Assessing the Language of Young Learners

Alison L. Bailey

Abstract

Developmental realities of young language learners (approximately ages 3–11 years) pose unique and challenging considerations for oral and written language assessment. This chapter first addresses construct definitions important for establishing a common understanding of language testing with young children. Both type (e.g., summative, formative) and purpose (e.g., accountability, diagnostic) of language testing commonly used in preschool and elementary (primary) school contexts are discussed. Developmental and cultural factors that need to be taken into account in assessing young learners are addressed. I conclude with work in progress and future directions for language test development involving empirically derived trajectories of language development, accounting for the intersection of language and content learning, the professional development of teachers around language learning and assessment, and further innovations with technology.

Contents

A.L. Bailey (✉)
Human Development and Psychology Division, Department of Education, University of California, Los Angeles, CA, USA
e-mail: abailey@gseis.ucla.edu; bruinprofab@gmail.com

Introduction

The field of language testing has recently seen new volumes and chapters dedicated to discussion of assessment with young language learners (e.g., Bailey et al. 2014; Nikolov 2016), compendiums characterizing and evaluating available language assessments for young learners (Barrueco et al. 2012; Jang 2014), and a burgeoning research agenda that as included studies of student self-assessment, the intersection of language and academic content learning, and further exploration of technology-assisted assessment with children. In part, this is likely a consequence of the increased communicative demands placed on school-age language learners. Young language learners not only encounter the assessment of their language development but also assessment of other learning and knowledge through the language(s) they are still acquiring. For example, students are now being expected to display academic content knowledge through oral and written explanations and argumentation both in daily classroom-based tasks and in summative assessments that are tied to new academic content standards (e.g., Bailey and Heritage 2014).

The shift in expectations for language demands has been led in the USA primarily by the Common Core State Standards Initiative of the National Governors Association Center for Best Practices, Council of Chief State School Officers (CCSSO 2010), and Next Generation Science Standards (NGSS Lead States 2013) and by similar accountability initiatives in other countries, for example, in the UK, new tests aligned to the National Primary Curriculum are expected in 2016 (Department for Education 2016)

Such explicit expectations for language competencies make a chapter devoted to the language assessment of the very youngest language learners more critical than ever. More than a decade ago, Inbar et al. (2005) called for "specific age appropriate and language level considerations" (p. 3.) in order to address the differences between testing the language learning population at large and testing young language learners. Much of what has been examined elsewhere in this volume is reconsidered

in this chapter from the points of view of those who must create valid (i.e., fair and effective) tests for assessing the language of young learners and those who must administer and interpret them. These viewpoints require familiarity with testing purposes and an understanding of developmental and cultural issues as they impact the design and use of language assessments with young children.

While not exclusively the case, this chapter deals predominantly with tests of students' English language development (ELD) or proficiency (ELP). This is a reflection of both the increasing number of young children learning English in various contexts around the world (Graddol 2006) and the fact that much research has been conducted on the assessment of English.

The chapter is organized around five main sections: First, I provide construct definitions that will prove important for establishing a common understanding of testing issues with young children, starting with a definition of the term "young learner" itself. Second, I review the types (e.g., summative, formative) and purposes (e.g., accountability, diagnostic) of language testing in preschool and elementary (primary) school contexts. Third, I address the developmental child level concerns that need to be taken into account in assessing this population of test takers, including a review of general guidelines and best practices for assessing young children. In the fourth section, I consider culture as an additional contextual factor that, while possibly impacting all language testing situations, may have particular significance for the testing of young children. Finally, in the fifth section, I conclude with updates on how the field has progressed over the past decade, current works in progress, and future directions for test development.

Construct Definitions

Many key constructs already encountered in other chapters will need special definition in the context of assessing young language learners.

Defining Young Language Learner

I start with the most crucial of all definitions for this chapter, that of the *young language learner*. Defining young language learner is complicated by the range of language learning experiences, the range of ages to be covered by the qualifier "young," and the fact that in different parts of the world, different school systems introduce students to second language and foreign language instruction at different points in their school careers. In Europe, young learner is often applied to students in only the very earliest school years (ages 5–7) or before. In the USA, where the introduction of foreign language teaching often does not take place until the secondary grades, the notion of a "young learner" can span the entire preschool and elementary years (ages 3–11). Obviously for second language learners, the

onset of a second language can start before the start of formal schooling or at any time during the primary school years for those who emigrate as school-age language learners.[1] Much of the focus of this chapter, however, will be on young learners from preschool through the earliest elementary years.

Defining the Language Learning Context

Turning next to language construct definitions, prominent among these are English as a second or additional language (ESL or EAL), bilingualism, and due to the demand for English in non-English- speaking countries around the world, English as a foreign language (EFL).[2] Second and foreign languages other than English will also be pertinent to a broader discussion of young language learners everywhere. Language assessment for young monolingual speakers is primarily confined to the literate uses of a language (e.g., reading and writing), with the exception of instances when a language disability is suspected or has been diagnosed for intervention and monitored for improvement.[3] There are few assessments of oral language proficiency in a first (and often only) language, and yet the increased language demands placed on students in schooling contexts will affect *all* students not just those learning an additional language.

