Self-Assessment *of* and *for* Young Learners' Foreign Language Learning

Yuko Goto Butler

Abstract Despite the recent focus on self-assessment (SA) as a tool for enhancing learning, some researchers and practitioners have expressed concerns about its subjectivity and lack of accuracy. Such concerns, however, originated from the traditional, measurement-based notion of assessment (assessment *of* learning) rather than the learning-based notion of assessment (assessment *for* learning). In addition, existing research on SA in second/foreign language education has been concentrated on adult learners, leaving us with limited information on SA among young learners. In this chapter, I address both sets of issues: the confusion between the two orientations for assessment and age-related concerns regarding SA. First, I clarify the two orientations of assessment—*assessment of learning* and *assessment for learning*—and demonstrate that most of the concerns about subjectivity and accuracy apply primarily to the former orientation. Second, I detail the current findings on SA among young learners and identify the most urgent topics for future research in this area. Finally, to help teachers and researchers examine and develop SA items that are most appropriate for their purposes, I propose five dimensions that characterize existing major SAs for young learners: (a) domain setting; (b) scale setting; (c) goal setting; (d) focus of assessment; and (e) method of assessment.

Keywords Self-assessment • Assessment for learning • Assessment of learning • Young learners • Foreign language learning • Summative assessment • Formative assessment • Age

1 Introduction

Coupled with the recent emphasis on learner-centered approaches to language teaching and self-regulated language learning, the use of various forms of SA is on the rise in language programs worldwide. According to Oscarson (1989), SA is a type of assessment where learners engage in "internal or self-directed" activities; as such,

Y.G. Butler (✉)
Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA
e-mail: ybutler@gse.upenn.edu

SA is "fundamentally different" from assessment conducted from the perspective of external agents such as teachers and test administrators (p. 1). In recent years, SA has gained popularity even among educators of young language learners (typically defined as children up to the end of primary school or sometime around 12 years old). It is no longer uncommon, for example, to see SA items in language textbooks for young learners. Primary school curricula often encourage teachers to use SA as a tool for evaluating students' performance.

The growing attention paid to SA in early language education may be due to the fact that SA is considered to be a low-stakes form of assessment and so is assumed to be less stressful for young learners. Researchers have developed various types of can-do statements, a form of SA, for young learners and have made the statements readily available for teachers. Major efforts of developing can-do statements include CILT's *Can-do Speech Bubble,* as part of Language Portfolio in the U.K. (CILT The National Center for Languages, 2006), and *Lingua Folio Junior* (National Council of State Supervisors for Languages, 2014)*,* based on the American Council on the Teaching of Foreign Language (ACTFL) Proficiency Guidelines; the *Lingua Folio Junior* has been implemented on a trial basis in select U.S. states.

Despite the growing interest in SA, it has not had as large a presence in second language (L2) and foreign language (FL) classrooms at the primary school level as policy makers may have expected. The reasons for the slow take-up of SA in practice are presumably varied. But, perhaps most importantly, teachers often do not seem to know how or why they should use SA. Concerns have been expressed about the extent to which SA accurately captures young learners' actual performance. Subjectivity has also been cited as a serious concern, particularly when SA is implemented in so-called "exam-driven" teaching and learning contexts.

Importantly, concerns regarding the accuracy and subjectivity of SA apply primarily to the traditional, measurement-based notion of assessment (assessment *of* learning) and thus are most relevant when SA is used primarily for summative purposes. When SA is implemented primarily for formative purposes, its accuracy may not be critical. From a process-oriented view of assessment (assessment *for* learning), assessment is considered to be a process of seeking relevant information, interpreting that information so that learners can reflect on their own learning, and making constructive decisions for further learning. As such, when the assessment is for learning, traditional psychometric notions of validity and reliability may not be suitable. Indeed, as Brookhart (2003) suggested, we need to sort out "classroometric" measurement concepts from psychometric measurement concepts (p. 8).

The major motivation for policies to promote SA for primary school teachers came from the theoretical association between SA and learning. Researchers agree that SA is a vital process for facilitating learners' autonomy and self-regulation (Black & Wiliam, 1998, 2009; Blanche & Merino, 1989; Butler & Lee, 2010; Dickinson, 1987; Oscarson, 1989). The premise that SA can be aligned with self-regulated learning sounds promising. However, we still have only a limited understanding of how SA can best be used to facilitate children's language learning. What kinds of feedback during and/or after SA would promote young learners'

self-reflection, which, in turn, would lead to further language learning? Researchers have just begun to explore these questions.

In addition, previous research on SA in L2/FL language education has predominantly dealt with adult learners and has paid little attention to the role of age in SA. Age-related concerns—such as the extent to which children can handle self-assessing their performance or abilities in L2/FL in the first place—should be addressed to inform practice.

In this chapter, I clarify the two assessment orientations (namely, *assessment of learning* and *assessment for learning*) while focusing on SA among young learners, and I discuss the possibilities and challenges of implementing SA among young learners from both points of view. I draw on examples from previous studies to illustrate my points. I then characterize major existing SA item types according to five dimensions, and discuss how different types of SA can be used for both assessment of learning and assessment for learning. I conclude by offering suggestions for future research on SA for young learners.

2 Two Approaches to SA for Young Learners

In the following sections, I discuss the two approaches for assessment (assessment *of* and *for* learning) in turn. They originated from different theoretical and epistemological traditions, and the distinctions need to be clarified. That being said, however, these approaches are not necessarily mutually exclusive but can instead be located on a continuum according to the degree of emphasis on learning. In practice, the same SA tool can be used for more evaluation-oriented means (assessment of learning) or for more learning-oriented means (assessment for learning).

2.1 *Self-Assessment* of *Learning*

In the assessment of learning orientation, assessment is a means of capturing a learner's true ability. Thus, the assessment is concerned with eliciting meaningful information for making accurate and consistent inferences about a learner's true ability. The learner is a subject being observed and is external to the inferences being made and the actions being taken as the result of the inferences (Brookhart, 2003).

2.1.1 SA as Assessment *of* Learning Among Adult Learners

Among adult learners, a great deal of research has been conducted with respect to the validity and reliability of SA as well as its use. With a few exceptions (e.g., Matsuno, 2009; Patri, 2002; Pierce, Swain, & Hart, 1993), there is ample evidence

indicating that SA results, at least among adults, are generally correlated with external criteria such as teachers' ratings, final grades in class, and objective tests (Bachman & Palmer, 1989; Blanche, 1990; Brantmeier & Vanderplank, 2008; Brantmeier, Vanderplank, & Strube, 2012; Dickinson, 1987; Hargan, 1994; Leach, 2012; Oscarson, 1997; Stefani, 1994). As a result, SA has been used for relatively high-stakes purposes, such as program placement (Hargan, 1994; LeBlanc & Painchaud, 1985) and choosing the appropriate level of tests (Malabonga, Kenyon, & Carpenter, 2005). However, the degrees of correlations with external criteria varied across studies. Factors that influenced accuracy of SA included the skill domain being assessed, the ways in which items were constructed, and learners' individual characteristics.

