reader instead of an avatar for listening items). However, this practice would be contingent upon the assessment's construct definition because the practice would have implications for the overall assessment claims. Further, additional research is needed to determine the appropriateness of ELP/ELD standards as they pertain to children with hearing impairments because it is possible children who are ELs with hearing impairments may acquire English at different stages (i.e., following different developmental pathways) compared to ELs who are able to hear (e.g., Hoffmeister & Caldwell-Harris, 2014).

Speech/Language Impairment

Speech and language impairment (SLI) is a disability category that encompasses a range of specific characteristics, including (but not limited to) lisping, stuttering, selective mutism, and significant language issues (Thurlow et al., 2009). Unlike other disabilities that may not be diagnosed until students have had more experience in a school setting, SLI is a disability that is more often identified in the elementary grades, compared to a specific learning disability that is more often identified in the middle or high school grades (see Zehler et al., 2003). Given the four domains typically assessed in an ELP assessment (listening, speaking, reading, and writing), it is clear that a student with this disability will possess characteristics that are likely to interact and interfere with an accurate measurement of his or her speaking ability. To enable students with SLI to participate in the ELP assessment to the extent they are able, certain accommodations may be identified in these students' IEPs that allow them access to the assessment as much as is deemed appropriate. In other words, instead of being required to speak responses, if the response options are represented by a visual or text, the test taker could be allowed to point to the desired response. Additionally, if the test taker uses any type of augmentive/assistive communication (AAC) device, his or her accommodations may allow for an AAC device or a scribe to help produce responses. Under a traditional construct definition and strict interpretation, these accommodations do not traditionally represent the speaking construct. However, this is how students who use these devices speak. This scenario illustrates the need

for accessibility considerations at the early stages of ELP test development, including the development of the conceptual framework and construct definition. This early awareness may better support stakeholders to make appropriate student participation decisions and more meaningful score interpretations for ELs using AAC devices.

Autism Spectrum Disorder

Students diagnosed with autism spectrum disorder (ASD) possess characteristics that vary across a wide spectrum of severity, ranging from mild social/communicative delays to some children being unable to speak, read, or write. Other characteristics that can manifest in students identified on the spectrum include echolalia (i.e., repeating utterances heard from other sources of input such as a person, television show, or song), and other speech delays associated with pragmatics (e.g., using social conventions of personal space or turn-taking behaviors in conversations), initiations (e.g., initiating conversations), and interrogatives (i.e., question asking). Echolalia has the potential to be problematic in domains requiring productive expression due to the students' tendency to repeat words or phrases of interest. Also, in the context of language assessment, any stimulus describing hypothetical scenarios (a common characteristic of some prompts) may have potential to lead the test-taker astray because of the possibility of literal interpretations and meaning making. Additionally, tasks that require the test taker to engage in simulated social interaction complete with turn-taking behaviors via a conversation, presentation, or interview context may also interfere with the social-communicative challenges typically associated with the disability. However, due to the largely individualized nature of this particular disability, participation decisions should be reserved for a one-on-one basis by the appropriate educational team (e.g., IEP team).

Intellectual Disability

This category is still referred to as mental retardation (MR) in most federal laws. This category is recognized now as *intellectual disabilities* after Rosa's Law was signed in 2010. Regardless of the label, this disability covers a range of presentation characteristics, from mild to moderate to significant. This range of abilities makes the measurement of ELP difficult at the domain, as well as the overall score level. ELs with intellectual disabilities are estimated to comprise approximately 7% of all students with significant cognitive disabilities and anywhere from 3% to 36% of selected states' populations of students with significant cognitive disabilities (Towles-Reeves et al., 2012). When surveyed, a majority of responding states indicated that their ELs with significant cognitive disabilities participated in either some or all of an ELP assessment. ELs with significant cognitive disabilities also were noted to participate in an alternate assessment (e.g., WIDA's Alternate ACCESS) or not participate in the ELP assessment process at all (Rieke, Lazarus, Thurlow, & Dominguez, 2013).