Second language acquisition (SLA) such as ESL is made more complex in the young learner context by the existence of bilingual first language acquisition (BFLA) (De Houwer 1998), in which children may be acquiring two languages, each as a native language. As they enter formal schooling environments, including preschool, these children may become literate in only one of their two languages if the schooling system favors one language over the other or if parents do not opt to enroll their children in dual language programming. The language learning experiences of young children may also be characterized by immersion in a second language they are yet to acquire. In Canada, for example, children have the opportunity to learn English and French (and other desired languages) in this environment from an early age (see Bailey and Osipova 2016 for review of educational options with young language learners).

EFL (and other foreign language acquisitions) characterizes learners who acquire a language after their native language has already been acquired, but do so outside an

---

[1]Young learners of English are variably referred to as English learners (ELs), English language learners (ELLs), English as a second (ESL) or additional language (EAL) students, students with non-English-speaking backgrounds (NESB), or, most recently, emergent bilinguals or dual language learners (DLLs) to reflect that many young students are acquiring English and a home language (García 2009). Throughout this chapter, I simply use young language learner.

[2]ELD will be used throughout this chapter to refer to the process of English acquisition regardless of whether it is being acquired as a second or foreign language.

[3]Review of such clinical assessments is outside the scope of this chapter (see Conti-Ramsden and Durkin 2012 and Dockrell and Marshall 2015 for recent reviews of language assessment with children with language learning disabilities).

English (or other target L2) environment. For the very youngest preliterate language learners, this may mean learning a foreign language without the aid of the print medium that is available to older children and adult learners. Older learners can garner literacy abilities in their L1 to augment their learning of oral English, as well as transfer print skills in their L1 to reading and writing in English. The latter is particularly enhanced if their L1 shares the same orthography and possibly even cognate words with English.

Defining Language Varieties

This chapter adopts a broad definition of language including all four modalities of listening, speaking, reading, and writing and, where relevant, further denotes sub-skills such as phonological awareness and pronunciation. Additional construct definitions that need to be taken into account in the assessment of young learners include the social and academic language constructs (Cazden 2001; Chamot 2005). While the distinction between the language used in a scholastic environment and the language used in everyday (out-of-school) contexts may not be as great during the early years of schooling as it is once children begin to take discipline-specific classes (e.g., history, algebra), the distinction arguably still exists. With increasing preschool enrollment worldwide, more young children have been affected by ties between opportunities for preschool language development and later academic outcomes. Working in preschool settings in Europe, Michel and Kuiken (2014) have found that preschool environments place unique demands on the language of young learners and consequently require appropriate ways to assess the language development of the very youngest of students.

The existence of an academic language construct is not without controversy, however, especially in what constitutes fair assessment of the obvious scholastic uses of language at this young age – emergent reading and writing. *Should the reading and writing skills in English of young learners be assessed differently from those of native English students who are also just beginning to learn to read and write?* If young English learners are already literate in their L1, there are implications for how we assess their literacy in English. Environmental print (i.e., sight words) from the content areas such as science, mathematics, and history may make appropriate content for assessing the literacy abilities of young school-age learners. However, there may be no positive transfer for literacy skills from children's L1 to their L2 if the orthographies of the two languages do not match (Bialystok et al. 2005), although more recently Gottardo et al. (2006) report significant correlations in the phonological processing of young language learners whose languages do not share orthographic systems (e.g., Chinese and English). Reading and writing modalities may, however, still be problematic in other ways when operationalized for testing their development in young children. For example, reading and writing are frequently tested orally which requires children to *listen* to directions not simply demonstrate their literacy abilities. Conflating these skills may result in ambiguous information for teachers wishing to effectively target their instruction. Finally, the

academic language construct may not be imperative for acquiring and displaying learning in content areas when children can effectively convey their mathematics, science, history learning, etc., using all linguistic resources at their disposal including use of L1, as well as everyday and nonstandard varieties of L2 (e.g., Faltis 2015).

Types and Purposes of Assessment

As with assessments developed for use with older children and adults, there is a range of purposes for language assessment with young learners. Due to the maturational constraints and the need for developmentally relevant measures, we witness far greater variety in the purpose and use of informal assessment in this young population.

High-Stakes Assessment

With standardized assessments, the content is a sampling of all that a student may have been taught in a given period. These assessments are summative of knowledge gain and are often considered "high stakes" for the student (e.g., a deciding factor in being reclassified as a fluent English speaker for instructional placement) or "high stakes" for those who educate them (e.g., evaluation of teacher or school performance). Also considered "high stakes" but not summative are assessments designed to screen a student's abilities for weaknesses that need immediate amelioration or flagged for possible future attention. Such screening purposes can also be considered "high stakes" for both the individual and the schooling system. An individual needs to be accurately identified for further instruction or services if these are necessary to their development. These are the cases when the schooling system also needs accurate information; providing services to individual students who are falsely identified as in need of services will not be cost effective, and those who are falsely identified as sufficiently able when they are not may require more costly remediation at a later point in time (Vellutino and Scanlon 2001). Technical quality of a test in terms of validity and reliability and the integrity of young learner language assessment systems as a whole are of course major considerations when the stakes for testing young students are high (McKay 2005; Bailey and Carroll 2015).