With respect to the skill domains being assessed, if we assume that productive skills (i.e., speaking and writing) require higher degrees of meta-awareness, such as pre-planning and self-monitoring, than receptive skills (i.e., listening and reading), we may expect that learners are better at self-assessing their productive skills than their receptive skills. Interestingly, in a meta-analysis of SA, Ross (1998) found the opposite to be the case: adult learners could self-assess their receptive skills (reading in particular) in L2/FL more accurately than their productive skills. It is not clear, however, if receptive skills are inherently easier to self-assess. In speculating about which factors might explain the surprising result, Ross suggested such things as learners' experiences (e.g., adult L2/FL learners at college are more likely to have engaged in reading activities more heavily than the other activities), the reference points that they used (e.g., the adult learners might have judged themselves in relation to the performances of other students in class), and the scales that were used in external measurements (e.g., writing assessments often use nominal or categorical scales that may not be readily applicable to correlational analyses). In general, people tend to more accurately self-assess lower order cognitive skills than they do higher order cognitive skills (Zoller, Tsaparlis, Fastow, & Lubezky, 1997).

Second, how the items are worded and constructed influences learners' responses to SA. College students' responses differed based on whether the items were negatively worded (e.g., "I have trouble with…" and "I cannot do….") or positively worded (e.g., "I can do …"), although the degree of inconsistency varied greatly depending on the items (Heilenman, 1990). Not too surprisingly, learners' SA accuracy improved when the items were provided in their L1 rather than the target language (Oscarson, 1997).

Finally, various factors associated with individual learners are also found to influence their SA accuracy. One of the factors studied most extensively is learners' proficiency levels and experiences with the target language (Blanche & Merino, 1989; Davidson & Henning, 1985; Heilenman, 1990; Orsmond, Merry, & Reiling, 1997; Stefani, 1994; Sullivan & Hall, 1997). These studies generally indicate that students with lower proficiency and/or less experience with the target language tend to overestimate their performance, whereas student with higher proficiency tend to be more accurate or underrate their performance. Other influential factors over the accuracy of SA responses include the ways in which learners understand and respond to scales and items (Heilenman, 1990), the ways in which learners retrieve

relevant memory to self-assess the given skills and performance (Ross, 1998; Shameem, 1998), learners' learning styles (Cassidy, 2007); their anxiety levels (MacIntyre, Noels, & Clément, 1997), and their levels of self-esteem and motivation (AlFallay, 2004; Dörnyei, 2001). Another important factor, which is of particular relevance to the current discussion, is the age of the learners.

2.1.2 SA as Assessment *of* Learning Among Young Learners

Research on SA as an assessment *of* L2/FL learning among young learners has been very limited so far. It is largely unclear if the results of studies among older learners described in Sect. 2.1.1 are applicable to young learners of L2/FL.

Responding to SA requires highly complicated mental processing. For example, consider the item "I can ask questions in class," which is included in O'Malley and Pierce's (1996) popular resource book for young learners of English as L2. In order to respond to this item using a 4-point scale (ranging from *not very well*, *okay*, *well*, to *very well*) as instructed, the children need to go through at least the following cognitive processes:

1. Comprehend what the item refers to (what it means to "ask questions");
2. Understand each scale level and differentiate them (what it means to say "I can ask questions *okay*" and how that statement differs from "I can ask questions *well*");
3. Retrieve and synthesize their recent linguistic performance of asking questions in class;
4. Set a reference point to make a judgment (making a judgment in relation to others in class, in relation to the learner's own goal, or based on some other criteria).

While Harris (1997) asserts that "younger learners may be less resistant to the concept of self-assessment" (p. 18), given the complexity of cognitive processing required for answering SA, one may wonder about the extent to which children can accurately assess their own performance and abilities. From the *assessment of learning* point of view, at least two major issues must be examined: (a) how we should interpret children's responses to SA items (interpretation-related issues); and (b) the factors that influence the accuracy of their SA responses (measurement-related issues). I discuss these issues, which are summarized in Fig. 1, in the following sections.

2.1.2.1 Interpretation-Related Issues in Young Learners' SA of Learning

Previous research on children's development of self-appraisal and competence indicates that young learners' self-appraisal has been consistently high regardless of their actual performance. More specifically, children's self-appraisal remains very positive during the pre-school and early primary school years, and it starts declining
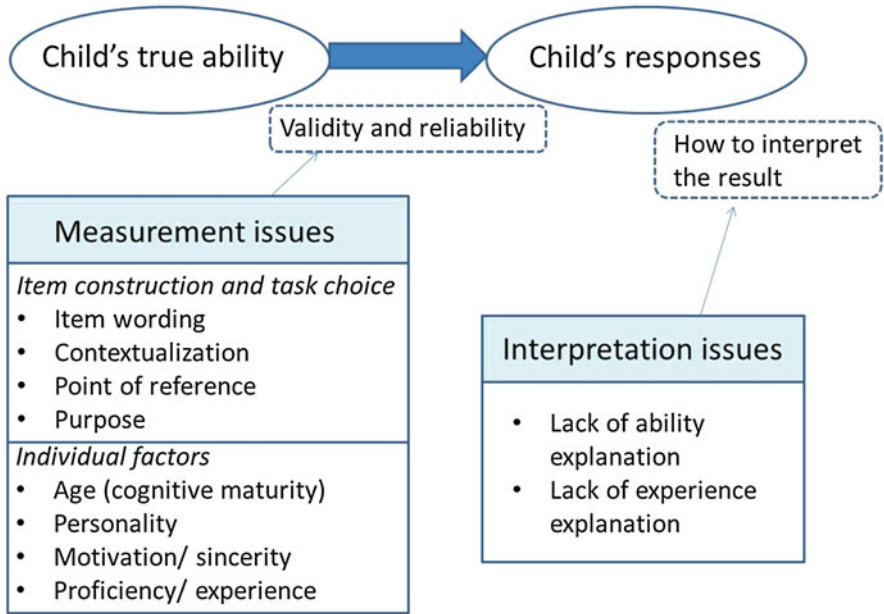
Fig. 1 Two major issues for self-assessment of learning for young learners (Note: SA of learning primarily concerns how best to elicit children's true abilities. In the process, there are two major issues: measurement issues and interpretation issues)

sometime around the ages of 7–9, with another drop around the ages of 11–13. The accuracy of children's perceived competence (examined by calculating correlations with external measures such as their teachers' ratings) increases after the age of 8, when they start using social-comparative information (information indicating that one's performance or ability is superior or inferior to others). Although social-comparative information begins to influence children's self-appraisal of *performance* by the time they are around 7 years old, it does not influence self-appraisal of their *abilities* until much later (around 11–12 years old) (R. Butler, 2005).

