



Assessment of Young English Learners in Instructional Settings

27

Yuko Goto Butler

Contents

Introduction	478
Ecological Layers Characterizing Young Learners	478
YLS' Changing Learning Environments	478
YLS' Age-Related Characteristics	480
Key Issues and Challenges for Assessing Young Learners	482
Targeted Language Abilities for Assessment	482
Age-Appropriate Assessment Formats and Procedures	484
Learner-Centered Approach: Assessment to Promote YLS' Learning	485
Challenges to Meet YLS' Diverse Needs	487
Future Directions	488
Child Second Language Acquisition (Child SLA) Research	489
Teacher's Role in Assessment to Assist YLS' Learning	490
Technology and Assessment	491
Conclusion	492
Cross-References	492
References	493

Abstract

With the growing number of children worldwide learning English in instructional settings, educators are increasingly concerned with how best to assess their language development. However, assessing young learners (YLS, defined in this chapter as children age 5–12) warrants a number of special considerations due to their age and other unique characteristics. After discussing young language learners' unique characteristics and needs with respect to assessment, this chapter addresses the following major issues that arise when assessing them: (a) targeted language abilities for YLS, (b) age-appropriate formats and procedures of assessment, (c) assessment

Y. G. Butler (✉)

Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA

e-mail: ybutler@gse.upenn.edu

for learning (e.g., feedback and diagnostic information in assessment and assessment for autonomy enhancement), and (d) challenges to meet YLs' diverse needs (e.g., students with multilingual backgrounds and students with disabilities). The chapter concludes with suggestions for future research directions, including (a) more child second language acquisition research to inform assessment theory and practice, (b) addressing teachers' role in assessment to better assist YLs' learning, and (c) greater attention to the use of technology for YLs' assessment.

Keywords

Young learners · Age · Assessment for learning · Feedback · Autonomy · Technology · Accountability · Standards

Introduction

English is considered to be a powerful global language, and the number of children who are learning English in instructional settings is on the rise worldwide. Accordingly, educators are concerned about how best to assess these children's language development. Assessing young learners (YLs) warrants a number of special considerations because of their unique age-related and environmental characteristics. YLs' needs are increasingly diversified, and their language use is changing as a result of technology development. Assessment for YLs has to reflect such changes and needs. Assessment can be used for a wide variety of purposes, but any assessment information should ultimately enhance a child's learning. Importantly, teachers play a critical role in assessing YLs.

Assessment for YLs had been a relatively unexplored area of research. In the last couple of decades, however, there has been a rapid increase in the number of empirical studies on language assessment of YLs, including recent books on the topic (e.g., Nikolov 2016a; Prošić-Santovac and Rixon [forthcoming](#); Saville and Weir 2018; Tsagari and Spanoudis 2013; Wolf and Butler 2017). Drawing on such research, the present chapter discusses key issues when assessing YLs, defined as children age 5–12 learning English in various instructional settings (as an additional and/or foreign language). After describing social and cultural environments where such learning is taking place, as well as YL's age-related characteristics, the chapter addresses the following key issues when assessing YLs: (a) targeted language abilities for assessment, (b) age-appropriate assessment formats and procedures, (c) assessment designed to directly assist YLs' learning, and (d) challenges to meet YLs' diverse needs. The chapter concludes with suggestions for future directions for research.

Ecological Layers Characterizing Young Learners

YLs' Changing Learning Environments

YLs' language learning does not simply reside in individual children but is embedded in contexts. When attempting to understand children's development, Bronfenbrenner's (1979) model of ecology of human development reminds us of

the importance of considering whole ecological systems where the development is taking place. According to this model, a child's development, including language development, is a complex interplay between the child and multilayered social and cultural environments where learning is situated. Some of these environments are more immediate, such as family and schools, and others are more remote macro-systems, such as the given culture's belief system and bodies of knowledge. The model also incorporates a *chronosystem*, which allows for the consideration of changes to a child as well as his/her environments over time, both during the individual life span and across history. As Bronfenbrenner's model makes clear, in order to understand a YL's language development and capture its processes and attainments through assessment, we need to pay close attention to the child's entire ecological system and how the different elements of that system interact with the child.

Among many environmental complexities associated with YLs' English development, the following two macrolevel general trends are worth mentioning. First, YLs' learning environments are increasingly diversified. Traditionally, the field of second language acquisition (SLA) has characterized English learning contexts as falling into two general categories, namely, English-as-a-second language (ESL) (English as an additional language (EAL) is also often used in some regions, including the United Kingdom) and English-as-a-foreign language (EFL) contexts. ESL learning takes place in an environment where English is the dominant language, whereas EFL learning occurs in places where English is not dominantly spoken outside of the classroom (i.e., input-limited conditions). The underlying assumption for these classifications is that ESL and EFL learners have different types and amount of exposure to English and distinctively different goals and needs. Though this classification still has some usefulness, it might oversimplify what's really going on, at least in a certain regions and contexts; YLs' learning environments are increasingly varied within as well as across geohistorical boundaries. (A similar argument can be made for any other geohistorical classification models, including the well-known three-circle model of English by Kachru (1992).) On the one hand, some immigrant children in ESL contexts may be locked in a so-called language ghetto with limited exposure to English. On the other hand, a growing number of children in EFL contexts can increase their exposure to the target language through attending various types of language immersion programs or content and language integrated learning (CLIL) programs (Anderson et al. 2015). Additionally, many middle-class children participate in early study abroad programs (e.g., Song 2011); these children move between ESL and EFL contexts. Moreover, recent expansion of English-medium information via media and other means allows some learners to have greater access to English, both within and outside of formal school contexts, and the opportunities to access to English often vary according to YLs' socioeconomic backgrounds even within a given region.

Second, the increased use of various types of digital technology, such as computers, cellphones, and tablets, has drastically changed the way people communicate, which in turn greatly influences the way they learn and teach language. Technology has become a tool of instruction as well as a target of communicative language use. It is natural, then, to use technology as a means of assessment and to

consider the technology skills involved in language use as part of the target language ability (namely, potential *construct*) in assessment. Communication through technology, however, is increasingly obscuring boundaries of traditional categorizations, such as spoken and written languages and print (a traditional notion of literacy) and nonprint. As more children grow up with digital technology in their lives, their mental processing and strategies, such as memory and strategies to generate relevant information through networks, may differ from previous generations. Although there has been limited empirical investigation, especially among children, to verify this claim, studies targeting primarily adolescents and young adults do suggest that digital technology's potential effects on cognition "are likely nuanced, but could strengthen specific cognitive strategies" (Mills 2014: 385). If such effects do exist, instructions and assessments for YLs should be matched well with their preferred cognitive styles.