Assessment for Learning

Assessment for instructional or diagnostic purposes can take the form of standardized summative assessments or classroom-based formative assessments. Standardized assessments will offer the language teacher information about a sample of items across a variety of domains to measure general language proficiency or within a single domain of language such as vocabulary or syntax and how well a student is

doing on these skills relative to either standards (i.e., criterion referenced) (e.g., the TOEFL Primary developed by Educational Testing Service for children aged 8 and older is mapped to the Common European Framework of Reference (CEFR, Council of Europe 2001)) or relative to other students his or her age, grade, or level of overall language proficiency (i.e., norm referenced) (e.g., the *pre*LAS developed by CTB McGraw-Hill for assessing 4–6-year-olds in both English and Spanish as either an L1 or L2). The information gained can be used to monitor annual progress or to categorize students for educational purposes, for example, to literacy instruction in a child's dominant language. However, the information from such standardized assessments is likely to be neither sufficiently refined nor contain a critical number of like items to effectively target specific subskills. Educators must guard against using information from tests designed for one purpose (e.g., annual growth in general language proficiency) with another purpose in mind (e.g., next-steps instructional decisions) (National Educational Goals Panel [NEGP] 1998). Alternative or formative assessment is, however, designed to closely guide student learning as Wiliam (2006) explains:

> What makes an assessment formative, therefore, is not the length of the feedback loop, nor where it takes place, nor who carries it out, nor even who responds. The crucial feature is that evidence is evoked, interpreted in terms of learning needs, and used to make adjustments to better meet those learning needs. (p. 285)

Assessment for learning, such as formative assessment, is especially pertinent in the case of young learners still acquiring a new language. Formative approaches to assessment can capture a broad array of relevant language information for teachers that is closely tied to the young learners' instructional needs (Davidson and Lynch 2002; Frey and Fisher 2003). Formative assessment can be conducted by teachers either informally while "on the run" as part of ongoing instruction, or it can be formal, that is, planned in advance to address certain aspects of student language knowledge (e.g., McKay 2006). A central focus of formative assessment is teacher feedback to students, as well as a focus on student monitoring of their own language learning through self-assessment (Bailey and Heritage 2008).

Formative assessment may also include extra-child characteristics such as the classroom environment, parental involvement, home literacy habits, etc., and take many different forms (see Tsagari 2004 for a brief overview of the nomenclature and strengths and weakness of alternative assessments in the language assessment context). The use of informal observations, for example, allows for a range of skills (e.g., peer-to-peer oral discourse) not always amenable to more formal or standardized assessment environments. Observations can also be made formally and used to evaluate the quality of the language environment of a classroom rather than individual students (e.g., the Sheltered Instruction Observation Protocol, SIOP, Echevarria et al. 2004). The use of progress maps on a developmental continuum in order to estimate a student's growth over time (Masters and Forester 1996) and the use of portfolios to create individual profiles of language learning progress and achievement (e.g., Butler and Stevens 1997; CEFR, Council of Europe 2011;

Puckett and Black 2000) are alternative methods well suited to documenting the language of young learners and facilitating teachers' decision making for further learning. Such approaches can even be adopted by students themselves. For example, the "language passport" supported by the Council of Europe's European Language Portfolio initiative (2011) is used by students to directly rate their own language proficiency, although see Hasselgreen 2005 for a critique of the CEFR with younger language learners to which the European Language Portfolio is mapped.

Developmental Considerations

Motivation for this chapter comes primarily from the recognition that there are developmental and contextual factors that must be taken into account with the assessment of young language learners (e.g., Inbar et al. 2005; McKay 2006; Rea-Dickins and Rixon 1997). As in the USA, initiatives in Australia, Canada, and the UK have placed increasing emphasis on school systems to be held accountable for monitoring progress in the language development of young students, particularly young immigrant or language minority students (Indigenous and nonIndigenous) (e.g., McKay 2005, 2006; Silburn et al. 2011). There has also been an increase in young children studying English as a foreign language in non-English-speaking countries. Graddol (2006) reports that:

> The age at which children start learning English has been lowering across the world. English has moved from the traditional 'foreign languages' slot in lower secondary school to primary school – even pre-school. The trend has gathered momentum only very recently and the intention is often to create a bilingual population. (Graddol 2006, p. 88)

An interesting prediction stemming from this situation is that in the future there will be only "young" learners of English as older members of societies will have acquired English earlier in life. Consequently, it is appropriate that learners in this young age range receive emphasis in future assessment development and research efforts.

In a review of research on the assessment of school-age language learners conducted in various parts of the world, McKay concludes that young learner assessment deserves to be established as a highly expert field of endeavor requiring, for example, knowledge of the social and cognitive development of young learners, knowledge of second language literacy development, and understanding of assessment principles and practices (McKay 2005, p. 256). Beginning with McKay's assertion that the field develop an understanding of assessment principles and practices, three main areas of test design with young children require special consideration: (1) format (whether individual, small group, or whole class), (2) choice of item and task types, and (3) choice of contextualized, age-appropriate stimuli (Inbar et al. 2005). Explicitly identifying these three areas raises specific challenges for test development practices with young children. In each

case, design decisions must take the learning context into account to establish a match between instructional environment and assessment.

Test Format

The language modality and age of the test taker will certainly dictate the appropriate format in which to assess young learners. Individual assessment will be necessary for coverage of many of the skills in the speaking and listening modalities. However, in the preschool setting, many classroom teachers also call upon children to respond in unison (e.g., sing-alongs, calling out keywords as a chorus, and providing en masse actions/enactments to stories and poems, Tabors 2008). A child's ability to both comprehend and participate in such group activities should be at least one focus of assessment with the youngest language learners in this early instructional context.

Assessment of early literacy may need to be carried out in individual or in small group contexts because test takers cannot be relied upon to be sufficiently proficient to read directions for responding to print items or tasks nor to maintain their attention in group settings. No matter the format, limiting the duration of the test to avoid testing fatigue will be of far greater concern with this young population than with older test takers.