Researchers' interpretations of children's self-appraisal behaviors have been changing in recent years. Traditionally, children's unrealistically high self-appraisal was mainly attributed to their lack of cognitive maturity for making accurate judgments about their performance and abilities. Piaget's (1926/1930) well-known stage theory of cognitive development certainly made a tremendous impact on researchers' interpretation. According to this theory, children at the preoperational stage (ages 2–7) struggle with logical thinking; instead, their thoughts are dominated by concrete reasoning and intuition. This theory also posits that children are egocentric and have a hard time taking other people's perspectives. The theory goes on to say that children at the concrete operational stage (ages 7–11) gradually begin to operate logical thinking and to differentiate their own thoughts from those of others. However, they still have difficulty handling multiple perspectives systematically

and forming abstract and causal thinking. In line with this theory, Stipek (1984) offered an explanation for why children are not only unrealistic but also excessively positive in their perceived competence by proposing their "wishful thinking" interpretation; namely, children cannot distinguish reality from their wishes, and they tend to make decisions based on the latter.

Similarly, interpretations based on achievement goal theory assumed that children's accuracy in evaluating their own abilities would be partially based on the development of their conception of *ability*. The theory proposed that there are two distinctive goal perspectives when perceiving one's ability: a *task-goal perspective* and an *ego-goal perspective*. The task-goal perspective is based on one's subjective assessment of task achievement and mastery. The ego-goal perspective relies on one's demonstration of superior performance compared to others (Dweck, 1986; Nicholls, 1989). According to this theory, children up to 7 years old cannot distinguish between ability and effort when it comes to determining performance on a task (referred to as *undifferentiated conception of ability*); for them, effort *is* ability. Thus, for young children, a person with high ability refers to a person who makes effort or obtains a high score in a given task, but they do not understand how to conceptualize a person who makes effort but achieves low in the given task, or vice versa. Researchers believed that young children are relatively invulnerable to failure and that they tend to respond to the failure by increasing effort. They also believed that children do not fully develop the concept of normative difficulty; instead, they tend to judge task difficulty in an egocentric fashion (e.g., this task is difficult because it was hard for me) (Nicholls & Miller, 1983). As they grow, children gradually understand that there is a cause-and-effect relationship between effort and outcome (outcome is a result of effort). But according to this theory, it is only after children reach the ages of 11–12 that they fully understand that one's performance (outcome) is also constrained by one's ability (referred to as *mature conception of ability*). After children reach this level, they can construct perceived competence in relation to other people's performance (Nicholls, 1978; also see Mihaljević Djigunović, 2016 in this volume.)

If children's self-evaluative abilities are mainly constrained by their underdeveloped internal mental structures, it makes sense to hold off on implementing SA of learning until they reach a cognitively mature state. However, in contrast to the results of experimental studies, anyone who spends sufficient time with children may notice that they appear to have more sophisticated self-evaluative knowledge and skills in naturalistic contexts than the cognitive-developmental theories predict. Indeed, neo- or post-Piagetian researchers indicate that children's self-evaluative abilities vary greatly depending on contexts, domains, and tasks at a given age level (see Flavell, 1999, for a review of such studies). Children's self-appraisal becomes more accurate if they can engage in familiar tasks and tasks that require lower levels of cognitive demand to perform. Experiences with different domains (e.g., math, music, language) help them develop distinct, domain-specific, and stable self-evaluative competence. Children who have intensive social contacts with other children can use normative information (information based on social comparison) more appropriately and are less ego-centric than those who don't, as we can see,

for example, in the work of Vygotsky (1934/1978). Children may also be more vulnerable to failure than was previously thought (Carless & Lam, 2014). R. Butler (2005) argued that:

> regarding competence assessment, one implication is that self-appraisal may indeed become more accurate, differentiated and responsive to relevant information with age, in large part, however, because of age-related changes in children's typical experiences and contexts, rather than their internal cognitive structures. (p. 208)

In addition, potential problems have been raised with respect to the methodologies of many earlier studies of cognitive development. Children's failure in tasks may not be a sign of their lack of abilities but may be due to their misunderstanding the researchers' questions or intentions. For example, children as young as 4–5 who were once thought to be incapable of rating their performance based on temporal comparison (i.e., comparing their current performance with that in the past) turned out to be able to so as long as the information provided to them for evaluation was meaningful and familiar to them (R. Butler, 2005). These more recent findings on and interpretations of children's assessment competence remind us that we need to pay careful attention to contexts, assessment task choice, and the ways in which SA is constructed and delivered.

2.1.2.2 Measurement-Related Issues in Young Learners' SA of Learning

It is also important to understand measurement-related factors that contribute to children's biases and influence the accuracy of SA responses during the administration of *SA of learning*. As shown in Fig. 1, such measurement-related factors can largely be classified into two types: (a) item construction and task choice issues; and (b) individual factors, such as the child's age, personality, and proficiency.

The factors listed in Fig. 1 are based on previous studies, which were conducted primarily among adult L2 learners. How these factors may influence children's self-assessment responses is largely unknown.

As I examine in detail in later sections, different formats of SA items have been used; some SA tools employ multiple-choice formats while others ask learners dichotomous questions (i.e., requiring either "Yes/Can do" or "No/Cannot do" responses). Many SA items for young learners are short and simple, but some SA items provide the learner with more detailed contextual information. There has been very limited research examining if children have response biases based on different SA formats and item wording when assessing their L2/FL abilities. In a clinical setting, Chambers and Johnston (2002) found that, when asked to rate their own feelings (referred to as a *subjective task*) and other people's feelings (referred to as a *social objective task*) in a Likert scale, younger children (5–6 year olds) tended to show more extreme responses in both tasks than older children (7–9 year olds and 10–12 year olds). However, this response bias was not observed when the same children were asked to rate physical characteristics described in pictures using a Likert scale (referred to as an *objective task*) even among the youngest group that

they examined. Interestingly, the response bias observed in the youngest group was *not* found to be a function of the number of choices in the Likert scales; their responses did not differ between the three-level and five-level Likert scales. We do not know, however, if dichotomous items would have made any difference on the children's responses. Judging from the previous studies conducted in domains other than L2/FL, children do not seem to handle negatively worded items well (e.g., "I am not good at doing math") compared with positively worded items (e.g., "I am good at doing math") (e.g., Marsh, 1986). Considering the possible domain specificity of children's responses, however, we need to examine whether a similar response bias is observed when children self-evaluate their L2/FL.