YLs' Age-Related Characteristics

Young learners are undergoing cognitive, socio-affective, and linguistic development. Their information-processing speed, memory span, and attention span show nonlinear and rather drastic improvement patterns until they reach their mid-teens. Preschool children's memory span is, on average, one third of that of adults. Memory capacity among younger children (children up to around age 6) differs not only quantitatively (i.e., memory span) but also qualitatively, and they use different memory strategies than older children and adults (Fry and Hale 2000). Children acquire various types of knowledge as they accumulate their daily experiences, and they gradually develop analytical abilities, metacognitive/metalinguistic abilities, and self-evaluative abilities throughout the school-age years. From birth, children also develop socio-cognitive interactive abilities, such as having joint attention (i.e., two individuals share their attention to an object through eye gazing, pointing, and other verbal and nonverbal means) and understanding other people's intentions. By the time children are 5–7 years old, they can collaborate with others in tasks and take turns in communicative exchanges. They are curious about objects and events around them, and they are also fond of fantasies and stories. Children around 9–10 years old begin to read for information, and by age 11 and 12, they are gradually able to sustain abstract topics in conversation (Clark 2009). It is important to note, however, that such general developmental patterns can vary significantly depending on children's sociocultural and physical environments. As discussed in detail below, all these developmental factors, in conjunction with contextual factors, greatly influence the design and administration of assessments, including the assessment content (e.g., age-appropriate tasks or item choices), formats (e.g., multiple-choice vs. performance-based assessment), procedures (e.g., individual-based vs. pair or group assessment, time requirements), the degree of autonomy (e.g., self-assessment and peer-assessment), and how best to offer scaffolding as part of assessment processes.

Another developmental factor that uniquely characterizes YLs is that they are developing their first language (L1). While we still have limited understanding of the bidirectional, cross-linguistic influence on language development among YLs, it may be different from that of adults, who have already firmly established their L1 categorizations in phonology, syntax, semantics, and pragmatics. Moreover, due to the increased mobility of people, a greater number of YLs are learning English in multilingual environments (both in the so-called ESL and EFL contexts) (Bailey and Osipova 2016). For example, a Kurd child in a primary school in Germany may primarily speak Kurdish at home, Turkish in the community, and German at school and learn English as an additional language for global communication. She may exercise *translanguaging* in multiple languages and may do so across multiple social spaces (García and Li 2014). In any of her given language-use contexts, multiple languages are activated (Grosjean 2010). We know little about how language development trajectories of bilingual or multilingual YLs differ from those of their monolingual L1 learning peers. No matter how the two groups differ (and multilinguals may be different from bilinguals as well), such differences challenge the traditional approach to assessment in which monolingual native speakers' performance is set as the stable norm.

Importantly, there are substantial individual differences in cognitive, socio-cognitive, and linguistic developments among children within the same age groups. Moreover, children's individual development in different domains may not occur in tandem; for example, a child may be more advanced in a certain cognitive domain but less developed in a social domain. Teachers often observe substantial variability among students with respect to their experience with assessment and their L1 proficiency (or any other languages in the case of multilinguals), in the literacy domain in particular. The use of home and community language(s), personality, and learning styles are all found to be influential over YLs' target language development (Ćatibušić and Little 2014). Such diversity among children makes it difficult to standardize assessment content, criteria, and procedures.

Finally, we must not forget that children are vulnerable to adults' attitudes about assessment and their expectations concerning children's performance on assessment. Carless and Lam (2014) reported that lower primary school students in Hong Kong expressed both negative and positive feelings in drawings they made to describe their feelings about tests as well as their parents' reactions to their test performance. In the United States, the United Kingdom, and Australia, recent standard-based educational policies mandate that young English language learners be assessed in English and other academic content, often through a series of high-stakes standardized tests, for accountability purposes. As a result, YLs can face substantial pressure (e.g., McKay 2006; Menken et al. 2014). Assessment results can have long-lasting effects, both positive and negative, on YLs' affective factors such as motivation, anxiety, and self-confidence, which in turn can greatly influence their learning. Early experience of failure tends to make children attribute their performance to ability rather than effort and leads to lower expectations for their future performance (Perry and VandeKamp 2000).

Key Issues and Challenges for Assessing Young Learners

Targeted Language Abilities for Assessment

Communicative Language Ability (CLA)

Different instructional models have varying degree of emphasis on “a language-content continuum,” with some programs focusing more on developing age-relevant language proficiency for communication and others concerned more about content knowledge acquisition through English (Inbar-Lourie and Shohamy 2009: 84). YLs’ English programs essentially aim at developing *communicative language ability (CLA)* – the ability to construct meanings and converse successfully in various social and academic contexts – but the targeted use domains and required age-appropriate proficiencies vary depending on the programs.

Researchers have taken different approaches to conceptualizing and operationalizing CLA. Purpura (2016: 193) identified four major approaches: (a) trait-centered, (b) task-based, (c) interactionist, and (d) sociointeractional. Detailed discussions of each approach are beyond the scope of this chapter, but many models assume the compositionality of some traits or abilities of CLA. (Socio-interactional models are an exception to this trend. They consider CLA to be co-constructed among individuals by participating in moment-by-moment interactions to achieve a goal-oriented activity.) These various models differ with respect to how to conceptualize the role of context in theorizing CLA. While these theoretical models have been influential in language assessment, theory-driven approaches to YLs currently face unique challenges. For example, when applying models that assume the compositionality of CLA, we have limited information about the interrelationships among components, how each component develops in relation to others, and how such development among YLs might differ from adults. Similarly, little is known about how YLs’ different traits interact with contextual factors and strategy use; such knowledge is critical for fully utilizing interactional models.