Despite the prevalence of ELs with intellectual disabilities, few approaches to measuring ELP standards against alternate performance standards exist that would guide appropriate item development to assess ELs with intellectual disabilities in the more moderate to significant range. Experts participating in focus groups to generate general principles and guidelines for assessment of ELs confirmed their belief that there is a need for an assessment of English proficiency specifically designed for students with intellectual disabilities (Thurlow, Liu, Ward, & Christensen, 2013). As noted previously, federal mandates now also make explicit the need for all ELs, including ELs with intellectual disabilities, to take part in the ELP assessment process. Access to a range of appropriate ELP assessments (including alternate screener and summative assessments measuring alternate achievement standards) is needed so that IEP teams can make appropriate participation decisions that benefit the test taker. Developing alternate assessments to measure ELP for ELs with significant cognitive disabilities would allow students to participate in the assessment and identification process to receive the supports needed to acquire the English language, rather than risk potential exclusion from language services because there is no assessment appropriate for ELs with intellectual disabilities.

Technology Innovations and Accessibility Challenges

Technology provides opportunities to make assessments more accessible for students with disabilities while also enhancing the measurement of the target construct. At the same time, technology expands item type presentation, response functionalities, and accessibility considerations, which in turn present unique accessibility challenges for both ELs and ELs with disabilities, as described below.

Young Test Takers and Technology Novices

Recent overviews of state ELP test administrations and the shift from state-based assessments to consortia-based assessments (e.g., ASSETS; ELPA21) suggest that ELP test administration is moving toward computer-based testing development and administration (Christensen et al., 2013). This poses significant new questions about the feasibility and validity of computer-based measures of ELP. In other words, technology should enhance the measurement of the construct, not to drive (or limit) the definition of the construct (Hauck et al., 2016). This is often exemplified using drag-and-drop opportunities (or click-and-click functionality) that may better measure students' listening skill for following directions (e.g., test takers can demonstrate the act of selecting and organizing appropriate materials for a class project as the teacher reads the list aloud).

Keyboarding

Writing can be assessed with both selected-response and constructed-response opportunities, mediated either by handwriting or keyboarding skills. In the United States, test takers in the target population begin taking ELP assessments in kindergarten, but students are also identified as ELs through the 12th grade (i.e., late arrivals, newcomers) and may be technology novices with limited exposure to formal educational opportunities. The heterogeneous nature of the EL population calls for critical consideration of whether computers, let alone keyboarding, are tools that young learners and technology novices are exposed to so that technology enhances the construct measurement and does not act as a source of construct-irrelevant variance.

The Digital Divide

The heterogeneity of the target population for ELP assessments is a variable that should be addressed when considering assessment mode of delivery. Previous research has found differences in the age at which students receive keyboarding instruction in Grades K–6 and found that overall, as grade level increased, access to keyboarding instruction slightly increased, peaking at 4th grade (Rogers, 1997). Other differences include a potential racial/ethnic and or economic divide in computer access and Internet use, as well as differences in device access, which includes computers in addition to laptops, smart phones, tablets, and gaming consoles (National Center for Education Statistics in 2003, as cited in Morgan & VanLengen, 2005; Project Tomorrow, 2014; U.S. Census Bureau, 2014). The heterogeneity of the population and variety in technology and device exposure suggests additional challenges for assessment development. The features of a computer, which require keyboarding and mousing skills, are very different from tablet features that support drag-and-drop and soft-touch (i.e., haptic) keyboarding.

Technology-Enhanced Accessibility Features

Administering assessments via computer provides opportunities for a more standardized approach to delivering student accommodations. In other words, there is less demand for a test administrator to administer an accommodation external to the test (e.g., a teacher reading the test aloud for test takers who are deaf or hard of hearing but read lips). Instead, item content can be flagged prior to test administration in a way that allows for any item content with a listening component to also call up the IEP-approved accommodation (in this hypothetical case, closed captioning), when the listening stimulus is presented. This is accomplished via accessible portable item protocol (APIP) content tagging (IMS Global Learning Consortium, n.d.). APIP promotes interoperability standards for transferring accessibility information with the item and recording the accessibility needs of the test taker. For ELP assessments in particular, there is a benefit with this approach, as various accessibility features or accommodations like a read aloud can be digitally delivered, minimizing the threat of nonstandard delivery or cuing from the reader. Additionally, in instances where the read aloud may cue the test taker to the correct answer, the read aloud can be automatically programmed to be shut off. For example, traditional ELP-item types are often accompanied by a graphic (this graphic can be necessary to answer the item correctly, or it can be decorative). For students with visual impairments, mousing over this graphic could show alt text (i.e., alternate text format, or ATF, a short picture description) describing the picture if it cannot be made accessible with a tactile graphic. For items with graphics that are construct relevant, it is possible that by providing alt text (i.e., labels) or voicing the labels for the visuals, the correct response may be given away.