SA items are often highly decontextualized—see, for example, the item "I can ask questions in class," which I quoted from O'Malley and Pierce (1996) in Sect. 2.1.2. However, depending on the age of children, the degree of contextualization can be a potential threat to the validity of SA of learning. In a study I did with a colleague (Butler & Lee, 2006), we compared children's (9–10 year olds and 11–12 year olds) responses to two formats of SA, an *off-task SA* and an *on-task SA*, concerning their oral performance in an FL. The off-task SA was a type of SA that asked learners to self-evaluate their general performance in a decontextualized fashion, as exemplified by the example item I quote above. The on-task SA was a contextualized SA in which learners were asked to self-evaluate their performance in a specific task immediately after the task was completed. We compared the children's responses to these two types of SA items with an objective proficiency measurement and an assessment of the children based on their teachers' classroom observations. We found that the children could self-assess their performance more accurately in the contextualized format than the decontextualized format. Not too surprisingly, the younger group (9–10 years) had a harder time with the decontextualized format than the older group. We also found that the children's responses to the contextualized format, compared with the decontextualized format, were less influenced by their attitudes and personality factors.

Considering the potential age- and experience-related challenges children may face when making temporal and/or normative comparisons while self-evaluating their abilities (see Sect. 2.1.2.1), it seems safe to assume that how researchers define reference points for SA (e.g., setting learners' own previous performance or other people's performance as a reference point) will influence children's responses to the SA items. Unfortunately, we know little about how children rely on different reference points when they assess their L2/FL abilities. In fact, our knowledge of the self-assessing process is quite limited, even when considering adult learners. Moritz's (1995) exploratory study based on a think-aloud protocol and retrospective interviews revealed that college students of French as FL used a variety of reference points (both temporal and normative information) when self-assessing their French abilities.

We can also assume that the extent to which young learners of L2/FL understand the purpose of SA influences the accuracy of their responses. In an intervention study of SA that I conducted with Lee (Butler & Lee, 2010), one of the challenges that the participating primary school teachers reported was how to provide their

students with initial guidance in order for them to treat SA seriously. It was particularly challenging to implement SA in a competitive, exam-driven environment. A teacher who taught in a competitive environment told us that she believed that SA had to be tied to other assessments or grades in order to ensure the accuracy of her students' responses. However, a teacher who taught in a much less competitive environment did not see such measures as necessary. We know from the research on the development of self-appraisal among children that their motivation for responding to SA accurately seems to increase with age but not in a linear fashion. Moreover, their motivation for accurate SA is also influenced by the amount of domain-specific knowledge they have acquired as well as by the context in which SA is conducted. For example, children's positive bias is motivated if the context and culture value positive self-appraisal. Accuracy of response is also constrained (more likely negatively biased) if the child realizes that there is a social cost for aggrandizing self-appraisal (R. Butler, 2005). In any event, we need more studies on how best to situate SA so that children of different ages can understand the purpose of SA and are motivated to respond to SA accurately in their specific learning environments.

In addition to the issues related to item construction and task choice, various individual factors likely influence the accuracy of children's SA responses. Such factors include cognitive maturity, personality, motivation, proficiency in the target language, and experience with SA. The role of individual differences in children's responses in SA is an unexplored area of inquiry, and so I can offer no practical, research-based suggestions for ensuring the accuracy of SA of learning among children.

2.2 *Self-Assessment* for *Learning*

While research on SA to date has been conducted primarily from an *assessment of learning* orientation, researchers have been giving increasing attention to SA as a formative assessment, with the goal of discovering its potential for influencing learning.

In taking the assessment for learning approach, the relationship between validity and reliability may need to be conceptualized differently. According to Sadler (1989), in the traditional assessment of learning, higher reliability is necessary but not sufficient for ensuring higher validity; a test can be highly reliable but can be off target. Thus, reliability serves as a precondition for validity. In contrast, with assessment for learning, validity should be a precondition for reliability because, according to Sadler (1989), "attention to the validity of judgments about individual pieces of work should take precedence over attention to reliability of grading in any context where the emphasis is on diagnosis and improvement" (p. 122).

Validity and reliability can themselves be conceptualized very differently depending on which approach is used. In the assessment for learning orientation, assessment is considered as part of instruction and "is usually informal, embedded in all aspects of teaching and learning" (Black, Harrison, Lee, Marshall, & Wiliam,

2003, p. 2). In assessment for learning, *validity* refers to the extent to which both the content of the assessment and the assessments' methods and tasks are matched with instruction. Thus, assessment for learning is deeply embedded in the particular context of the assessment. In assessment for learning, learners are no longer merely objects being measured; they are active participants who make inferences and take actions, together with the teachers, for formative purposes. According to Brookhart (2003), the validity concerns of assessment for learning include the degrees and the ways in which learners can self-reflect and benefit from having assessment enhance their learning. Similarly, teachers' knowledge, beliefs, and practices are all part of the validity concerns as well. In assessment for learning, *reliability* refers to the degree of stability of "information about the gap between students' work and 'ideal' work (as defined in students' and teachers' learning objectives)" (p. 9).

By engaging learners in self-reflection, SA is considered to be effective for developing their self-regulation, which can be defined as "the self-directive process by which learners transform their mental abilities into academic skills" (Zimmerman, 2002, p. 65), and should enhance their motivation and learning. However, empirical studies examining the effect of SA on learners' motivation and learning have been limited, particularly in relation to L2/FL.

2.2.1 SA as Assessment *for* Learning Among Adult Learners

Among adult learners, intervention studies of SA indicate that learners' perceived effects of SA were generally positive. For example, Orsmond, Merry, and Reiling (1997) found that, out of 105 college-level biology students, 98 % of them thought that SA made them think more and 71 % thought that they learned more, and 90 % found SA beneficial. Similarly, in Stefani (1994), out of 87 college students who conducted SA and 67 students who conducted peer-assessment in biochemical studies, nearly 100 % said that SA or peer-assessment procedures made them think more, and 85 % said they could learn more using these procedures than using the traditional tutor-lead assessment.

A number of studies on adults employed objective measures, such as external tests, grades, and teachers' or tutors' evaluations, in order to examine the effectiveness of SA on learning, and they identified some factors that led to positive outcomes. Such factors included receiving sufficient training to conduct SA (McDonald & Boud, 2003), setting clear criteria or rubrics (Andrade, Wang, Du, & Akawi, 2009), and having feedback (Taras, 2002). To facilitate learners' understanding of criteria and rubrics, researchers have suggested that presenting descriptive statements along with examples (e.g., writing examples for writing rubrics) would be effective. Having opportunities to discuss the meaning of the criteria with the teachers and tutors made the learners think more. Learning outcomes were different when the learners were allowed to construct their own criteria and when they were given criteria (Orsmond, Merry, & Reiling, 2000). Because peer-assessment should help learners understand the criteria better, it has been suggested that peer-assessment be implemented before SA (e.g., Nicol & Macfarlane-Dick, 2006).

