Book Review

English language proficiency assessments for young learners

Editors:
Mikyung Kim Wolf, Yuko Goto Butler

Language assessment has recently attracted a great deal of attention of both researchers and practitioners, which is evidenced, among other things, by a number of well-known monographs (Brown & Abeywickrama, 2010; Coombe et al., 2012; Gordon & Rajagopalan, 2016; Gottlieb, 2006; Komorowska, 2002; Tsagari & Banerjee, 2016, to name just a few), as well as a proliferation of journals oriented towards language testing and assessment (e.g., *Language Testing, Assessing Writing, Language Assessment Quarterly, International Journal of Language Testing and Assessment,* and *Educational Assessment*). In recent years, great popularity of computers and easy access to the Internet have made it possible to move testing to a new dimension, through enabling Web-based testing (delivered via the internet) as well as computer-adaptive testing (see Krajka, 2016; Malec, 2018; Marczak et al., 2016). The use of computers has enhanced the assessment of not only target language skills and subsystems, which could be easily predicted, but also more complex constructs, such as intercultural communicative competence (Marczak & Krajka, 2014; Wilczyńska et al., 2019). *Formative assessment*, often referred to as *assessment for learning* (Black et al.,

2003), *dynamic assessment* (Shohamy, 2015) or *alternative assessment* (Alismail & McGuire, 2015; Tedesco et al., 2014) is redefining the way school teachers think about assessment, moving them away from testing towards more comprehensive ways of evaluation. At the same time, even though a great number of publications have appeared on teaching young learners, also with a focus on assessment, this does not necessarily translate into widespread awareness of these assessment issues among teachers. The question might arise, then, whether there is a need for a new publication dealing with the complex nature of language assessment, and if yes, what kind of reader to aim at, how to bridge the gap between what is available and what might be desired, and how to structure it to respond to the changing educational reality.

Few of the existing publications actually bring three important perspectives together: language assessment, young learners and standardized testing. It is in this context that the current review attempts to evaluate the newly published book *English Language Proficiency Assessments for Young Learners* edited by Mikyung Kim Wolf and Yuko Goto Butler. According to the series editors John Norris, James Purpura, Steven Ross and Xiaoming Xi (the series is Innovations in Language Learning and Assessment at ETS),[1] it "provides a comprehensive overview of issues related to assessing the language proficiency of young English second and foreign language learners," covering "a host of issues that should be of central interest to the many stakeholders in young language learner assessment, including not only test developers, but also researchers, administrators, practitioners, and policy makers" ("Series Editors' Foreword," p. viii). The purpose of this review, then, will be to see whether the book does what it claims to do, namely, discuss innovative ways of assessing young learners' English language abilities, offer stimulating ideas for future research and development work, and constitute a useful resource for a number of participants of the teaching-learning-assessment process (students, test developers, educators and researchers).

The book is subdivided into fifteen chapters, which are preceded by the series editors' foreword, acknowledgements, a list of contributors and a list of illustrations, and followed by a subject index. The chapters are arranged in five separate sections: "Introduction" (Chapter 1), "Theoretical Basis and Assessment Frameworks" (Chapters 2-5), "Empirical Studies for Validity Evidence" (Chapters 6-10), "Future Assessments and Innovations for Young Learners" (Chapters 11-14) and "Conclusion" (Chapter 15). Overall, the structure of the book is transparent and helps guide the reader from general description of young learner assessment and overall characteristics of young learners as recipients

---

[1] The fact that the book series is concerned with assessment at the Educational Testing Service (ETS), an American language services provider, is significant, as will be made clear later in the review.

of assessment, through a report on two recent ETS tests (TOEFL Junior and TOEFL Primary), design characteristics for young learner assessments, empirical studies aimed at validating the two exams mentioned, to innovative assessment designs in the fourth section. This way of structuring the book is justified and makes a presentation of the ETS offerings highly convincing, first by describing them in much detail in some chapters in Section 2, and then providing solid empirical evidence for their effectiveness, validity and reliability in Section 3.

The chapters differ considerably in their format and level of accessibility for readers. Chapters 1 and 2 are theoretical overviews of definitions, typologies, characteristics and approaches used in assessing young learners' English language ability. Well-documented in sources, covering relevant and topical literature, they are of great use for teacher trainers and practicing teachers interested in broadening their knowledge about assessing children and adolescents in a more general sense.

On the other hand, Chapters 3 and 4 have a clearly informative and reporting value, dealing with the process of defining constructs subject to testing; selecting target language use domains; selecting tasks; and designing, piloting and validating TOEFL Junior and TOEFL Primary. With little literature review and no empirical research, these two chapters are clearly structured more like technical reports, of great interest to those who want to find out the mechanics of the two tests. Also, the two chapters would be invaluable for test developers planning future standardized assessments for learners of different ages.