While computer delivery allows for these technology-enhanced accessibility features, additional challenges exist for a feature such as alt text due to the need to make the alt text accessible for beginning readers (kindergarten, first grade, second grade) with varying levels of ELP. Students also need to be familiar with the use of alt text, and so it is imperative that classroom instruction also include opportunities to interact with alt text. This is particularly important for students who are in the early primary grades or students (i.e., newcomers, late arrivals) who may have had less experience with formalized schooling or who may be technology novices.

Although the benefit of APIP appears obvious for the test administrator, there remain some challenges to development and implementation. For example, cost and development procedures are a challenge for many test development efforts. Another challenge is the lack of student familiarity with the manner of the delivered accommodations. Also important to mention is that currently, despite its overwhelming popularity across the assessment consortia, APIP is not required by federal law and has not fully integrated the W3C Web Content Accessibility Guidelines (WCAG) that are required for accessible digital delivery of computer-based assessments. Regardless, APIP remains a desirable component for computer-based testing administration. Finally, with the supposed ease of implementing APIP, it is critical that test developers and clients conceptualize the effect APIP could have on the target construct and provide support for students if they have regular access to these features in the context of their everyday classroom. That is, APIP should not be adopted and administered as a cure-all for accommodation delivery, but instead the accommodations should be administered judiciously to preserve the validity of measurement of the target construct. In other words, allowing certain accommodations on an ELP test may change what the item is intended to measure and may have an impact on the overall calculation and interpretation of both overall composite score and domain scores.

Possible Directions to Pursue

Mandates issued from the U.S. Department of Education make explicit the federal requirement to include ELs and ELs with disabilities in general ELP assessment or alternate ELP assessment practices, without exception. Accordingly, current and next-generation ELP assessments must address the accessibility needs of the diverse population of students who will participate in these ELP assessments. Supporting the accessibility needs of these populations would ensure valid assessment of the skills being tested and support the high-stakes uses of the assessment, since these ELP test scores are used for federal accountability purposes, as well as for EL identification, instructional placement, and reclassification or exiting the EL category.

Attention to accessibility at the early stages in test development and throughout the development process can help minimize the effects of construct-irrelevant variance (i.e., information the test is not intended to measure), resulting in a more valid and reliable assessment experience. This paper addressed these accessibility considerations and challenges for ELs and ELs with disabilities and provided suggestions for ongoing and future test development and policy considerations to improve the ELP testing experience for all users. The following section will elaborate on possible directions to pursue for future test development, policy consideration, and research practices.

Accessibility

Accessibility is for all test takers. Many of the challenges discussed in this paper apply to ELs with disabilities, as well as the entire range of the EL population. For ELP assessment purposes, there is a need to broaden the definition of accessibility to include young (transitional kindergarten, kindergarten) test takers and first-time test takers (students arriving at all grade levels, not just at the kindergarten level), including technology novices. Acknowledging this heterogeneity within this EL population calls attention to the need for thoughtful and intentional identification of the construct definition and assessment purpose. Knowledge of the characteristics of the target population should help guide decisions surrounding elements such as mode of delivery, item types, and allowable accommodations to support the construct definition and intended assessment purpose. Additionally, accessibility features and accommodations (e.g., technology-embedded supports, alt text) should take into account this population's heterogeneity (e.g., technology novices) early in the test-design process to validate claims that the accessibility features actually work as intended to promote EL's accessibility to the content being tested.

Additionally, there is a need to reconcile the tension between assessment innovations to better measure ELP and the need to be accessible for a wide-range of test takers, as required for federal accountability purposes. These innovations

must include accessibility and accommodations experts from the beginning to identify (a) challenges and (b) paths to resolution to ensure the assessment innovation is not limited, but rather so the innovation can be accessible to all students. This includes the need to consider supports to promote accessible test development, such as redefining the universal accessibility features used in the test development process (e.g., heavy use of visuals) so that assessments are truly accessible to the entire test-taking population.