This may make sense, particularly considering that peer-assessment was found to be psychometrically more internally consistent and to have higher correlations with external measures than SA (Matsuno, 2009; Patri, 2002) but that SA helped to increase learning more than peer-assessment (Sadler & Good, 2006).

Feedback is an essential part of SA for it to be effective for learning (Sadler, 1989), but having feedback itself does not guarantee positive outcomes. Hattie and Timperley's (2007) meta-analysis on feedback showed that there were substantial differences in effect sizes across studies, indicating that the quality and timing of the feedback greatly influenced learners' performance. Nicol and Macfarlane-Dick (2006) listed seven principles for good feedback practice:

> (1) helps clarify what good performance is (goals, criteria, expected standards); (2) facilitates the development of self-assessment (reflection) in learning; (3) delivers high quality information to students about their learning; (4) encourages teacher and peer dialogue around learning; (5) encourages positive motivational beliefs and self-esteem; (6) provides opportunities to close the gap between current and desired performance; and (7) provides information to teachers that can be used to help shape teaching. (p. 205)

Nicol and Macfarlane-Dick also stated that once learners have developed their self-evaluative skills to the point where they are able to engage in self-feedback, they can improve themselves even if the quality of external feedback is "impoverished" (p. 204).

In order to benefit from SA, learners themselves need to meet certain conditions. Sadler (1989) identified three such conditions: "(a) possess a concept for the *standard* (or goal, or reference level) being aimed for; (b) compare the *actual* (or current) *level of performance* with the standards; and (c) engage in appropriate *action* which leads to some closure of the gap" (p. 121). From a constructivist view of learning, such as that of Vygotsky (1934/1978), such learners' abilities are cultivated through having dialogues with and receiving assistance from their teachers or capable peers. Orsmond et al. (1997) also showed that learners' thorough understanding of the subject matter makes the SA results more useful.

In the field of L2/FL, empirical studies on the effect of SA on learning are limited. Among adult learners of French in Australia, de Saint Léger (2009) found that SA had a positive influence on their perceived fluency, vocabulary, confidence, and sense of responsibility for their own learning. Similarly, de Saint Léger and Storch (2009) found that SA had a positive influence on adult learners' willingness to communicate in an FL (e.g., perceived participation of class activities).

It is important to note, however, that many studies that examined the effect of SA on learning conceptualized learning as one-dimensional, sequential, and largely knowledge-based. Sadler (1989) reminded us that not all learning can be conceptualized as such, and stated that "the outcomes are not easily characterized as correct or incorrect, and it is more appropriate to think in terms of the quality of a students' responses or the degree of expertise than in terms of facts memorized, concepts acquired or content mastered" (p. 123). Indeed, we need more research examining the effect of SA on learning when learning is conceptualized as multidimensional, nonlinear, and nonstatic processes.

2.2.2 SA as Assessment *for* Learning Among Young Learners

When applied to young learners, empirical studies on SA from an assessment for learning orientation are scarce, particularly in the context of L2/FL. Thus, it remains unclear if most of the basic issues addressed in the previous section apply to young learners.

Figure 2 illustrates a conceptual model of SA as assessment for learning for young learners. Compared with Fig. 1, which shows a model for SA as assessment of learning, there are a few important points to note. First, in assessment for learning, SA for learning is embedded in specific social and educational contexts. Second, the emphasis is placed on a circular process of SA, which is carried out through repeated interactions between children and their teachers or peers. We can assume that the teachers or other capable peers would play greater roles in the process for young learners than they would for adult learners. Third, by having learners engage in self-reflection, SA ultimately aims to help them be self-regulated and autonomous learners. While young learners may have limited abilities to self-regulate their learning, depending on their cognitive maturity and experience (Zimmerman, 1989), children generally show substantial development in self-regulatory abilities during the preschool and primary school years (Morrison, Ponitz, & McClelland, 2010).

Before implementing SA, teachers need to (a) make sure that the assessment is consistent with the instruction and (b) choose tasks for assessment carefully. Some tasks or domains may be more difficult for children to self-evaluate than others. In Dann's (2002) case study, primary school students (ages 10–11) found it particularly
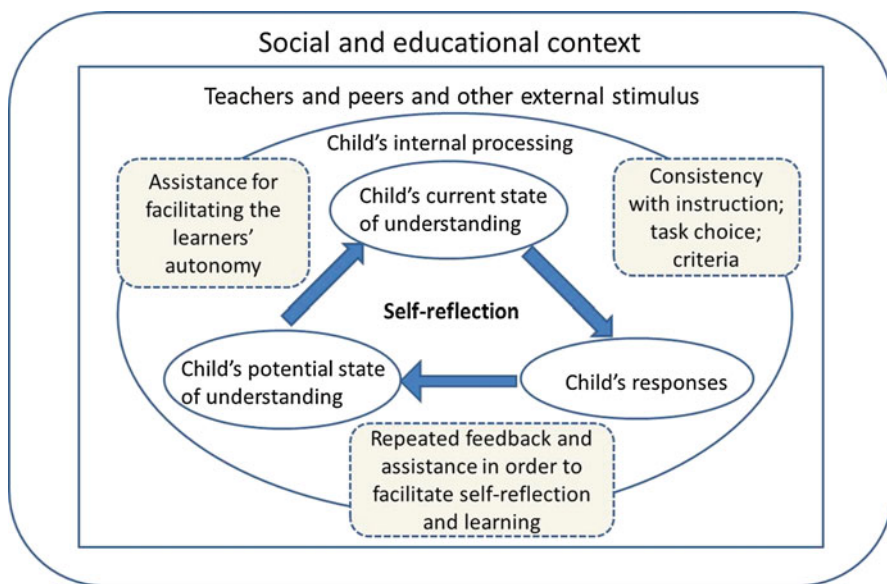
Fig. 2 The process of self-assessment for learning for young learners (Note: Components in SA described in *dotted squares* are key driving forces to facilitate learners' self-reflection processes)

difficult to assess listening compared with other domains. (Note, however, that Dann's study was conducted in a language arts context as opposed to an L2/FL context.) Unfortunately, we know very little about the kinds of tasks and performances that would be suitable for children—based on their cognitive maturity and experience—to engage in during SA.