Test Item and Task Types

Choice of item and task types will need to correspond to the cognitive processing capabilities and degree of task familiarity of young learners. For example, Weir (2005) provides a language test item that requires making meaning from a bar chart. This type of task presents statistical information in a way that primary school children will encounter in graphics during mathematics, science, or social studies lessons. An item type that requires responding to a series of questions based on information extracted from a graphic would be appropriate once children have, as part of the school curriculum, received explicit instruction in "reading" graphic displays of information, otherwise the item type would be unfamiliar, and the task demands too great for the young learner; simply put, the demands of the assessment should match the demands of the curriculum. Other cognitive developments that will restrict the range of tasks include attention span and memory. For example, multistep items that require sequential manipulations of information or lengthy passages of text followed by comprehension questions may be outside the cognitive capacity of the youngest language learners.

To lessen the negative impact of processing demands and to capitalize on the degree of assistance learners may require from more expert others (Vygotsky 1978), assessments can award partial credit based on the verbal scaffolding necessary to elicit a response from young ESL test takers. This strategy allows for diagnostic information to be generated. The differing levels of response reveal how much

knowledge a child has and how much they still need to learn to succeed without assistance.

Task Content

The content of the tasks needs to be relevant to the young learner in terms of cognitive demands and cultural specificity (culture is addressed further in the later section). The younger the learner, the more contextualized the items will need to be in order for the test taker to make meaning of them. That is, items will need to be topically appropriate for the target age of the test taker, and the ability to answer the items should not require knowledge of information not already provided in the tasks or test items. Cognitive developments impacting these considerations include an awareness of testing procedures or the "test genre" (i.e., cooperation in attempting to answer all items and providing adequate constructed responses), as well as an understanding of an opportunity to use decontextualized language – that is, responses will be sufficiently explanatory for the absent test grader to make meaning of them. For the youngest learners, a number of easy "warm-up" items can be used to familiarize the child to the tester and testing procedures and should not be scored. Manipulatives (i.e., toy farm animals, dolls) can be incorporated in both item questions and response formats. According to research, young children are more successful on both production and comprehension tasks if the tasks use objects rather than pictures (e.g., Cocking and McHale 1981; Serna 1989 cited in Beck 1994); objects help contextualize the task in the cognitively less demanding "here and now."

Choice of age-appropriate content in test construction is made more complex with young learners than in other testing target groups because language development is concurrent with developments in other areas (e.g., scholastic, cognitive, and social developments). Because a child may begin learning a second or foreign language at any point in their early school years, the development of the language can be asynchronous with developments in other areas. The beginner status of young learners in the later elementary grades makes choosing content difficult (i.e., restrictions on availability of age-appropriate topics from which to select beginning level vocabulary and simple discourse contexts). This is also the situation if a test is to span an age range rather than be targeted at individual grades or ages.

Impact of Child Development on Test Interpretation

Cognitive and social developments not only impact test design but also the manner in which tests are administered and interpreted. Assumptions upon which validity arguments are made with standardized assessments (e.g., Davidson and Lynch 2002; Weir 2005) are often compromised when administering such tests with young children. For example, the assumption of uniformity of the testing experience for all the test takers is not met with young children whose attention abilities and

familiarity with test taking can vary tremendously (Powell and Sigel 1991). Moreover, what is considered "typical" for this young age range also varies. This raises the issue of whether using certain types of assessment with very young children is desirable. If the purpose of assessment is accountability of the program, then group level and classroom-related indicators (e.g., amount of student engagement, ESL experience of teaching staff) may be most appropriate. If the purpose is diagnostic, then information on individual students may be preferred. However, caution is required because of the compromises outlined earlier. Interpretations from formative assessment approaches rather than from administration of standardized assessments may be more meaningful.

Guidelines for Assessing Young Children

While not specifically targeting the assessment of language, there are several general test administration guidelines for use with young learners. For the very youngest learners in the USA, the *Principles and Recommendations for Early Childhood Assessments* assembled by the NEGP (1998) still hold. These include but are not limited to the following four guidelines: (1) assessments should be tailored to a specific purpose and should be reliable, valid, and fair for that purpose, (2) policies should be designed recognizing that reliability and validity of assessments increase with children's age, (3) assessments should be age appropriate in both content and the method of data collection, and (4) parents should be a valued source of assessment information, as well as an audience for assessment results. Specific guidelines for assessment practices with young English learners have also been published by the National Association of the Education of Young Children (NAEYC 2009).

As the basis for all assessment (clinical, scholastic, and linguistic) in K-12 education in the USA, the *Standards for Educational and Psychological Testing*, published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (2014), includes comprehensive guidelines for the testing of individuals of diverse linguistic backgrounds, and researchers working with language minority students have additionally made contributions to guide fair and valid assessment practices of both the language and other knowledge areas of language learners (see also Sireci and Faulkner-Bond for review 2015).