Chapter 5, despite its fairly general title, focuses on the interconnection between the work of ETS and US-K12 assessments, showing in particular how the evidence-centered design framework is operationalized in the latter. It also analyzes task types and test items in reference to the current context of English language education in the USA. The chapter also provides insight into challenges encountered and solutions proposed (at ETS) while developing English language proficiency assessments for U.S. K-12 students. Hence, this chapter has yet another target readership – educational authorities in different sectors of the U.S. educational system.

Chapters 6 and 7, as noted above, are of yet another kind in the sense that they constitute reports of in-depth empirical studies, with very little theoretical background and a bulk of space devoted to reporting all the steps of design and procedure, giving results, performing statistical analyses and discussing findings. Both chapters take under scrutiny the two new assessment formats introduced by ETS – TOEFL Primary (Chapter 6) and TOEFL Junior (Chapter 7).

Chapter 8, another one of a more descriptive and theoretical nature and thus devoid of empirical research, deals with the topic of the usefulness of the *Common European Framework of Reference for Languages* (CEFR) in interpreting TOEFL Junior and Primary scores. The chapter also contains a discussion of

the challenges that might be encountered by test developers wishing to use the CEFR to back up, interpret or justify their assessments.

Chapter 9, for a change, is similar in nature to Chapters 6 and 7 as it reports upon empirical research intended to verify the validity of TOEFL Junior in the process of monitoring young learners' English language proficiency. Employing a longitudinal data collection design, the authors analyzed the scores of a considerable number of students (a few thousand) who took the test more than once to prove its validity. On average, repeat test takers scored higher on the second administration. The longer the interval between testing was, the greater the score gain. Even though formidable in sample size, the study's findings have to be interpreted with caution due to the fact that as many as 65% of the participants came from one country – South Korea.

In Chapter 10 another initiative of ETS, a large-scale technology-enhanced language proficiency (TELP) project, is given as an assessment option for K-12 English language learning. Instituted in 2011, the TELP project undertook a series of prototyping studies in which a set of design principles were empirically investigated with prototype tasks integrating various technology-enhanced features (multimedia, animation, touch-screen interactivity, immediate feedback, and embedded scaffolding functions). The core of the chapter is the report of a comparative study between the linguistic performances of English-language (EL) and non-English language students, focusing on the speaking tasks. Quite interestingly, both EL and non-EL students in kindergarten, and even Grades 1 and 2, produced linguistic errors and demonstrated an insufficient command of discourse devices to retell a story or describe events coherently.

Chapter 11 opens the section devoted to future assessments and innovations for young learners. First of all, it deals with cognitive diagnostic assessment as one of the promising alternatives for language assessment. The chapter reports a study investigating how struggling young learners respond to one-on-one diagnostic assessment interventions. The final part of the chapter contains recommendations on how to design young learner assessments. As evidenced by the study, the diagnostic assessment intervention proved to be instrumental in providing mediated scaffolding for individual students whose diagnostic profiles had indicated the most needs by facilitating their metacognitive control over their reading strategy use.

In Chapter 12 computerized dynamic assessment (C-DA) is elaborated upon as an innovative alternative to traditional assessment in a young learner context. An explication of the theoretical principles behind C-DA as well as presentation of examples from two recent projects conducted with L2 learners can be found therein. In practice, C-DA is delivered in a computerized system focused on selected language skills (e.g., listening and reading abilities), with

assessments delivered online through a web-based interface written in PHP and HTML. Both the listening and reading tests in Chinese, French and Russian are comprised of multiple-choice format questions. Quite interestingly, upon completion of the test, students immediately receive an automatically generated report that includes a set of scores and a learning profile, which is what one would expect of C-DA. Thanks to the use of a learning profile, it is possible to capture both the abilities that have already developed and those that are still emerging. The strength of the chapter lies in showing the usefulness of C-DA in highly diverse contexts (L2 Spanish in elementary school vs. L2 Chinese at university level).

The final innovative approach to assessing young learners presented in the book is scenario-based assessment (SBA), tackled in Chapter 13. According to the chapter authors, SBA combines a number of techniques for delivering a set of tasks and items that provide a platform for measuring reading skills while simultaneously enabling an increase in instructional relevance of the assessment. In reading, for instance, SBA involves placing learners in authentic context of a series of interrelated activities that take the learner from more guided to more independent performance. What is important about SBA is that those activities or items provide scaffolds and guidelines to better understand and model the target performance. Thus, as evidenced in the chapter, SBAs do not only measure selected abilities but also support their development and increase learner metacognitive awareness.