As with adults, children need to understand the reasons for doing SA and have a clear understanding of the criteria. Children need to understand the goals and be invested in them in order to advance themselves (Torrance & Pryor, 1998). This appears to be the first hurdle to deal with, as indicated by Black et al.'s (2003) comment about young learners: "the first and most difficult task is to get students to think of their work in terms of a set of goals" (p. 49). In order to overcome this challenge, teachers may need to talk with children individually, perhaps on an ongoing basis. Although we have limited information on how children interpret the criteria for SA and make judgments using the criteria, it has been reported that children do not necessarily make judgments rationally—at least from the point of view of adults (Dann, 2002).

As suggested for adult learners, peer-assessment can help children understand the criteria better, and so it may be effective to implement peer-assessment before SA or along with SA (for a related discussion, see Hung, Samuelson, & Chen, 2016 in this volume). Dann's (2002) case study indicated that when children engaged in SA, they tended to draw on personal elements such as the *effort* that they had put into it in order to complete the work. Evaluating their peers' work (peer-assessment) seemed to help them objectify the criteria. In conducting peer-assessment with young learners, however, careful oversight is necessary. Research indicates that children who evaluate their peers' work and realize that their own progress and learning are limited compared to others are likely to lower their self-efficacy (Bandura, 1997), which in turn could negatively influence their further learning. In my studies in China (Butler, 2014, 2015), by the 8th grade (ages 13–14), some children started lowering their self-efficacy in FL learning at relatively early stages, and their level of self-efficacy turned out to be a major predictor of their FL performance.

It is also important to note that in assessment for learning, we do not necessarily adhere to the criteria in a strict sense. Instead, Dann (2002) suggested that "the priority given to pupil learning required a large degree of sensitivity in balancing the promotion of specific criteria with personal and individual factors" (p. 96–97). In other words, instead of considering the criteria to be absolute and fixed and expecting everybody to follow it uniformly, in assessment for learning the criteria should be flexible so that it can be adjusted according to the specific learning goals and needs of individual learners. Depending on the children's cognitive maturity and experience, they might even be able to actively participate in the process of developing criteria, in collaboration with their teachers.

SA can help teachers understand the gaps in a child's current state of understanding and his or her potential level of understanding (or an optimal goal for learning). It is important to note that children's judgment about their current understanding can be very different from the teachers' judgment, and thus dialogues are needed to

close the perceptional gaps between students and teachers. In order to become competent self-regulated learners, children have to develop metacognition to figure out *what they know* and *what they don't know*. As Harker (1998) stated, "only when students know the state of their own knowledge can they effectively self-direct learning to the unknown" (p. 13). And importantly, young learners are capable of monitoring their knowledge when they are provided with sufficient training. To facilitate the development of children's monitoring skills, SA should include items that capture the *process* of learning in addition to those that capture the learning outcome itself (Butler & Lee, 2010). After the gaps are understood by both the learner and the teacher, the teacher can help the learner set a goal within the zone of proximal development (ZPD, to use a Vygotskian term) and offer concrete assistance to help the learner reach the goal.

SA for learning is a recursive process. By repeating the process, SA ultimately aims to help children become self-regulated and autonomous learners. SA should be designed in such a way that learners can understand the goals of the tasks, self-reflect on their learning in relation to the goals, monitor their process of learning, and figure out what it takes to achieve the goals.

The teachers' role in the process of SA for learning is substantial. Y. G. Butler and Lee (2010) found that SA improved Korean primary school students' (ages 11–12) learning in English as well as their confidence but, importantly, the effects differed depending on individual teachers' attitudes toward assessment and their teaching context. When the teaching context was exam-driven and competitive, and if the teacher could not fully subscribe to the spirit of the assessment for learning, the effect of SA on the students' learning was limited. In other words, in order for SA to be effective, fostering a learning culture and the teachers' understanding of the assessment for learning appear to be indispensable.

3   Types of Major SAs

Various types of SA items have been developed for young learners in recent years. Some items are clearly designed for *SA of learning*, others are clearly designed for *SA for learning*, and still others can be used for either purpose, depending on the students' and teachers' needs and objectives. In this section, I examine major types of existing SAs, classifying them based on the following five dimensions and where they fall on the continua associated with those dimensions. These dimensions should be helpful for teachers and students as well as researchers when using existing SA items or developing their own items.

Domain setting
More general (open ended) ------------------------------------- More specific
Scale setting
Fewer levels   ------------------------------------- More levels
More general (open ended) ------------------------------------- More specific

Goal setting

More externally regulated	------------------------------	More self-regulated
More static	------------------------------	More dynamic

Focus of assessment

More product oriented	------------------------------	More process oriented

Method of assessment

More individual based	------------------------------	More collaborative based

3.1 Domain Setting

SAs can vary in terms of domain specifications. In Example 1, the domain is defined very generally (i.e., speaking), and the assessment focuses only on fluency. Oskarsson (1978) called this type of SA "global assessment" (p. 13). It allows us to get only a rough picture of learners' abilities.

Example 1 (Oskarsson, 1978, p. 37)[1]
SPEAKING
Put a cross in the box which corresponds to your estimated level.

☐ 10← I am completely fluent in English
☐ 9
☐ 8
☐ 7
☐ 6
☐ 5
☐ 4
☐ 3
☐ 2
☐ 1
☐ 0← I cannot speak English at all.

However, in this format, the domain can be easily defined with increasing specificity, as in examples 2 and 3: "I can ask questions in class" (Example 2) is more specific than "speaking" (Example 1), and "I can ask where someone lives" (Example 3) is even more specific (ignore the scales of these examples for the time being).

---

[1]This item did not include descriptions for each level of the scale, and was not meant for children. All other examples in this chapter were designed for young learners.

Example 2 (O'Malley & Pierce, 1996, p. 70)
I can ask questions in class

1. Not very well
2. Okay
3. Well
4. Very well

I can understand TV shows

1. Not very well
2. Okay
3. Well
4. Very well

Example 3 (CILT, European Language Portfolio, 2006, p. 11–12)[2]
Color in the speech bubbles when you can do these things.

I can play 'Simon says'

I can name colors

I can say the alphabet

I can ask where someone lives

From an assessment of learning perspective, the more concretely defined the domain specification, the more accurate the assessment, particularly among young learners. We can even set domains in a specific task that the children engaged in, as in Example 4. From an assessment for learning perspective, the assessment has to be embedded in context, as noted above; thus, contextualizing domain specificity is a critical condition for SA for learning.

---

[2] Some of the items in the European Portfolio are more or less specific. The items listed here are relatively specific.