Adaptive English Language Proficiency Assessments

To meet the needs of a diverse K – 12 student population, it is important to consider the advantages of adaptive testing over a linear form. Task difficulty is a common consideration when creating item pools for linear and adaptive test forms. Difficulty in traditional selected-response items is easier to design and manipulate compared to tasks measuring integrated skills. For example, it is necessary to define the task difficulty on several levels: Will the item employ comparable difficulty across the skill domains (listening, speaking, reading, and writing)? Will the item employ social language in the beginning and progress to academic language? Will the item be designed to elicit receptive skills initially and then progress to more productive skills? These decisions impact the types of items and task designs necessary to elicit authentic language interactions from test takers, as several items could be embedded in tasks that may assess one or more standards.

While some states or consortia may opt to provide a linear form to students, others may choose an adaptive assessment. If states or consortia do choose to create a computer adaptive assessment, there are particular challenges to create accessible test content (e.g., the need to have larger item pools for content spanning several difficulty levels), but having to produce accessible content on demand is an added challenge. One approach taken by consortia and states for their content assessments (Smarter Balanced, Oregon, and Utah) has been to create digital alternate format materials for a pool of items and then print (large print or paper version of item) or emboss (braille or tactile drawing) the item in real time. This approach allows students with disabilities to reap the benefits of computer adaptive testing, like reduced testing time and increased opportunity to interact with appropriately leveled testing content, although the challenges include printing or embossing that is not as high quality as test materials printed ahead of time (e.g., tactile embossers that are affordable for school purchase cannot emboss curved lines as well as professional grade embossers). Another approach, used in the development of the Kansas content assessments, is to create a multistage adaptive assessment and then produce one braille or large print form for each stage of the assessment. This approach requires that teachers enter student responses into the computer and then retrieve the correct test form for each section of the test, but allows for professional grade embossing and for the students to participate in the adaptive assessment.

Policy Implications for English Learners With Disabilities

Previous research has noted the high variability in identifying, tracking, and accommodating ELs with disabilities (Christensen et al., 2013). Given the federal legislation mandating annual assessment and accountability for students identified as ELs, students with disabilities, and ELs with disabilities (IDEA, 2004; No Child Left Behind Act of 2001), there is a critical need to identify and track ELs and ELs with disabilities.

Consolidated state performance reports for 2011 – 2012 indicated the number of ELs with IEPs ranged from 0% to 31% of the EL student population in the 50 states. Although disability classification is indicative of a diverse category of students, this huge range suggests there may be differences in identification practices and subsequent instructional service allocation for ELs with disabilities across the United States. Another possibility for the difference may be due to the proportionality of EL identification and special education referrals within and across states. However, the recent request for information to identify research areas to protect against overrepresentation of minorities in special education (Yudin, 2014) suggests states may be grappling with disproportionality and overrepresentation (Artiles & Klingner, 2006) related to issues associated with accurate and appropriate EL identification and classification for ELs at risk (Guzman-Orth, Nylund-Gibson, Gerber, & Swanson, 2014; Klingner, Artiles, & Barletta, 2006). More research is clearly needed in this area to identify ELs from ELs with disabilities to be able to better serve them in the instructional and assessment context. This identification and subsequent tracking has implications for district, state, and federal data management systems as systems will have to communicate across high-stakes assessments (e.g., content assessment, ELP assessment) to ensure that students' scores and accommodations are being collected appropriately. Additional implications exist for the federal funding allotted to schools and districts serving students in these special subgroups, as well as at the practitioner level, with the need to ensure appropriate instructional service allocation to meet students' needs.