Suggested practices include use of test accommodations. Accommodations such as extra time and dictionaries are thought to vitiate the interpretation of test results obtained with ELL students as long as these have been empirically proven not to alter the language construct to be measured (e.g., reading comprehension). Increasingly technology can play a role in test accommodations such as the use of computer-administered bilingual dictionaries or glossaries (e.g., Abedi 2014). If the construct is not language ability itself but the test uses language as the medium (e.g., an assessment of mathematics requiring both reading and writing skills), then accommodation options include not only extra time and bilingual dictionaries/glossaries but also the option to test in a student's native language if this matches the language

of instruction (Pennock-Roman and Rivera 2011) (for a general discussion, see also Abedi, chapter "▶ Utilizing Accommodations in Assessment," Vol. 7). However, interpretation of accommodated results as valid indicators of academic content area knowledge has real consequences for students; we should still be cautious in our interpretations because as Davidson (1994) points out, norming studies of such academic achievement assessments have often not included learners who reflect the full range of language proficiencies found in schools.

Other impacts on administration and interpretation of language assessments include the training needs of teachers who often must administer assessments to school-age language learners. General education teachers have been found to have little training in language development and assessment (e.g., Téllez and Mosqueda 2015). Scoring and reporting the test performance of young learners also proffer challenges to teachers and test developers alike. Scoring concerns include the degree of teacher variance in what is considered an acceptable answer. For example, children's immature articulatory abilities or their productions influenced by L1 may make responses difficult to decipher and thus score reliably.

Multiple sources of evidence should be used to increase the validity of inferences about student language performances (e.g., Conti-Ramsden and Durkin 2012; Dockrell and Marshall 2015). Employing multiple measures helps prevent overreliance on any one assessment that may yield a biased view of performance due to cognitive or social development constraints. This does not however entail administering large batteries of standardized assessment that could quickly lead to test fatigue in young children. Rather, studies of expert teachers suggest that they use their knowledge of teaching and learning to create an ongoing cyclic process of teaching and assessment involving a repertoire of both formal and informal assessments (Rea-Dickins 2001). For example, evidence of language proficiency can come from combining a student's performance on formal assessments and informal quizzes and from teacher observation during class time (e.g., Frey and Fisher 2003), as well as utilizing self- and peer assessment.

Finally, reporting the results of a test performance to young children also needs to be carefully considered and made age appropriate to avoid issues of demotivation or threats to a child's self-esteem. However, reporting results to children and reporting results to teachers and parents need not be the same process, and teachers will need item level or subskill level information from assessments in order to make effective instructional modifications.

Impact of Culture on Administration and Interpretation

Culture impacts the fair and valid testing of young children's language abilities when there is a mismatch between home practices in communication and those practices commonly used for assessment. For example, Peña and Quinn (1997) report that Latina and African-American mothers typically do not label objects in their children's environment (as is the case for most vocabulary assessment), but rather engage in games that more often require descriptions. Thematic content of an

assessment also needs to be compatible with children's home culture (at least culturally appropriate for the majority of learners taking the test, if known). Alternatively, assessors have successfully administered dynamic assessments using a test-teach-test design with preschool children to reduce bias from lack of cultural familiarity with vocabulary (Peña et al. 2001). In addition, many children come from backgrounds where they might be expected to learn from observation rather than overt participation and to demonstrate their comprehension nonverbally (e.g., Beck 1994; Scollon and Scollon 1981). Collectively, this research should impact test development design, encouraging more development of dynamic assessments, or in the case of listening comprehension, creation of items that do not rely exclusively on verbal responses to signal accurate comprehension. Furthermore, early childhood education agencies, such as the Head Start in the USA, recommend caution with the interpretation of language assessments with young children noting the need to assess and take into account all the languages a young learner knows during educational decision making (Office of Head Start 2010) (for discussion, see Mahon et al. 2003).

Work in Progress and Future Directions

The twenty-first century began with a new era of educational accountability impacting young language learners in terms of the language demands now placed on them in schooling contexts. The field has shifted from recognizing that assessing young learners entails the assessment of their language development and the assessment of their academic content learning through language to also recognizing the need to take account of the language practices of the academic content areas on those very language assessments themselves. This has led to recent comprehensive language test development efforts in many parts of the world. In the USA, "next-generation" ELD assessment is under way under federal government initiatives to align ELD assessment with the academic content standards. Add to this mandate the anticipated expansion of publicly funded education to young, preschool-age children, many of whom are the children of immigrants from non-English-speaking countries. In Australia, add to this the increased focus on the language learning needs, indeed rights, of Indigenous students (Silburn et al 2011). In Europe, add in the expansion of the European Union with the increased mobility this brings, as well as, most recently, asylum seekers from the Middle East and North Africa, and collectively large numbers of families with young children are settling in areas of Europe where they do not speak the dominant language. Much research and test development has still to be done to improve assessment of the wide range in language demands now facing young learners.

Recommendations made in the 2008 version of this chapter of the encyclopedia to meet the language assessment needs of young language learners were organized around three aspects of the mission statement of the Committee on Early Childhood Pedagogy (National Research Council 2001). Reviewing those aspects now (technical quality of assessments, teacher professional development around language assessment, and integration of technology) reveals to what extent advancements

have been made nearly a decade on and which areas still need the attention of researchers and educators.