Efforts also have been made to identify the knowledge and skills necessary for successful communication based on experts’ judgments. Experts have developed various curricula and assessment standards and frameworks, with specific knowledge and skills corresponding to a certain proficiency level. The *Common European Framework for References for Languages (CEFR)* (Council of Europe 2001) is one well-known example. The CEFR’s can-do statements, composed of six major proficiency levels (ranging from A1 [breakthrough] level to C2 [mastery] levels), indicate what learners are able to do with the target language at each proficiency level. The CEFR has been widely used among YL educators and assessment developers not only in Europe but also in other parts of the world. It is critically important to remember, however, that the CEFR descriptors were not originally meant for use with YLs and that they were written in a context-free fashion; namely, the descriptors are written in a general fashion so that they can be used in wide contexts (Hasselgreen 2005). Thus, major adaptation is indispensable to make the CEFR descriptors age- and content-appropriate when using them for YLs. There has been some effort to

create CEFR-based can-do descriptors for YLs, including creating a pre-A1 level, further-dividing A1 and A2 levels and changing wordings in the descriptors (e.g., Benigno and de Jong 2016; Hasselgreen 2005; North 2014). Little (2007) noted concern that modifying higher levels (C1 and C2) for YLs would be particularly challenging because they assume learners' cognitive maturity and academic and professional experiences that go beyond YLs who are in immersion and/or CLIL contexts. After all, we still don't know the extent to which children follow the same developmental path as the one outlined by the descriptors. Major international, large-scale, standardized, proficiency tests – including the Cambridge Young Learners of English tests, the Pearson Test of English Young Learners, and TOEFL Primary Test – also indicate alignment with the CEFR. But as Papageorgiou and Baron (2017) warned, test users “should not misinterpret any type of linking as a sufficient indicator of the overall quality of an assessment or as confirmation of the validity of its scores for their intended use” (148).

Language Ability for Academic Contexts

YLs in immersion contexts (e.g., immigrant children in ESL/EAL contexts, YLs in bilingual and CLIL contexts) need to acquire content knowledge such as math, science, and social science through the target language. Theorists, including Vygotsky (1962) and Bruner (1975), have addressed the significant role that language plays in developing concepts in academic domains. Cummins's (1979) classic distinction between basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP) has shed light on the importance of developing CALP in academic studies, which could take a long time for YLs to develop. Given increasingly demanding accountability requirements in many regions and institutions, concerns have been expressed about the validity of content subject assessments if a child has not yet developed the language skills necessary for acquiring the content knowledge. Bailey and Butler (2007) proposed developing a test of academic language for L2 learners that could serve as a prerequisite for taking content assessments.

Although researchers acknowledge the important role of language in academic studies, they disagree about how to conceptualize *academic language*. Some researchers consider academic language to be diverse sets of discourses and genres associated with academic disciplines (Johns 1997), while others argue that it involves “language used to navigate school setting more generally” (Bailey and Huang 2011: 343) in addition to lexicon, sentential structures, and discourse cohesions used to teach academic subjects. Indeed, Gu (2014), based on a large-scale proficiency test of YLs' English in immersion contexts, provided some empirical evidence indicating that academic and social languages are not distinct constructs. Scarcella (2003) included not only linguistic components but also cognitive and social factors associated with using language in academic contexts (e.g., values, attitudes, and motivation) as well as various learning strategies (e.g., higher order thinking and metalinguistic awareness). Others have questioned a static conceptualization of academic language, instead suggesting dynamic and evolving multiple literacies (e.g., Street 1996).

There are many challenges to identifying and mapping YLs' English development in academic content studies. For example, if we accept that academic language for YLs consists of discipline-specific elements as well as common core elements, as in Bailey and Butler (2007), then we need to identify academic language for each subject area (e.g., math and science) and grade level. In the United States, such efforts can be seen in English language proficiency standards developed or adopted by states. The English language development (ELD) standards developed by the World-Class Instructional Design and Assessment (WIDA) consortium (adopted by a number of states), in order to meet the requirements of No Child Left-Behind (NCLB) policy in 2001 initially and the K–12 Common Core State Standards (CCSS) in 2010, conceptualize academic English as being at the intersection of ELD standards and core standards in each subject area. WIDA thus linked ELD standards to content standards in five areas (language of socialization and instruction, language arts, math, science, and social studies) at each corresponding grade level (WIDA n.d.). Such linkages can provide teachers with blueprints for how YLs generally develop language in academic contexts, but as Bailey and Heritage (2014) pointed out, “[ELD standards] lack the specificity needed to describe the language learning and development that must occur for students to use language as both a goal in itself and in the service of content learning” (482). As discussed below, this can be partially due the field's insufficient empirical-based understanding of YLs' language development in academic contexts. Even if the goal is to make sure that YLs can develop sufficient academic language, however defined, before taking content-subject tests, as suggested by Bailey and Butler (2007) above, there is little information to rely on to determine such a level. Finally, poststructuralist researchers question the norm that serves as the foundation for current standard-based approaches; a particular idealized monolingual norm is used as the standard, and there is no room for fostering YLs' dynamic bilingual/multilingual development (Flores and Schissel 2014).

Age-Appropriate Assessment Formats and Procedures

In designing assessments for young learners, the tasks, formats, and procedures should be appropriate for their age and their life and classroom experiences. When using tasks for assessment, it is important to consider what cognitive demands the tasks entail. For example, the cognitive demands for “telling a story” based on pictures can be manipulated by increasing or decreasing the number of pictures used, showing or not showing children the pictures in the right order, using a story with a simpler or more complicated plot line, and adjusting the amount of planning time offered to children (Pinter 2015). Cognitively rich tasks can serve as a tool to elicit various linguistic, cognitive, and metacognitive resources that YLs obtain to complete tasks. A certain degree of cognitive challenge also can motivate YLs (Jang 2014). However, if the cognitive demands exceed children's capacity (or what Vygotsky (1978) called their *zone of proximal development*), the assessment will not only fail to give teachers accurate and meaningful information to assist the children's language learning, but also it can potentially dampen the children's motivation and confidence.

Educators need to understand that tasks that work well as classroom activities may not necessarily be effective assessment tasks. Once YLs realize that they are being assessed, they may behave differently from usual. It is also important to remember that children are sensitive to the pragmatic role of teachers or other assessors in the assessment process. Carpenter et al. (1995) found that, during a teacher-child pair task assessment, children ages 5–10 were puzzled during the assessment when their teacher asked them what they saw in a picture when they knew that the teacher could also see the picture. Indeed, children need to be socialized into the world of assessment in order to perform “appropriately” during the assessment (Butler and Zeng 2014), but understanding what they are expected to do during the assessment requires a certain level of social-cognitive maturity and experience.