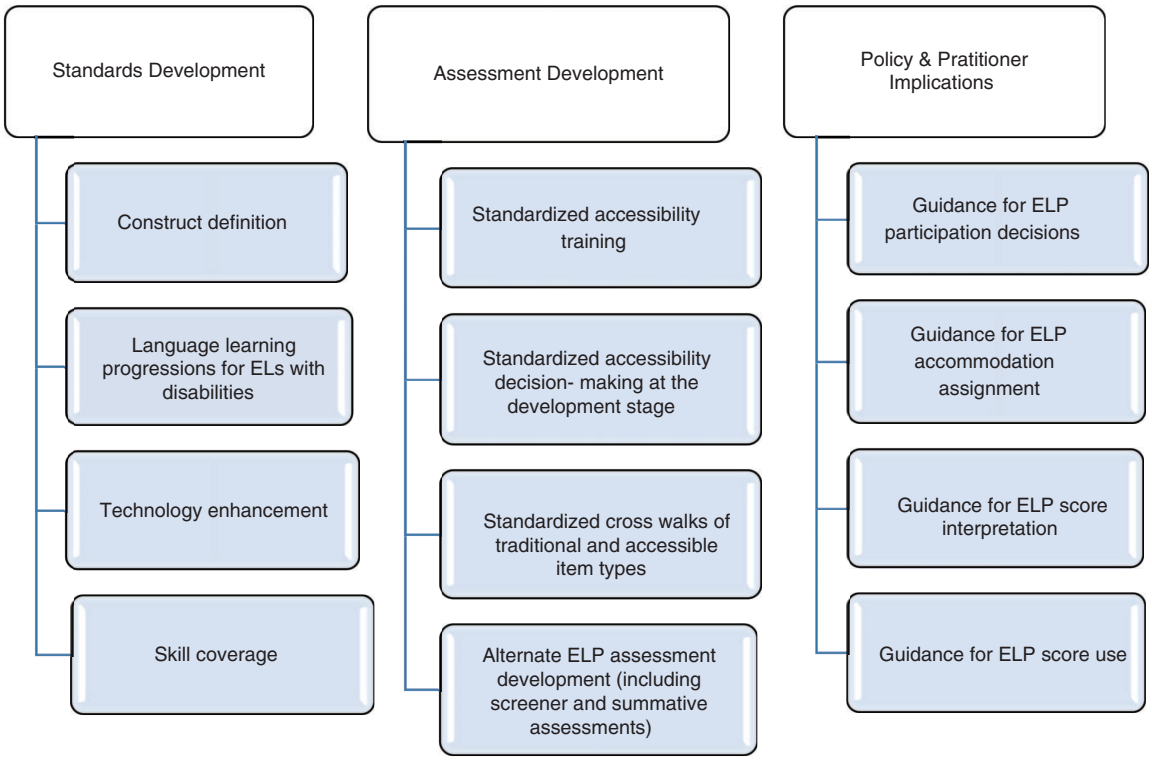

Figure 2 Suggested paths for research and policy to address challenges of diverse accessibility needs for English learners and English learners with disabilities taking English language proficiency assessments.

Research Implications

Although accessibility is an area in need of research to validate assessment development and administration practices for ELs and ELs with disabilities, a clear research and policy agenda has emerged revolving around three main issues of standards development, assessment development, and practitioner implications (see Figure 2).

More research is necessary to establish an empirical foundation to validate the ELP construct definition and learning progressions (i.e., sequential pathways for students to learn English for various purposes) for ELP assessments for children who are ELs with disabilities. Disability experts know that language acquisition for atypically developing children may progress differently from language acquisition for typically developing ELs (e.g., children with hearing impairments may learn English grammar much later compared to same-aged peers who are able to hear). Additionally, assessment development needs to consider standardization of accessibility training for ELP assessment developers, in addition to a standardized approach to a decision tree to guide test developers about when to focus efforts on traditional ELP measurement approaches and when to consider accessible item development alternatives, such as twin items (i.e., items measuring similar constructs in different ways), separate forms (i.e., a separate, accessible, assessment form generally intended to be comparable to the original form), adaptive testing, or an alternate assessment for ELs with significant cognitive disabilities. Additionally, assessment developers are in need of cross walks, that is, empirically validated accessible item alternatives for traditional (nonaccessible) items within each domain of listening, speaking, reading, and writing to support the practice of developing accessible assessments from early in the test-design blueprint stage. Although existing practices of not administering items or whole domains exist (Christensen et al., 2013), there is also a need to develop empirically validated guidelines to support these practices, as well as the practice of developing separate, comparable forms and scoring tables to ensure the resulting score is still valid for high-stakes accountability purposes. Additional practices, such as using decision trees to guide decision making to identify whether items are accessible with certain accommodations or within each domain, are also important to validate to maintain a principled approach to accessibility decisions for ELs and ELs with disabilities. Practitioner implications call for guidance for participation decisions, accommodation assignment, score interpretation, and score use (e.g., decision making for appropriate service allocation) to improve the ELP assessment experience for stakeholders at the federal, state, and local levels.