In terms of advancing the technical quality of assessments and how they are used to support the learning of young learners, much has been achieved in articulating construct definitions of necessary language knowledge and skills. The call for the revision or creation of ELD standards for the preschool through school-age levels has not only been met by several individual US states and consortia, for example, but CCSSO created the *Framework for English Language Proficiency Development Standards* (CCSSO 2012) to guide such revisions based on a synthesis of research around language and content learning (e.g., Lee et al. 2013) that has led to the identification of key *language practices or performances* found to be common across the new language arts and mathematics and science standards (CCSSO 2010; NGSS 2013). These practices and performances are a "combination of communicative acts (e.g., saying, writing, doing, and being) used in the transmission of ideas, concepts, and information in a socially mediated context" (p. 2) that include, among others, for language arts the support of "analyses of a range of grade level complex texts with evidence," for mathematics "construct viable arguments and critique the reasoning of others," and for science the necessary language to "plan and carry out investigations" and "engage in argument from evidence." Continued research at the intersection of content knowledge and language will no doubt help to refine the construct for future assessment development with this age range. Uccelli and colleagues (2014) are focusing on language that is common across various disciplines at the upper elementary level and how best to assess this construct, whereas others are looking at the intersection of content knowledge and ELD in the preschool context (e.g., the Literacy and Academic Success for English Learners through Science, or LASErS program of the Education Development Center) which could aid us in understanding how content knowledge itself shapes language use.

Accommodations research has also continued apace with, as mentioned, new meta-analyses providing details about the efficacy of accommodation use under different conditions (e.g., Pennock-Roman and Rivera 2011). This nuanced information has informed new principled accommodation guidelines or algorithms for use with school-age students still acquiring the language in which their academic content knowledge will be assessed (Abedi and Ewers 2013; for a general discussion, see also Abedi, chapter "▶ Utilizing Accommodations in Assessment," Vol. 7).

In other areas dealing with the technical quality of language assessments with young learners, work is still in progress. Despite our calling for developmental trajectories for language acquisition, the field still knows little about the progression of language in young school-age language learners (Hoff 2013). The characterization of language development is paramount in the creation of effective language proficiency assessments. Work under way by the Dynamic Language Learning Progressions (DLLP) project (Bailey and Heritage 2014) addresses the lack of empirically derived trajectories of language development by sampling oral and

written language practices as outlined by the CCSSO *framework* (e.g., explanations of mathematics task procedures) with students aged 5–12 who have varying proficiencies of ELD. This kind of evidence-based approach to creating language progressions needs to be extended to additional language practices (e.g., argumentation), a wider range of academic content areas (e.g., science, history), and of course to students across all grades.

Language learning progressions hold promise not only for informing development of standardized assessments but also for the area of formative assessment. While traditional notions of validity and reliability cannot be easily applied to establishing the technical quality of formative assessment approaches, criteria for establishing the effectiveness of formative assessment in the classroom can be created, discussed, tried out, and refined. Indeed, recently Heritage (2013) has highlighted the immediacy or proximate timing of evidence of student learning as a key facet of what makes formative assessment valid, along with the need for formative approaches to assessment to yield insights into students' current learning that are sufficiently tractable to be useful in instruction. Also within the area of formative assessment, there is accumulating evidence from a program of research on self-assessment that this population of language learners is not too young to benefit from the self-reflection entailed by self-assessment practices. Butler and Lee (2010), for example, found that 11–12 -year-old students in an EFL context were able to improve their English performances and increase their confidence in learning English with regular use of self-assessment in a classroom context. And in-progress work by Pitsoulakis and Bailey (2016) is revealing that children as young as 7 years are able to self-assess with the appropriate scaffolds to notice features of their own language productions.

Research on teacher professional development remains a key area in language assessment with young learners. Many of the guidance documents cited thus far have educators in mind for special caveats to the administration and interpretation of standardized assessments with young children (e.g., NAEYC 2009). Developing teacher capacity around practices that generate evidence of student learning and lead to accurate interpretations remain important topics for language assessment research specifically (Téllez and Mosqueda 2015; Michel and Kuiken 2014) and for formative assessment research more broadly (Wiliam 2006; Heritage 2013). While some early research has shown expert teachers to effectively use assessment for learning (e.g., Rea-Dickins 2001), more research is needed in the area of professional development to answer the question: *How do teachers effectively implement and use a wide repertoire of assessments for a variety of summative and formative purposes?*

Within the past decade, technology has changed the landscape of language assessment, and this is as true for the youngest learners we have considered here as it is with the assessment of older children and adults. The move to computer-based assessment has been made by the standardized assessments already mentioned (e.g., TOEFL Primary Speaking section), as well as by other newly released assessments

such as the Test of English Language Learning (TELL) progress monitoring application from Pearson and the revised ACCESS for ELLs, the annual summative assessment used by the WIDA ELP assessment consortium in more than 35 states.

Electronic devices are also readily available for continuous digital documentation of student progress (Pellerin 2012) with the possibility for even very young children to use the same tablet devices deftly for assessment purposes. Technology is especially suited to the assessment of young children (see also Chapelle and Voss, chapter "▶ Utilizing Technology in Language Assessment," Vol. 7); the graphic capabilities that technology offers can also provide a child-friendly context for assessment, with testing made enjoyable for young test takers by mimicking familiar games or cartoons.

Technology has solved a key issue in formative assessment – that of data capture, storage, and management. Data management systems can help make formative assessment practices more effective by systematizing the information that teachers may record formally or "on the run." Language corpora can now also be accessed to provide audiovisual and transcript data that provide teachers who lack familiarity with students from diverse language backgrounds with ways to more accurately compare and evaluate their young language learners. The DLLP project in progress has this as an explicit goal of the project (Bailey and Heritage 2014; Bailey et al. 2016). Authentic language use found in linguistic corpora can also be used to guide test item writers in the production of stimuli texts and test questions However, in standardized assessment development contexts with young learners as it has in test development with adult learners (Frantz et al. 2014).