A teacher-child oral interview format (a popular assessment format in primary school) certainly has some advantages in that it can allow teachers to tailor questions to individual learners’ proficiency levels and interactive styles and to stretch the learners’ abilities. Thus, this individual assessment format may work particularly well for younger children or children who are less proficient and less proactive. However, the individual format can easily fall into an initiation-response-evaluation (IRE) discourse pattern – a typical classroom discourse pattern – that may lead to relatively limited responses from the children, such as simply answering teachers’ questions (Butler and Zeng 2011). Conversely, child-child paired or group assessment formats can elicit a wider range of language use, such as asking questions, disagreeing, and suggesting (Butler and Zeng 2011), as well as a variety of interactional strategies, including repetitions and comprehension checks, all the while creating more balanced power relationships among the participants (Oliver 2002). Paired and group assessment formats are also better aligned to classroom activities. Depending on the nature of task contents and formats (e.g., pairing and grouping, familiarity of tasks), however, children up to around age 10 may find it hard to work collaboratively during the assessment (Carpenter et al. 1995; but also see García Mayo and Agirre (2016) who found a U-shape development pattern of group dynamics).

Learner-Centered Approach: Assessment to Promote YLs’ Learning

As noted above, YLs are in the midst of developing not only their languages but also various knowledge and skills in academic and nonacademic domains, and this development is nonlinear and dynamic. Thus, assessment should be primarily designed to assist the development of the targeted abilities while focusing on the processes of learning. And this should be achieved through flexible and multiple means and in an ongoing fashion. Measuring YLs’ abilities and performances at a single point in time makes sense only if such information is clearly used for further learning and enhancing targeted abilities. In other words, *assessment for learning* (Black and Wiliam 1998), or assessment primarily used for formative and diagnostic purposes, has a particular relevance for young learners. Similarly, assessments that are designed to foster children’s development of self-regulation during early childhood have been strongly promoted, given the fact that early

self-regulation predicts children's long-term success in various academic learning (e.g., McClland et al. 2014).

Based on her investigation of teachers' assessment practices for young English learners in England and Wales, Rea-Dickins (2001) suggested that "good 'assessment for learning' thus motivates learners to become engaged in the interaction through which they are enabled to develop skills of reflection (as a basis for self- and peer-monitoring), as well as providing them with an ability to reflect metacognitively on their own learning" (452–3). *Dynamic assessment (DA)* is one type of assessment that focuses on the role of scaffolding in assisting learners' learning during interaction. Based on Vygotsky's (1978) notion of the *zone of proximal development*, DA, by providing various supports individually to learners during interaction as an assessment procedure, aims to determine the extent to which scaffoldings are necessary for the given learner to improve his/her performance. In other words, DA intends to capture not only a learner's current ability to complete a task independently but also his/her emergent abilities (Poehner et al. 2017). Researchers have also sought to uncover YLs' cognitive processing and strategies for solving assessment tasks during *cognitive diagnostic assessment (CDA)*, an assessment approach designed to provide learners with feedback on their cognitive and metacognitive strengths and weaknesses for successfully completing the assessment tasks. In an intervention study of CDA in reading among young English learners in Canada, for example, Jang et al. (2017) identified a number of practical tips for effective mediation. Tips included (a) YLs' emotional responses to feedback indicated a sign of their cognitive and metacognitive abilities, (b) self-questioning promoted YLs' metacognitive control, (c) YLs who chose texts based on their interest were more responsive to the intervention, and so forth.

Self-assessment is increasingly viewed as a way to promote learners' self-reflection and autonomy. As such, it has become common practice to include self-assessment items in textbooks and other resource books for teachers. And yet self-assessment does not seem to have much of a presence in practice in YLs' classrooms (e.g., Becker 2015), which might be due, in part, to the fact that teachers and parents often perceive the primary function of assessment to be summative. Thus, the relative unpopularity of self-assessment in practice might reflect their concerns that self-assessment is too subjective and unreliable, especially for YLs. Indeed, some evidence indicates that children up to around age 9 or 10 are more likely less accurate in self-assessing their performance in L2/FL compared with older children (e.g., Butler and Lee 2006). Given the complex nature of the act of self-assessment, age differences in response may be due to cognitive and metacognitive developmental factors as well as various social and affective factors, including children's experiences with self-assessment and their personality (Butler 2018a, b). Therefore, it is important for teachers to (a) select age-appropriate activities for self-assessment items, (b) provide clear wording, (c) construct items in a contextualized fashion (e.g., asking "Can I sing the ABC song well?" after children sing the song in class instead of asking, "Can I sing English songs well?"), and (d) offer children sufficient experience with self-assessment. Children also need to clearly understand the purpose of doing self-assessment (Butler 2016).

Critically, however, the accuracy of children's responses may be less important if one focuses on the potential merits of enhancing YLs' self-regulation. From a formative perspective, self-assessment should be designed to help children understand the goal of the task, reflect on and monitor their process in relation to that goal, and foresee the next step. Self-assessment should highlight children's accomplishments and promote their confidence. The teacher's role during this process is significant. Research has indicated that self-assessment can lead to improvements of YLs' confidence and English-learning but that if teachers do not subscribe to the spirit of assessment for learning and do not see the value of self-assessment for children's learning, the effect of self-assessment on children's learning remains limited (Butler and Lee 2010). Combining peer-assessment with self-assessment may facilitate YLs' autonomy over their own learning (Hung et al. 2016).

Challenges to Meet YLs' Diverse Needs

Current assessment practices fall far short of meeting the diverse needs of YLs. Test reliability and validity can vary by test-taking group. In the United States and the United Kingdom, high-stakes standardized tests of academic subjects tend to have lower reliability and validity among YLs compared with monolingual English-speaking counterparts (Espinosa 2013). According to Espinosa, YLs should first be assessed in their dominant language. However, this can be challenging due to a lack of valid and reliable assessment for identifying YLs' dominant language. And even if one's dominant language is identified, with a few exceptions, compatible tests in academic domains are not available in other languages. Translating existing tests into the students' dominant language is not an easy solution. A translated version usually does not ensure a compatible level of validity and reliability with the original test. Moreover, translated versions are often normed on children who do not share similar characteristics with YLs in immersion contexts (e.g., monolingual speaker of the YLs' L1 or dominant language).