The book concludes with a final chapter which sums up most of the discussions while elaborating upon challenges and future directions for young learner language assessments. Well-documented and based on the previously published studies, this chapter makes a strong case for using English language proficiency standardized tests. It shows that since in reality more and more young learners are taking exams such as TOEFL Primary and Cambridge Movers, the responsibilities of teachers, test developers and researchers are greater than ever before.

Despite a great number of strong points addressed above, however, the book has two minor shortcomings. One is the somewhat striking and disappointing confinement to the American perspective only, without coverage of other standardized exams or assessment frameworks for young learners (e.g., Cambridge Young Learner English Starters, Movers, Flyers, Cambridge PET or KET, and European Language Portfolio, to name just some). Quite clearly, the American perspective prevails – if there are references to some European or Australian language policy landmarks (such as the CEFR in Chapter 8), this serves the purpose of confronting these with ETS's TOEFL Junior and Primary, rather than dealing with them in their own right. Hence, while the book is immensely informative and argues strongly in favor of the ETS exams for young learners, its more general coverage and impact is more limited. Hence, the book would have benefitted

from a subtitle clearly indicating this US/TOEFL orientation to prevent some disillusionment of potential readers coming from non-US contexts who might not be able to see the name of the book series on the cover, where the reference to TOEFL is explicitly given.

In addition, the second minus of the book is the shift in the readership profile across the chapters. For instance, due to its high level of specificity, and close reference to U.S. curricula and standards, Chapter 5 might be less accessible to readers unfamiliar with the U.S. context to such a great extent. On the other hand, just the following chapter (Chapter 6) is a research study with high sophistication of statistical measurement, which, of course, strengthens the points made by the researchers but, at the same time, makes the chapter difficult to read for non-researchers.

To conclude, it goes without saying that offering a comprehensive, well-documented and strongly-data-driven discussion of various aspects involved in English language assessments of young learners, especially standardized testing, is the purpose that the book manages to achieve to a great extent. Especially those chapters that are not focused only on TOEFL Junior and Primary are a true asset of the book – C-DAs and SBAs are only two examples of innovations in assessment convincingly presented in the book. Because it is not mainly (or exclusively) focused on presenting and promoting the two newly implemented tests, the book assumes a very universal character, becoming highly relevant and applicable in diverse young learner contexts around the world.

Reviewed by
Jarosław Krajka
Maria Curie-Skłodowska University, Lublin, Poland
jarek.krajka@gmail.com

References

Alismail, H. A., & McGuire, P. (2015). 21st century standards and curriculum: Current research and practice. *Journal of Education and Practice*, *6*(6), 150-154.

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

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practice* (2nd ed.). Pearson Education.

Coombe, C., Davidson, P., O'Sullivan, B., & Stoynoff, S. (Eds.). (2012). *The Cambridge guide to second language assessment*. Cambridge University Press.

Gordon, E. W., & Rajagopalan, K. (2016). *The testing and learning revolution: The future of assessment in education*. Palgrave Macmillan.

Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Corwin.

Komorowska, H. (2002). *Sprawdzanie umiejętności w nauce języka obcego. Kontrola – ocena – testowanie*. Fraszka Edukacyjna.

Krajka, J. (2016). E-assessment of teaching skills: On evaluating teacher competence in blended learning. In M. Krawiec (Ed.), *New insights into learning and teaching practices* (pp. 191-206). Sprachlit.

Malec, W. (2018). *Developing web-based language tests*. Wydawnictwo KUL.

Marczak, M., & Krajka, J. (2014). Electronic assessment tools in Learning Management Systems: Perceptions of student teachers of foreign languages. In E. Smyrnova-Trybulska (Ed.), *E-learning and intercultural competence development in different countries* (pp. 265-278). Uniwersytet Śląski.

Marczak, M., Krajka, J., & Malec, W. (2016). Web-based assessment and language teachers: From Moodle to WebClass. *International Journal of Continuing Engineering Education and Lifelong Learning, 26*(1), 44-59.

Shohamy, E. (2015). From cloze to consequences and beyond: Interview by Anne Lazaraton. In A. J. Kunnan (Ed.), *Talking about language assessment: The LAQ interviews* (pp. 150-183). Routledge.

Tedesco, J. C., Opertii, R., & Amadio, M. (2014). The curriculum debate: Why it is important today. *Prospects*, *44*, 527-546.

Tsagari, D., & Banerjee, J. (Eds.). (2016). *Handbook of second language assessment*. De Gruyter Mouton.

Wilczyńska, W., Mackiewicz, M., & Krajka, J. (2019). *Komunikacja interkulturowa: Wprowadzenie*. Wydawnictwo Naukowe UAM.