We can assuredly claim that the assessment of young language learners has made large strides toward evolving into what McKay called "a highly expert field of endeavor." The field has the attention of many national governments due to the increased accountability placed on the role of language in the educational outcomes of all young learners but especially those speaking languages other than English or their society's dominant language. This situation has posed challenges on how best to design assessments that are fair and valid with young children, illuminated gaps in our understanding of the intersection of language and academic content learning, required that we continue to learn how to build the capacity of teachers to both summatively and formatively assess their students' language learning, and led us to continue to leverage technology to meet these myriad objectives. the second decade of the twenty-first century is first and foremost an exciting time to be working with young language learners and their dynamic assessment needs.

Cross-References

► Cognitive Aspects of Language Assessment
► Critical Language Testing
► Washback, Impact, and Consequences Revisited

Related Articles in the Encyclopedia of Language and Education

Sheena Gardner, Aizan Yaacob: Role Play and Dialogue in Early Childhood Education. In Volume: Discourse and Education

Christine Hélot: Awareness Raising and Multilingualism in Primary Education. In Volume: Language Awareness and Multilingualism

Kathryn M Howard: Language Socialization and Language Shift Among School-Aged Children. In Volume: Language Socialization

Amy Kyratzis, Marjorie Goodwin: Language Socialization and Social Practices in Children's Peer Interactions. In Volume: Language Socialization

---

References

Abedi, J. (2014). The use of computer technology in designing appropriate test accommodations for English language learners. *Applied Measurement in Education, 27*(4), 261–272.

Abedi, J., & Ewers, N. (2013). *Accommodations for English language learners and students with disabilities: A research-based decision algorithm.* Smarter Balanced Assessment Consortium. Available at http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-represented-students.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological measurement*. Washington, DC: American Educational Research Association/American Psychological Association/National Council on Measurement in Education.

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), 480–506.

Bailey, A. L., & Carroll, P. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education, 39*, 253–294.

Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy, Grades K-6: Building reading and academic language skills across the curriculum*. Thousand Oaks: Corwin/Sage Press.

Bailey, A. L., & Osipova, A. V. (2016). *Children's multilingual development and education: Fostering linguistic resources in home and school contexts*. Cambridge, UK: Cambridge University Press.

Bailey, A. L., Heritage, M., & Butler, F. A. (2014). Developmental considerations and curricular contexts in the assessment of young language learners. In A. J. Kunnan (Ed.), *The companion to language assessment*. Hoboken: Wiley-Blackwell.

Bailey, A. L., Blackstock-Bernstein, A., Ryan, E., & Pitsoulakis, D. (2016). Data mining with natural language processing and corpus linguistics: Unlocking access to school-children's language in diverse contexts to improve instructional and assessment practices. In S. El Atia, O. Zaiane, & D. Ipperciel (Eds.), *Data mining and learning analytics in educational research*. Malden: Wiley-Blackwell.

Barrueco, S., Lopez, M., Ong, C., & Lozano, P. (2012). *Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures*. Baltimore: Brookes Publishing Company.

Beck. (1994). *Development of the Alchini Bizaad comprehension test of Navajo and English for young children.* Unpublished manuscript, University of Northern Arizona.

Bialystok, E., McBride-Chung, C., & Luk, G. (2005). Bilingualism, language proficiency and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580–590.

Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5–31.

Butler, F. A., & Stevens, R. (1997). Oral language assessment in the classroom. *Theory Into Practice, 36*(4), 214–219.

Cazden, C. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Portsmouth: Heinemann.

Chamot, A. U. (2005). The cognitive academic language learning approach (CALLA): An update. In P. Richard-Amato & M. A. Snow (Eds.), *Academic success for English language learners*. White Plains: Longman.

Cocking, R. R., & McHale, S. (1981). A comparative study of the use of pictures and objects in assessing children's receptive and productive language. *Journal of Child Language, 8*, 1–13.

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review, 22*(4), 384–401.

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.* Washington, DC: CCSSO.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Council of Europe. (2011). *European Language Portfolio.* Available at http://www.coe.int/t/dg4/education/elp/Default_en.asp

Davidson, F. (1994). Norms appropriacy of achievement tests: Spanish-speaking children and English children's norms'. *Language Testing, 11*(1), 83–95.

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven: Yale Press.

De Houwer, A. (1998). By way of introduction: Methods in studies of bilingual first language acquisition. *International Journal of Bilingualism, 2*(3), 249–263.

Department for Education. (2016). *The national curriculum.* Available at https://www.gov.uk/national-curriculum

Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health, 20*(2), 116–125.

Echevarria, J., Vogt, M., & Short, D. (2004). *Making content comprehensible for English Language Learners: The SIOP model*. Needham Heights: Allyn & Bacon.

Faltis, C. (2015). Language advocacy in teacher education and schooling. In M. Bigelow & J. Ennser-Kananen (Eds.), *The handbook of educational linguistics* (pp. 65–77). New York: Routledge.

Frantz, R. S., Bailey, A. L., Starr, L., & Perea, L. (2014). Measuring academic language proficiency in school-age English language proficiency assessments under new college and career readiness standards in the U.S. *Language Assessment Quarterly, 11*(4), 432–457.

Frey, N., & Fisher, D. (2003). Linking assessment with instruction in a multilingual elementary school. In C. A. Coombe & N. J. Hubley (Eds.), *Assessment practices*. Alexandria: TESOL, Inc.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Oxford, UK: Wiley-Blackwell.