Under monolingual assessment contexts where YLs are required to take standardized tests in the target language, various types of test accommodation have been employed through modifying the test itself (e.g., using plain language without changing the content of the test) or the test procedures (e.g., providing extra time). According to Abedi et al. (2004), the accommodation should minimize the test takers' potential source of difficulty, but that source has to be measurement irrelevant. The authors found that the most common standardized test accommodation practice in the United States was not made based on empirical evidence. They also found that the effectiveness of accommodations is largely learner- and context-dependent, leading them to conclude that there is no "one-size-fits-all" approach to accommodation (1).

The accommodation approach, an inclusive approach to mainstream assessment practice, rests on the premise that as long as the source of difficulty for YLs is removed, the test should measure the same ability among both YLs and monolingual speakers of the target language. However, this premise itself is questionable if we

accept the view that bilinguals'/multilinguals' abilities are *qualitatively* different from monolinguals' abilities (e.g., Cook 1992); if there is a qualitative difference, then assessing YLs through tests normed on monolingual students would raise serious validity concerns. A number of social consequences and implications, or *washback*, as a result of score interpretations and use of standard-based high-stakes tests (e.g., influences on instruction, students' grade promotion, and school and teacher evaluation), have been reported (Menken et al. 2014). Such washback effects are in fact considered to be an important part of test validity (Messick 1996). To respond to this problem, some researchers proposed multilingual assessments that allow YLs to use their multilingual resources by engaging in *translanguaging* during the assessment so that the assessment result can better represent their true understanding. This practice should reflect YLs' actual language practice more accurately as well (Menken and Shohamy 2015). At this point, however, little practical information is available for teachers due the scarcity of empirical research on the effectiveness and feasibility of this proposal.

Finally, diagnosing specific learning difficulties (SLDs) among YLs poses a serious challenge. When YLs do not meet academic standards in English-medium school contexts, they are often misidentified as having SLDs (over-representation) or judged as lacking sufficient English proficiency when in fact SLDs exist (under-representation) (Ballantyne 2013). Both over- and underrepresentations of SLDs invite serious consequences. Assuming that SLDs appear across languages, it is suggested that SLDs should be identified through diagnostic tests in the child's dominant language; however, such diagnostic tests are often not available in children's dominant languages. Moreover, even if the diagnosis is possible in the child's dominant language, "the lack of an official diagnostics of SLDs does not exclude the possibility of having L2 learning difficulties" (Kormos 2017: 36). Learning difficulties in L2 can be due to multiple factors, not only cognitive and metacognitive factors (e.g., working memory, naming speed, and attention control) but also social and affective factors (e.g., instructional contexts and motivation). After all, we still have limited knowledge about how L1 and L2 learning difficulties overlap. Complicating matters, SLDs encompass various types of difficulties. Any given sources of difficulties may also influence L1 and L2 differently depending on modalities (spoken and written language modes), types of language processing (implicit and explicit processing), stages of development, and the combination of L1 and L2 languages (Kormos 2017).

Future Directions

Assessment for YLs is still relatively understudied, and there are many agendas for future research. I focus on three critical areas in this section: (1) child second language acquisition research, (2) teachers' role in assessment to promote learning, and (3) technology and assessment.

Child Second Language Acquisition (Child SLA) Research

Researchers assume that child SLA is different from adult SLA; however, research on child SLA is still limited, and we do not know how, exactly, child SLA differs from or is similar to cases of adult SLA (Oliver and Azkarai 2017). First, we need to better understand how YLs develop communicative language abilities in the target language (in relation to other language(s) that they speak). Recent longitudinal research efforts among immigrant children in the United States (Bailey and Heritage 2014) and Ireland (Ćatibušić and Little 2014) are promising. As we have discussed in the previous section, considering that the monolingual approach to assessment in multilingual contexts is more likely to have serious validity threats, we need to better understand how bilingual/multilingual children *uniquely* develop their English as well as other language(s) in their own right. Since learning trajectories and speed of development may be influenced by instructional environments as well as the characteristics of L1 and the target language, we need more information from diverse learning contexts as well as learners with various linguistic backgrounds.

Second, we need more information about the relationship between the quality of input (including feedback) and YLs' target language development. Such information is particularly important in language-focused instructional contexts (or what are traditionally referred to as EFL contexts). This is because, in those input-poor contexts, it has been found that the frequency and quality of instruction, rather than the age of onset of learning, is more influential over YLs' target language development, contrary to a widely held assumption that "the younger the better" for language learning (Muñoz 2017). It would be valuable to develop corpora capturing classroom interactions (teacher–student and student–student interactions), just like Child Language Data Exchange System (CHILDES) – a corpus that has been used widely among first-language acquisition researchers.

Third, more research on individual differences in child SLA is necessary. In particular, information concerning individual differences in YLs' cognitive processing and strategies when they use language(s) would be valuable because it can inform teachers when designing assessment tasks, developing scaffolding techniques, or offering diagnostic feedback during/after assessments. We have limited research on *cognitive validity* (Field 2011) among YLs, which is research examining cognitive demands for completing tasks (in academic contexts in particular) and comparing cognitive processes between the assessment and real-life contexts. Such information is critical for YLs, whose cognitive resource availability when completing a given assessment task may be greatly influenced by their age, proficiency level, background knowledge, L1 background, and affective states. The information on cognitive validity may also provide foundational knowledge for developing valid diagnostic assessments for YLs with learning difficulties.