Conclusion

Accessibility is an important variable to recognize from the early stages of test development for all assessments and assessment purposes. Further, there is a need to recognize accessibility and administration procedures as multifaceted and fluid variables that need to be considered in order to meet the needs of the heterogeneous population of ELs. This paper only focused on accessibility of ELP assessments as this is an area that has unique challenges and considerably less research has been conducted on accessibility of ELP assessments. Although this paper illuminates challenges and possible directions to pursue to address accessibility challenges for ELP test administration, a critical need remains for established empirical evidence to validate accessible ELP-item development for ELs and ELs with disabilities, practices for assessing ELP for ELs and ELs with disabilities, and appropriate ELP assessment uses for students who are ELs and ELs with disabilities.

Acknowledgments

This paper is a collaborative effort between ETS and NCEO. It is supported by research funding from ETS and a cooperative agreement from the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education (#H326G110002) awarded to NCEO. The contents of this report represent the views of the authors and do not necessarily represent the views of the funding agencies. The authors would like to thank the reviewers for their thoughtful comments on the earlier versions of this paper: Louis Danielson (AIR), Vincent Dean (ETS), David Egnor (NCEO), Georgia Garcia (University of Illinois), Kenji Hakuta (Stanford University), Maurice Hauck (ETS), Alexis Lopez (ETS), Scott Paris (ETS), Rose Payan (ETS), Don Powers (ETS), Charlene Rivera (George Washington University), Mikyung Wolf (ETS), and John Young (ETS). A previous version of this report was published in October 2014 on ets.org.

References

Artiles, A. J., & Klingner, J. K. (2006). Forging a knowledge base for English language learners with special needs: Theoretical, population, and technical issues. *Teacher College Record, 108*(11), 2187 – 2194.

Buzick, H. M., & Stone, E. A. (2014). A meta-analysis of research on the read aloud accommodation. *Educational Measurement: Issues and Practice*, *33*(3), 17-30.

Christensen, L. L., Albus, D. A., Liu, K. K., Thurlow, M. L., & Kincaid, A. (2013). *Accommodations for students with disabilities on state English language proficiency assessments: A review of 2011 state policies*. Minneapolis: University of Minnesota, Improving the Validity of Assessment Results for English Language Learners with Disabilities (IVARED).

García Bedolla, L., & Rodriguez, R. (2011). *Classifying California's English learners: Is the CELDT too blunt an instrument?* Retrieved from the Center for Latino Policy Research website: http://www.escholarship.org/uc/item/2m74v93d

Guzman-Orth, D. A., Nylund-Gibson, K., Gerber, M. M., & Swanson, H. L. (2014). *The classification conundrum: Identifying English language learners at-risk*. Manuscript submitted for publication.

Hauck, M. C., Wolf, M. K., & Mislevy, R. (2016). *Creating a next-generation system of K-12 English learner (EL) language proficiency assessments* (Research Report No. RR-16-06). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12092

Higgins, J., & Katz, M. (2013). A comparison of audio representations of mathematics content. *Journal of Special Education Technology, 28*(3), 59 – 66.

Higgins, J., Russell, M., & Hoffman, T. (2005). Examining the effect of computer-based passage presentation on reading test performance. *The Journal of Technology, Learning, and Assessment, 3*(4), 1 – 35.

Hoffmeister, R. J., & Caldwell-Harris, C. L. (2014). Acquiring English as a second language via print: The task for deaf children. *Cognition, 132*(2), 229 – 242. doi:10.1016/j.cognition.2014.03.014

IMS Global Learning Consortium. (n.d). *Accessible portable item protocol (APIP)*. Retrieved from http://www.imsglobal.org/apip/index.html

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English language learners who struggle with reading: Language acquisition or LD? *Journal of Learning Disabilities, 39*(2), 108 – 128. doi:10.1177/00222194060390020101

Laitusis, C. C. (2010). Examining the impact of audio presentation on tests of reading comprehension. *Applied Measurement in Education, 23*, 153 – 167. doi:10.1080/08957341003673815