Gottardo, A., Chiappe, P., Yan, B., Siegel, L., & Gu, Y. (2006). Relationships between first and second language phonological processing skills and reading in Chinese-English speakers living in English-speaking contexts. *Educational Psychology, 26*(3), 367–393.

Graddol, D. (2006). *English next: Why global English may mean the end of English as a foreign language*. London: The British Council.

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing, 22*, 337–354.

Heritage, M. (2013). Gathering evidence of student understanding. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 179–196). Thousand Oaks: Sage.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*(1), 4–14.

Inbar, O., Shohamy, E., & Gordon, C. (2005). Considerations involved in the language assessment of young learners. *ILTA Online Newsletter, 2*, 3.

Jang, E. E. (2014). *Focus on assessment*. Oxford, UK: Oxford University Press.

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to Next Generation Science Standards and with implications for Common Core State Standards for English language arts and mathematics. *Educational Researcher, 42*(4), 223–233.

Mahon, M., Crutchley, A., & Quinn, T. (2003). New directions in the assessment of bilingual children. *Child Language Teaching and Therapy, 19*, 237–243.

Masters, G., & Forester, M. (1996). *Developmental assessment*. Melbourne: Australian Council for educational Research Ltd.

McKay, P. (2005). Research into the assessment of school-age language learners. *Annual Review of Applied Linguistics, 25*, 243–263.

McKay, P. (2006). *Assessing young language learners*. Cambridge, UK: Cambridge University Press.

Michel, M. C., & Kuiken, F. (2014). Language at preschool in Europe: Early years professionals in the spotlight. *European Journal of Applied Linguistics, 2*(1), 1–26.

National Association of the Education of Young Children (NAEYC). (2009). *Where we stand on assessing Young English-Language Learners*. Washington, DC: National Association of the Education of Young Children (NAEYC). Available at http://www.naeyc.org/files/naeyc/file/positions/WWSEnglishLanguageLearnersWeb.pdf.

National Educational Goals Panel (NEGP). (1998). In L. Shepard, S. L. Kagan, & E. Wurtz (Eds.), *Principles and recommendations for early childhood assessments*. Washington, DC: National Educational Goals Panel (NEGP).

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.

National Research Council. (2001). In B. Bowman, M. S. Donovan, & M. S. Burns (Eds.), *Eager to learn: Educating our preschoolers*. Washington, DC: Committee on Early Childhood Pedagogy/National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Nikolov, M. (2016). *Assessing young learners of English: Global and local perspectives*. Berlin: Springer.

Office of Head Start. (2010). *The head start child development and learning framework: Promoting positive outcomes in early childhood programs serving children 3–5 years old*. Arlington: Administration for Children and Families, U.S. of Health and Human Services.

Pellerin, M. (2012). Digital documentation: Using digital technologies to promote language assessment for the 21st century. *OLBI Working Papers*, *4*, 19–36.

Peña, E. D., & Quinn, R. (1997). Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools, 28*(4), 323–332.

Peña, E. D., Iglesia, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech- Language Pathology, 10*, 138–154.

Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and Non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10–28.

Pitsoulakis, D., & Bailey, A. L. (2016). *Extending learning progressions to self-assessment: Students finding "best fit" for next-steps learning and instruction.* Paper presented at the annual

meeting of the American Educational Research Association, Annual Conference, Washington, DC. Available at www.dllp.org

Powell, D. R., & Sigel, I. E. (1991). Searches for validity in evaluating young children and early childhood programs. In B. Spodek & O. N. Saracho (Eds.), *Issues in early childhood education* (Yearbook in Early Childhood Education, Vol. 2, pp. 190–212). New York: Teachers College Press.

Puckett, M. B., & Black, J. K. (2000). *Authentic assessment of the young child*. Upper Saddle River: Prentice Hall.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*, 429–462.

Rea-Dickins, P., & Rixon, S. (1997). The assessment of young learners of English as a foreign language. In C. Clapham & D. Corson (Eds.), *The encyclopedia of language and education* (Language Testing and Assessment, Vol. 7, pp. 151–161). Dordrecht: Kluwer.

Scollon, R., & Scollon, S. (1981). *Narrative, literacy, and face in interethnic communication*. Norwood: Ablex.

Silburn, S. R., Nutton, G., McKenzie, J., & Landrigan, M. (2011). *Early years English language acquisition and instructional approaches for Aboriginal students with home languages other than English: A systematic review of the Australian and international literature*. Darwin: The Centre for Child Development and Education, Menzies School of Health Research.

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education, 39*(1), 215–252.

Tabors, P. (2008). *One child, two languages: A guide to preschool educators of children learning English as a second language* (2nd ed.). Baltimore: Brookes Publishing.

Téllez, K., & Mosqueda, E. (2015). Developing teachers' knowledge and skills at the intersection of English language learners and language assessment. *Review of Research in Education, 39*(1), 87–121.

Tsagari, C. (2004). Alternative assessment: Some considerations. *Language Testing Update, 36*, 116–125.

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sanchez, E. (2014). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics, 36*, 1–33.

Vellutino, F. R., & Scanlon, D. M. (2001). Emergent literacy skills, early instruction, and individual differences as determinants of difficulties in learning to read: The case for early intervention. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 295–321). New York: Guilford Press.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach* (pp. 143–161). Basingstoke: Palgrave Macmillan.

Wiliam, D. (2006). Commentary: Formative assessment: Getting the focus right. *Educational Assessment, 11*(3–4), 283–289.