Teacher's Role in Assessment to Assist YLs' Learning

Concerning the centrality of assessment for learning for YLs, there is no question that teachers play a critical role in conducting assessment and using the results to assist YLs' learning. Limited research to date, however, suggests that teachers do not seem to make use of assessment directly to enhance YLs' learning. A series of international surveys by Rixon and her colleagues concerning teachers' assessment practices (Papp and Rixon forthcoming; Rea-Dickins and Rixon 1999, both cited in Rixon 2013, 2016) revealed that the teachers used assessment primarily for summative purposes, especially to see their own teaching effectiveness, and that their assessment practice was often constrained by beliefs and traditions of local teaching and assessment cultures. Similarly, Becker's (2015) implementation study of European Language Portfolio (ELP), which was designed to document YLs' language development based on the CEFR and to promote learners' autonomy and self-efficacy in their learning, was not successful in Germany. The teachers did not use ELP systematically or regularly; they found ELP too complex, time consuming, and unreliable. Becker concluded that "large-scale ELP use and assessment can only be established if teachers readjust their traditional ways of teaching and make changes at the level of lesson and learning and assessment culture"(275). Indeed, assessment for learning requires teachers to undergo conceptual changes. Teachers have to alter their conceptualizations regarding the relationship between learning and assessment, and the role of teachers in the assessment. They might also need to reconsider notions of validity, reliability, and fairness (Davison and Leung 2009). To make the matter more complicated, such conceptual changes need to take place in specific teaching contexts, which may no longer be construct-irrelevant (as in the psychometric tradition) but are often constrained by external factors that go beyond individual teachers (e.g., policy requirements).

In light of assessment for learning, teachers are expected to develop assessment literacy or diagnostic competence. Edelenbos and Kubanek-German (2004), based on their observations of teachers' practice in primary school English classrooms in the Netherlands and Germany, as well as on interviews with the teachers, addressed the importance of developing "diagnostic competence," which they defined as "the ability to interpret students' foreign language growth, to skillfully deal with assessment material and to provide students with appropriate help in response to this diagnosis" (260). According to the authors, diagnostic competence is composed of multiple elements, including the ability to observe and interpret students' performance (including nonverbal responses such as silence and facial expressions); various assessment-related skills such as selecting, analyzing, and adapting diagnostic materials; and abilities to scaffold students' learning (see Edelenbos and Kubanek-German 2004: 277–279 for complete descriptions of diagnostic competence). Although the authors identified the components of diagnostic competence in a primary school foreign language context, they should be applicable to any language teaching context.

Developing such competence does not seem to be easy, however. Torrance and Pryor (1998) reminded us that the initiation-response-evaluation (IRE) exchange, a

very popular classroom discourse initiated by the teacher, can help teachers detect if the student knows what the teacher had in mind, but it does not elicit information about what the student knows or provide the student with useful diagnostic feedback to promote their learning. Butler (2015) examined teachers' engagement in their YLs' task-based assessment in the context of China and found that there was substantial variability in the way that teachers elicited and diagnosed the YLs' English performance during the assessment as a result of their engagement styles. Studies examining teachers' practice of standard-based classroom assessment have also often reported the inconsistency of teachers' interpretations of standards as well as scoring (e.g., Llosa 2011 for a case in the United States).

For the future, given the fact that classroom assessment is deeply embedded in context, we first need more studies describing teachers' daily practice of assessment from different instructional contexts. For example, despite the growing popularity of CLIL, we hardly know how assessment is conducted in CLIL programs (Nikolov 2016b). Second, considering the importance of professional training, more research is needed to address potential gaps between what professional training offers and what teachers actually do in their classrooms. Longitudinal investigation, which follows teachers' cognition and assessment practice before, during, and after trainings, would be of particular interest. Third, in addition to teachers, we need to know more about children's views of assessment. Applied linguists have long treated children as merely an object of observation or treatment when they should be treated as autonomous and active agents (Pinter 2014). It would be a fruitful area of research to investigate how children develop self-regulation of and autonomy over their language learning over time as a result of assessments (e.g., by examining the process of interacting with teachers and allowing YLs to take initiative in assessment through self- and peer-assessment).

Technology and Assessment

As communication through digital devices is expanding, the role of technology in language learning for YLs is growing. Surprisingly, however, empirical investigation on digital technology-mediated assessment for YLs, including assessment using multimedia and digital games, is limited; we know little about how YLs interact with various technologies, their attitudes toward technology, and the potential influence of technology over their performance during assessment. Macaro et al. (2012) review of computer-assisted language learning (CALL) among primary and secondary school students learning English indicated that, although the direct effects of technology on YLs' English learning is rather slim, technology can positively influence YLs' attitudes and behaviors and facilitate collaboration, which all in turn influence their English learning positively. Assuming that technology is increasingly used in language instruction, technology-mediated assessment should be aligned well with instruction. Technology-based assessment also seems to be particularly suitable for YLs. First, considering substantial individual differences among YLs, technology makes it easier to have assessment tailored to individual

children's needs. Its potential for offering YLs instant feedback is an advantage as well. Second, multimodal capabilities can be useful for attracting and maintaining YLs' attention and motivating them to complete assessment tasks. Teachers can capture students' learning processes and trajectories over time through technology without making YLs feel anxious or self-conscious while being assessed. And lastly, using language technology itself likely corresponds well with the cognitive styles of children who grow up with technology. In any event, more empirical research is necessary to test such assumptions. We need to better understand how best to design and administer user-friendly and age-appropriate assessment tasks through technology, what factors may potentially influence validity and reliability of the assessment, how to provide feedback effectively through technology, the impact of technology-mediated assessment over instruction, and fairness issues in technology-mediated assessment. Critically, we may not be able to simply assume that technology-mediated assessment works well for all YLs. Papp and Walczak (2016), in a computer-based standardized English proficiency test, found that YLs who showed a preference for taking a computer version of the test performed better than those who didn't. Lee and Winke (2017) examined children's eye movements on the computer screen during a computerized speaking test. They found that English-learning YLs tended to look at a countdown timer (a potentially distracting element) while their monolingual-English-speaking counterparts tended to look longer at onscreen pictures, which would help them produce speech, suggesting that children's proficiency levels and anxiety levels may have unevenly affected their performance on the computer-based test. More research is needed to better understand how individual factors may interact with YLs' performance when using technology.

Conclusion

Although the number of studies on assessment for YLs has been on the rise, there are still far more questions than answers. Assessment for YLs requires careful considerations of age-related and environmental factors as well as individual differences. Given the vulnerability of YLs, it might be necessary to consider *all* assessment for YLs as having high stakes potentially. The teacher's role in assessment is particularly important for meeting the increasingly diversified and changing needs of YLs and to assist individual children's learning through assessment. And most importantly, we need to keep in mind that children are the center of the whole ecological system of learning and assessment.