Liu, K. K., & Anderson, M. (2008). Universal design considerations for improving student achievement on English language proficiency tests. *Assessment for Effective Intervention, 33*, 167 – 176. doi: 10.1177/1534508407313242

Lopez, A. A., Pooler, E., & Linquanti, R. (2016). *Key issues and opportunities in the initial identification and classification of English learners* (Research Report No. RR-16-09). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12090

Morgan, J. N., & VanLengen, C. A. (2005). The digital divide and K-12 student computer use. *Issues in Informing Science and Information Technology, 2*, 705–722. doi:10.1016/0378-7206(93)90038-u

No Child Left Behind Act of 2001, 20 U.S.C. § 6301 *et seq.* (2002).

Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.

Project Tomorrow. (2014). *The new digital learning playbook: Understanding the spectrum of student's activities and aspirations*. Retrieved from http://www.tomorrow.org/speakup/pdfs/SU13StudentsReport.pdf

Rieke, R., Lazarus, S. S., Thurlow, M. L., & Dominguez, L. M. (2013). *2012 survey of states: Successes and challenges during a time of change*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Rogers, H. D. (1997). A longitudinal study of elementary keyboarding computer skills. *Academy of Educational Leadership Journal, 1*(2), 55–57.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (NCEO Synthesis Report No. 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Liu, K. K., Lazarus, S. S., & Moen, R. E. (2005). *Questions to ask to determine how to move closer to universally designed assessments from the very beginning, by addressing the standards first and moving on from there*. Retrieved from the Partnership for Accessible Reading Assessments website: http://www.readingassessment.info/resources/publications/QuestionstoAskUniversallyDesignedAssessmen

Thurlow, M. L., Liu, K. K., Ward, J. M., & Christensen, L. L. (2013). *Assessment principles and guidelines for ELLs with disabilities*. Minneapolis: University of Minnesota, Improving the Validity of Assessment Results for English Language Learners with Disabilities (IVARED).

Thurlow, M. L., Moen, R. E., Liu, K. K., Scullin, S., Hausmann, K. E., & Shyyan, V. (2009). *Disabilities and reading: Understanding the effects of disabilities and their relationship to reading instruction and assessment*. Minneapolis: University of Minnesota, Partnership for Accessible Reading Assessments.

Towles-Reeves, E., Kearns, J., Flowers, C., Hart, L., Kerbel, A., Kleinert, H., … & Thurlow, M. (2012). *Learner characteristics inventory project report: A product of the NCSC validity evaluation*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

U.S. Census Bureau. (2014). *Computer and Internet trends in America*. Retrieved from http://www.census.gov/hhes/computer/files/2012/Computer_Use_Infographic_FINAL.pdf

U.S. Department of Education. (2014a, July). *Questions and answers regarding inclusion of English learners with disabilities in English language proficiency assessments and Title III annual measurable achievement objectives*. Retrieved from http://www2.ed.gov/policy/speced/guid/idea/memosdcltrs/q-and-a-on-elp-swd.pdf

U.S. Department of Education. (2014b, April). *SY 2011–2012 consolidated state performance reports part I*. Retrieved from http://www2.ed.gov/admins/lead/account/consolidated/sy11-12part1/index.html

Wolf, M. K., Guzman-Orth, D., & Hauck, M. C. (2016). *Next-generation summative English language proficiency assessments for English learners: Priorities for policy and research* (Research Report No. RR-16-08). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12091

Yudin, M. K. (2014, June 16). Request for information on addressing significant disproportionality under Section 618(d) of the Individuals With Disabilities Education Act (IDEA) [online Federal Register]. Retrieved from https://www.federalregister.gov/articles/2014/06/19/2014-14388/request-for-information-on-addressing-significant-disproportionality-under-section-618d-of-the

Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., & Sapru, S. (2003). *Policy report: Summary of findings related to LEP and SPED-LEP students*. Washington, DC: U. S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement of Limited English Proficient Students.

Suggested citation:

Guzman-Orth, D., Laitusis, C., Thurlow, M., & Christensen, L. (2016). *Conceptualizing accessibility for English language proficiency assessments* (Research Report No. RR-16-07). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002.ets2.12093

Action Editor: Donald Powers

Reviewers: Alex Lopez and John Young