Example 4 (Hasselgren, 2003, p. 79)

What I did in this task:

| |

How true are these? Ring round the best number
(4= true, 3= more or less true, 2= partly true, 1= not true)

I managed to say what I wanted	4	3	2	1
I understood the others	4	3	2	1
I managed to 'keep the talk going'	4	3	2	1
I knew how to pronounce words	4	3	2	1
I knew enough words & phrases	4	3	2	1
I knew enough grammar	4	3	2	1
I managed not to mix languages	4	3	2	1
I liked doing this	4	3	2	1

Things I managed to do well:

| |

Examples of words or phrases I learnt:

| |

Things I still need to work on:

| |

3.2   Scale Setting

The scale setting can be examined in two ways: (a) the number of levels and (b) the degree of specificity of each level. As I mentioned above, from the assessment of learning point of view, we don't know how many levels are optimal for young learners (i.e., yielding the most accurate responses). We can easily assume that the answer to this question depends, in part, on the degree of specificity of each scale level. Providing simple descriptions of each level, as in examples 2 and 4, may not necessarily contribute to higher accuracy. The scales still may be interpreted differently across children and, within a child, across items. It is important to make sure that children understand what each level means. While dichotomous SA items (can-do items), such as in Example 3, are increasingly popular at the primary school level, we still know very little about how children process and respond to dichotomous SA items, as discussed above.

Some SAs have detailed descriptions for each scale; such scales are often referred to as "descriptive rating scales" (Oskarsson, 1978, p. 16). In Example 5 (European Language Portfolio), each scale description corresponds to the *Common European*

Framework for Reference for Languages (CEFR, 2001). In general, the more detailed the descriptors, the easier it is for learners to respond. However, children may need assistance in comprehending the descriptors. Providing some concrete examples, as in Example 5, enhances children's comprehension of the descriptors.

Example 5 (CILT, European Language Portfolio, 2006, p. 32)
SPEAKING AND TALKING TO SOMEONE

A1 level: I can use simple phrases and sentences to describe where I live and people I know.

> Grade 1: I can say/repeat a few words and short simple phrases
> e.g., what the weather is like; greeting someone; naming classroom objects…

> Grade 2: I can answer simple questions and give basic information.
> *e.g., about the weather; where I live; whether I have brothers and sisters, or a pet…*

> Grade 3: I can ask and answer simple questions and talk about my interests
> *e.g., taking part in an interview about my area and interests; a survey about pets or favorite foods; talking to a friend about what we like to do and wear…*

From the assessment for learning point of view, scales can be useful if they are designed in such a way that learners can see the process or progress of their learning, or can identify the gaps in the current and potential levels of their learning. Scales can be set flexibly, according to individual learners' needs and learning trajectories.

3.3 Goal Setting

Goal setting refers to the process of identifying the goals of the SA, and it can be further divided into two sub-dimensions: (a) the extent to which learners have autonomy to identify the goals; and (b) the degree of flexibility with which goals can be defined. Granting autonomy and flexibility in goal setting may be a threat to the validity in the traditional assessment of learning approach, but it can be a critical feature for SA for learning, in order to help children to become autonomous and self-reflective learners. In Example 6, learners can choose from a list of predefined goals which goals they should aim for next. In Example 7, while some sample goals are listed, children can either come up with their own goals or choose their goals from the examples provided. The goals can be changed upon negotiation with the teacher.

3.4 Focus of Assessment

SAs can be designed to be more product-oriented or more process-oriented. SAs that are designed for assessment of learning are concerned mainly with what children can do (product), as exemplified in many can-do statements. Can-do items are also able to capture the degree of mastery by allowing for progressive responses (e.g., "I can do it all the time," "I can do it most of the time," "I can do it sometimes," and "I can rarely do it").

In assessment for learning, however, as we have seen already, it is critical to capture the process of learning—to make the learning process visible. We can see some attempt to capture the process in examples 4 and 7. Example 7 asks children to keep a record of their self-reflection on their performance. Upon receiving feedback from their teachers, the children can set a goal for the next class. By repeating this process and documenting it, the SA is designed to see the children's progress over time.

Example 6 (Hasselgren, 2003, p. 78)
SPOKEN INTERACTION CHECKLIST: LEVEL A2.2

	Can you <u>usually</u> do these things?[3] Use these symbols: column 1 ✓ =I think I can ✓✓ =I know I can column 2 ✓ =I aim to do this soon column 3 write the date when you've done an example of this	yes	myaim	example
1	I can understand what is said to me about everyday things if the other person speaks slowly and clearly and is helpful.			
2	I can show that I am following what people say, and can get help if I can't understand.			
3	I can say some things to be friendly when I meet or leave someone.			
4	I can do simple ask-and-answer tasks with a partner in class, using expressions we have learnt.			
5	I can ask or tell the teacher about things we are doing in class.			
:				
:				

Example 7 (Kato, n.d., p. 6)[4]

1. *Indicate today's date*
2. *Write down your own goal(s) today*
3. *Indicate your performance in ( ) using symbols below:*
     Ⓞ = super! ○ = Good Δ = Almost x = not done yet

---

[3] The underline was original. There are 12 items for each category.

[4] The original was in Japanese, and the select part was translated by the author.

4. *Write down your own reflection and submit it to your teacher*

Date	Your goal (write one or two)		Teachers' comments
		( )	
		( )	
	Your reflection		
Date	Your goal (write one or two)		Teachers' comments
		( )	
		( )	
	Your reflection		

Example goals
To try my best to engage in conversations, songs, and games in class
To speak (English) confidently
To talk to a foreign teacher
To effectively use gestures when speaking
To make eye contact to the partner when speaking
To used newly-learned words in conversation…….

3.5   Method of Assessment

SAs can be designed as an individual assessment activity or can be meant for more collaborative work. Although it is possible to use SAs for collaborative work even though they were originally meant to be carried out individually, SA items can also be designed in such a way that they invite other people's participation. This is particularly important for an assessment for learning orientation, in which it is critical to have a greater degree of collaboration (assistance from other capable individuals) in the SA process, especially during initial stages of children's SA practices. As children develop higher self-regulated skills, SAs can be conducted more independently.

4   Conclusion and Implications

Although recent policies often strongly encourage primary school language teachers to implement SA as a tool for helping children to gain greater ownership of their learning, many people continue to express concerns about the accuracy and subjectivity of SA. Such concerns, however, primarily originate from the traditional, measurement-based notion of assessment rather than learning-based notion of assessment. In addition, the age factor has not been sufficiently discussed in the previous research on SA. In this chapter, therefore, I clarified two notions of

assessment—*assessment of learning* and *assessment for learning*—while focusing on the case of SA among young learners. I also proposed five dimensions to characterize major SA items for young learners in order to help teachers and researchers to identify existing SA for use or develop SA items according to their own needs.