Cross-References

- ▶ [Developing Language Curricula for Young Language Learners](#)
- ▶ [Teaching Young Language Learners](#)
- ▶ [Using Assessment to Enhance Learning in English Language Education](#)

References

- Abedi J, Hofstetter CH, Lord C (2004) Assessment accommodations for English language learners: implications for policy-based empirical research. *Rev Educ Res* 74(1):1–28
- Anderson CE, McDougald JS, Cuesta L (2015) CLIL for young learners. In: Giannikas CN, McLaughlin L, Fanning G, Muller ND (eds) *Children learning English: from research to practice*. Garnet Education, Reading, pp 137–151
- Bailey AL, Butler FA (2007) A conceptual framework of academic English language for broad application to education. In: Bailey AL (ed) *The language demands of school: putting academic English to the test*. Yale University Press, New Haven, pp 68–102
- Bailey AL, Heritage M (2014) The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Q* 48(3):480–506
- Bailey AL, Huang BH (2011) Do current English language development/proficiency standards reflect the English needed for success in school? *Lang Test* 28(3):343–365
- Bailey AL, Ospova AV (2016) *Children’s multilingual development and education: fostering linguistic resources in home and school context*. Cambridge University Press, Cambridge, MA
- Ballantyne KG (2013) Disproportional representation of English learners among students identified with disabilities; equity and the federal accountability system. In: Tsagari D, Spanoudis G (eds) *Assessing L2 students with learning and other disabilities*. Cambridge Scholars, Newcastle upon Tyne, pp 3–25
- Becker C (2015) Assessment and portfolios. In: Bland J (ed) *Teaching English to young learners: critical issues in language teaching with 3–12 year olds*. Bloomsbury, London, pp 261–278
- Benigno V, de Jong J (2016) The “global scale of English learning objectives for young learners”: a CEFR-based inventory of descriptors. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 43–64
- Black PJ, Wiliam D (1998) *Assessment and classroom learning*. *Assess Educ* 5:7–73
- Bronfenbrenner U (1979) *The ecology of human development: experiments by nature and design*. Harvard University Press, Cambridge, MA
- Bruner JS (1975) Language as an instrument of thought. In: Davies A (ed) *Problems of language and learning*. Heinemann, London, pp 61–88
- Butler YG (2015) Task-based assessment for young learners: old meets new cultures. In: Thomas M, Hayo R (eds) *Contemporary task-based language teaching in Asia*. Bloomsbury, London, pp 348–365
- Butler YG (2016) Self-assessment of and for young learners’ foreign language learning. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 291–315
- Butler YG (2018a) The role of context in young learners’ processes for responding to self-assessment items. *Mod Lang J* 102(1):242–261
- Butler YG (2018b) Young learners’ processes and rationales for responding to self-assessment items: cases for generic can-do and five-point Likert-type formats. In: Davis J et al (eds) *Useful assessment and evaluation in language education*. Georgetown University Press, Georgetown, pp 21–39
- Butler YG, Lee J (2006) On-task versus off-task self-assessment among Korean elementary school students studying English. *Mod Lang J* 90(4):506–518
- Butler YG, Lee J (2010) The effect of self-assessment among young learners. *Lang Test* 17(1):1–27
- Butler YG, Zeng W (2011) The roles that teachers play in paired-assessments for young learners. In: Tsagari D, Csépes I (eds) *Classroom-based language assessment*. Peter Lang, Frankfurt, pp 77–92
- Butler YG, Zeng W (2014) Young foreign language learners’ interactions during task-based paired assessment. *Lang Assess Q* 11:45–75
- Carless D, Lam R (2014) The examined life: perspectives of lower primary school students in Hong Kong. *Education* 3–13(42):313–329
- Carpenter K, Fujii N, Kataoka H (1995) An oral interview procedure for assessing second language abilities in children. *Lang Test* 12:158–181

- Ćatibušić B, Little D (2014) *Immigrant pupils learn English: a CEFR-related empirical study of L2 development*. Cambridge University Press, Cambridge
- Clark E (2009) *First language acquisition*. Cambridge University Press, Cambridge
- Cook VJ (1992) Evidence for multicompetence. *Lang Learn* 42(4):557–591
- Council of Europe (2001) *Junior European Language Portfolio*. Available via http://www.primarylanguages.org.uk/resources/assessment_and_recording/european_languages_portfolio.aspx
- Cummins J (1979) Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* 19:121–129
- Davison C, Leung C (2009) Current issues in English language teacher-based assessment. *TESOL Q* 43(3):393–415
- Edelenbos P, Kubanek-German A (2004) Teacher assessment: the concept of ‘diagnostic competence’. *Lang Test* 21(3):259–283
- Espinosa L (2013) Assessment of young English-language learners. In: Chapelle CA (ed) *The encyclopedia of applied linguistics*. Blackwell, London. <https://doi.org/10.1002/9781405198431.wbeal10057>
- Field J (2011) Cognitive validity. In: Taylor L (ed) *Examining speaking: research and practice in assessing second language speaking*. Cambridge University Press, Cambridge, pp 65–111
- Flores N, Schissel JL (2014) Dynamic bilingualism as the norm: envisioning a heteroglossic approach to standards-based reform. *TESOL Q* 48(3):454–479
- Fry AF, Hale S (2000) Relationships among processing speed, working memory, and fluid intelligence in children. *Biol Psychol* 54:1–34
- García O, Li W (2014) *Translanguaging: language, bilingualism and education*. Palgrave Macmillan, London
- García Mayo MP, Agirre AI (2016) Task repetition and its impact on EFL children’s negotiation of meaning strategies and pair dynamics: an exploratory study. *Lang Learn J* 44(4):451–466
- Grosjean F (2010) *Bilingual: life and reality*. Harvard University Press, Cambridge, MA
- Gu L (2014) Language ability of young English language learners: definition, configuration, and implications. *Lang Test* 32(1):21–38
- Hasselgreen A (2005) Assessing the language of young learners. *Lang Test* 22(3):337–354
- Hung Y, Samuelson BL, Chen S (2016) Relationship between peer- and self-assessment and teacher assessment of young EFL learners’ oral presentations. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 317–338
- Inbar-Lourie O, Shohamy E (2009) Assessing young language learners: what is the construct? In: Nikolov M (ed) *The age factor and early language learning*. Mouton de Gruyter, Berlin, pp 83–96
- Jang EE (2014) *Focus on assessment*. Oxford University Press, Oxford
- Jang EE, Vincett JM, van der Boom EH, Lau C, Yang Y (2017) Considering young learners’ characteristics in developing diagnostic assessment intervention. In: Wolf MK, Butler YG (eds) *English language proficiency assessments for young learners*. Routledge, New York, pp 193–213
- Johns AM (1997) *Text, role, and context: developing academic literacies*. Cambridge University Press, Cambridge
- Kachru BB (ed) (1992) *The other tongue: English across cultures*. University of Illinois Press, Urbana
- Kormos J (2017) The effects of specific learning difficulties on processes of multilingual language development. *Annu Rev Appl Linguis* 37:30–44
- Lee S, Winke P (2017) Young learners’ response processes when taking computerized tasks for speaking assessment. *Lang Test*. <https://doi.org/10.1177/0265532217704009>
- Little D (2007) The common European framework of reference for languages: perspectives on the making of supranational language European policy. *Mod Lang J* 91(4):645–655
- Losa L (2011) Standards-based classroom assessments of English proficiency: a review of issues, current developments, and future directions for research. *Lang Test* 28(3):367–382
- Macaro E, Handley Z, Walter C (2012) A systematic review of CALL in English as a second language: focus on primary and secondary education. *Lang Teach* 45(1):1–43

- McClelland MM, Cameron CE, Duncan R, Bowles PP, Acock A, Miao A, Pratt ME (2014) Predictors of early growth in academic achievement; the head-toes-knees-shoulders task. *Front Psychol* 5. <https://doi.org/10.3389/fpsyg.2014.00599>
- McKay P (2006) *Assessing young language learners*. Cambridge University Press, Cambridge
- Menken K, Shohamy E (2015) Invited colloquium on negotiating the complexities of multilingual assessment, AAAL conference 2014. *Lang Teach* 48(3):421–425
- Menken K, Hudson T, Leung C (2014) Symposium: language assessment in standards-based education reform. *TESOL Q* 48(3):586–614
- Messick S (1996) Validity and washback in language testing. *Lang Test* 13(3):241–256
- Mills KL (2014) Effects of Internet use on the adolescent brain: despite popular claims, experimental evidence remains scarce. *Trends Cogn Sci* 18(8):385–387
- Muñoz C (2017) Tracing trajectories of young learners: ten years of school English learning. *Annu Rev Appl Linguis* 37:168–184
- Nikolov M (ed) (2016a) *Assessing young learners of English: global and local perspectives*. Springer, New York
- Nikolov M (2016b) Trends, issues, and challenges in assessing young language learners. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 1–17
- North B (2014) *The CEFR in practice*. Cambridge University Press, Cambridge
- Oliver R (2002) The patterns of negotiation for meaning in child interactions. *Mod Lang J* 86(1):97–111
- Oliver R, Azkarai A (2017) Review of child second language acquisition (SLA): examining theories and research. *Annu Rev Appl Linguis* 37:62–76
- Papageorgiou S, Baron P (2017) Using the common European framework of reference to facilitate score interpretations for young learners' English language proficiency assessments. In: Wolf MK, Butler YG (eds) *English language proficiency assessments for young learners*. Routledge, New York, pp 136–152
- Papp S, Walczak A (2016) The development and validation of a computer-based test of English for young learners: Cambridge English young learners. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 139–190
- Perry NE, VandeKamp KJO (2000) Creating classroom contexts that support young children's development of self-regulated learning. *Int J Educ Res* 33:821–843
- Pinter A (2014) Child participant roles in applied linguistics research. *Appl Linguis* 35(2):168–183
- Pinter A (2015) Task-based learning with children. In: Bland J (ed) *Teaching English to young learning: critical issues in language teaching with 3–12 year olds*. Bloomsbury, London, pp 113–127
- Poehner ME, Zhang J, Lu X (2017) Computerized dynamic assessment for young language learners. In: Wolf MK, Butler YG (eds) *English language proficiency assessments for young learners*. Routledge, New York, pp 214–233
- Prošić-Santovac D, Rixon S (forthcoming) Integrating assessment into early language learning and teaching practice. *Multilingual Matters*, Bristol
- Purpura JE (2016) Second and foreign language assessment. *Mod Lang J* 100(supplement 2016): 190–208
- Rea-Dickins P (2001) Mirror, mirror on the wall: identifying processes of classroom assessment. *Lang Test* 18(4):429–462
- Rixon S (2013) British Council survey of policy and practice of primary English teaching worldwide. Retrieved from www.britishcouncil.org
- Rixon S (2016) Do development in assessment represent the 'coming of age' of young learners English language teaching initiatives?: the international picture. In: Nikolov M (ed) *Assessing young learners of English: global and local perspectives*. Springer, New York, pp 19–41
- Saville N, Weir CJ (2018) *Examining young learners: research and practice in assessing the English of school-age learners*. Cambridge University Press, Cambridge

- Scarcella R (2003) *Academic English: a conceptual framework*. University of California Linguistic Minority Institute, Santa Barbara
- Song J (2011) Globalization, children's study abroad, and transnationalism as an emerging context for language learning: a new task for language teacher education. *TESOL Q* 45(4):749–758
- Street BV (1996) Academic literacies. In: Baker D, Clay J, Fox C (eds) *Alternative ways of knowing: literacies, numeracies, sciences*. Falmer Press, London, pp 101–134
- Torrance H, Pryor J (1998) *Investigating formative assessment: teaching, learning and assessment in the classroom*. Open University Press, Buckingham
- Tsagari D, Spanoudis G (eds) (2013) *Assessing L2 students with learning and other disabilities*. Cambridge Scholars, Newcastle upon Tyne
- Vygotsky LS (1962) *Thought and language*. MIT Press, Cambridge, MA
- Vygotsky LS (1978) *Mind and society: the development of higher mental processes*. Harvard University Press, Cambridge, MA
- WIDA (n.d.) English language development (ELD) standards. Available via <https://www.wida.us/standards/eld.aspx>. Accessed 1 Nov 2017
- Wolf MK, Butler YG (eds) (2017) *English language proficiency assessments for young learners*. Routledge, New York