Research on SA among young learners of L2/FL is limited, and a number of important issues remain unresolved. With respect to *assessment of learning*, we need to uncover how item construction influences the way that children interpret and respond to items (e.g., what response bias we may observe depending on the number of scales and scale descriptors; how children use reference points; how the item wording may influence children's interpretation, etc.); and how various individual factors may influence the validity and reliability of SAs. From the *assessment for learning* point of view, we need to better understand children's *process* of engaging with SAs and its impact on their learning (e.g., how SAs enhance children's self-reflection, how both children and their teachers make inferences about the children's current and potential level of understanding, what kinds of actions were taken and their impact on children's learning, etc.). Importantly, we need more research that conceptualizes learning as a dynamic and non-linear process.

References

AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System, 32*, 407–425.

Andrade, H., Wang, X., Du, Y., & Akawi, R. (2009). Rubric-referenced assessment and self-efficacy for writing. *The Journal of Educational Research, 102*(6), 287–302.

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing, 6*(1), 14–29.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. New York: Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFC study. *Language Testing, 6*(1), 14–29.

Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning, 39*(3), 313–340.

Brantmeier, C., & Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System, 36*(3), 456–477.

Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me? Individual self-assessment by skill and level of language instruction. *System, 40*, 144–160.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5–12.

Butler, R. (2005). Competence assessment, competence, and motivation between early and middle childhood. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 202–221). New York: The Guilford Press.

Butler, Y. G. (2014). Parental factors and early English education as a foreign language: A case study in Mainland China. *Research Papers in Education, 29*(4), 410–437.

Butler, Y. G. (2015). Parental factors in the children's motivation for learning English. *Research Papers in Education, 30*(2), 164–191.

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessment among Korean elementary school students studying English. *The Modern Language Journal, 90*(4), 506–518.

Butler, Y. G., & Lee, J. (2010). The effect of self-assessment among young learners. *Language Testing, 17*(1), 1–27.

Carless, D., & Lam, R. (2014). The examined life: Perspectives of lower primary school students in Hong Kong. *Education 3–13, 42*(3), 313–329.

Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education, 32*(3), 313–330.

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*(1), 27–36.

CILT (The National Center for Languages). (2006). *European language portfolio – Junior version: Revised edition.* Retrieved from http://www.primarylanguages.org.uk/resources/assessment_and_recording/european_languages_portfolio.aspx

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

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. New York: Routledge.

Davidson, F., & Henning, G. (1985). A self-rating scale of English difficulty. *Language Testing, 2*, 164–169.

de Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals, 42*, 158–178.

de Saint Léger, D., & Storch, N. (2009). Learners' perceptions and attitudes: Implications for willingness to communicate in an L2 classroom. *System, 37*, 269–285. doi:10.1016/j.system.2009.01.001.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge, UK: Cambridge University Press.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge, UK: Cambridge University Press.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*, 21–45.

Hargan, N. (1994). Learner autonomy by remote control. *System, 22*, 455–462.

Harker, D. J. (1998). Definitions and empirical foundations. In D. J. Harker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates.

Harris, M. (1997). Self-assessment of language learning in formal settings. *System, 51*(1), 12–20.

Hasselgren, A. (2003). *Bergen 'Can-Do' project*. Retrieved from http://blog.educastur.es/portfolio/files/2008/04/bergen-can-do-project.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effect. *Language Testing, 7*(2), 174–201.

Hung, Y.-J., Samuelson, B. L., & Chen, S.-C. (2016). The relationships between peer- and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives*. New York: Springer.

Kato, Y. (n.d.). Shogakko eigo katsudo-no hyoka-no arikata: jido-ga iyokuteki-ni torikumu jikohyoka-no kufu-to hyoka-no kanten-no meikakuka [*How to evaluate English activities at*

primary school: Self-assessment that children are motivated to engage in and the clarification of evaluation criteria]. Retrieved from http://www.kyoiku-kensyu.metro.tokyo.jp/09seika/reports/files/kenkyusei/h18/k-31.pdf

Leach, L. (2012). Optional self-assessment: Some tensions and dilemmas. *Assessment & Evaluation in Higher Education, 37*(2), 137–147.

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*(4), 673–687.

MacIntyre, P., Noels, K., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*(2), 265–287.

Malabonga, V., Kenyon, D., & Carpenter, H. (2005). Self-assessment, preparation and response time on a computerized oral proficiency test. *Language Testing, 22*(1), 59–92.

Marsh, H. W. (1986). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*, 37–49.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1), 75–100.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examination. *Assessment in Education, 10*(2), 209–220.

Mihaljević Djigunović, J. (2016). Individual differences and young learners' performance on L2 speaking tests. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives*. New York: Springer.

Moritz, C. E. B. (1995). *Self-assessment of foreign language proficiency: A critical analysis of issues and a study of cognitive orientations of French learners*. Doctoral dissertation, Cornell University, New York.

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 203–224). Washington, DC: American Psychological Association.

National Council of State Supervisors for Languages. (2014). *Lingua Folio*®. Retrieved from http://www.ncssfl.org/LinguaFolio/index.php?checklists

Nicholls, J. G. (1978). The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development, 49*, 800–814.

Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.

Nicholls, J. G., & Miller, A. T. (1983). The differentiation of the concepts of difficulty and ability. *Child Development, 54*, 951–959.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.

O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York: Longman.

Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: Tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education, 22*(4), 357–368.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 25*(1), 23–38.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(1), 1–13.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 175–187). Dordrecht, the Netherlands: Kluwer.

Oskarsson, M. (1978). *Approaches to self-assessment in foreign language learning*. Oxford, UK: Pergamon Press.

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing, 19*(2), 109–131.

Piaget, J. (1926/1930). *The child's conception of the world*. New York: Harcourt, Brace & World.

Pierce, B. N., Swain, M., & Hart, D. (1993). Self-assessment, French immersion and locus of control. *Applied Linguistics, 14*(1), 25–42.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing, 15*(1), 1–19.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 145–165.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington into-Fijians. *Language Testing, 15*(1), 86–108.

Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69–75.

Stipek, D. J. (1984). Young children's performance expectations: Logical analysis or wishful thinking? In J. G. Nicholls (Ed.), *Advances in motivation and achievement: Vol. 3. The development of achievement motivation* (pp. 33–56). Greenwich, CT: JAI Press.

Sullivan, K., & Hall, C. (1997). Introducing students to self-assessment. *Assessment and Evaluation in Higher Education, 22*, 289–305.

Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment and Evaluation in Higher Education, 27*(6), 501–510.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Maidenhead, UK: Open University Press.

Vygotsky, L. S. (1934/1978). *Thought and language* (Rev. and edited by A. Kozulin). Cambridge, MA: MIT Press.

Zimmerman, B. J. (1989). Models of self-regulated learning and academic achievement. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated and academic achievement* (pp. 1–24). New York: Springer.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Zoller, U., Tsaparlis, G., Fastow, M., & Lubezky, A. (1997). Student self-assessment of higher-order cognitive skills in college science teaching. *Journal of College Science Teaching, 27*, 99–101.