

Voices in the field: An interview with Yuko Goto Butler

By David Allen

Ochanomizu University

Bio

Yuko Goto Butler is Professor of Educational Linguistics at the Graduate School of Education at the University of Pennsylvania. She is also the director of the Teaching English to Speakers of Other Languages (TESOL) program at Penn. She received her Ph.D. in Educational Psychology from Stanford University. Her research interests are primarily focused on the improvement of second/foreign language education among young learners in the U.S. and Asia as well as assessment methods for them in response to the diverse needs of an increasingly globalizing world.

Keywords: Self-assessment, formative assessment, young learners

Interview

Can you describe the different areas of work you've done in language assessment?

I'm looking at assessment for young learners (YLS). The definition of YLS is a little bit tricky, but I tend to work with children from two to 12 years old (up to the end of primary school) who are learning an additional language or languages in an instructional setting. I'm not good at handling teenagers so I'm going to stop there! When it comes to assessment, I'm primarily interested in formative assessment. I'm looking at self-assessment, task-based assessment and speaking assessment. I'm interested in second language learners of English as well as learners of other languages. Most recently I started working with YLS of Japanese as a second language.

Can you tell us how you got into the field of language assessment?

I've always been interested in students' language development; my master's thesis was on first language and bilingual acquisition among children. Since I did my graduate work in the U.S., I was working with bilingual children in the U.S. originally. After I got a job at Penn, I realized that I was the only Asian or Asian American faculty member in the School of Education. At that time, the school started having more and more Asian students and people began sending them to me! I realized that even though I'm originally from Japan, I didn't know much about Asian education. So, I became interested in looking at language learning and teaching in Asian contexts. Coincidentally at that time, many Asian countries started introducing English at the primary school level. Since I'd been working with bilingual children and second language YLS in the US, it was a natural transition for me to start looking at Asian primary school students. Soon I realized that assessment is the least researched area in primary school language education: Teachers are looking for answers in what kind of assessment they should use – there are lots of questions but few answers. That's how I began to look at assessment for YLS.

You've researched self-assessment, mainly for use with young learners of English (e.g., Butler & Lee, 2010; Butler, 2018) – what place do you think self-assessment has in language learning situations in general and specifically in Japan?

I believe that self-assessment (SA) has potential as a type of formative assessment. SA is aligned very well with the notion of learner-centered approaches to language learning. SA should promote self-regulated learning. In principle, SA should help learners develop greater autonomy and self-regulatory abilities, which in turn should help them learn better in the long run. SA is also considered less stressful for learners. This is particularly important for YLS because we don't want to give them unnecessary stress especially at the beginning of their English language learning. People talk about SA being suitable in so-called 'exam-oriented cultures' because in those cultures learners tend to have high anxiety. Since SA is considered less stressful, it's well suited for learners who have high anxiety for English language learning. Finally, SA is less constrained by time and a large class size. Considering that Japanese classes still tend to have many students in general, SA has some practical merits as well.

But SA is not the ultimate solution for everything – we have to be very careful when we implement it. There are a series of considerations that are required for successful implementation of SA. First, especially when you

administer it to YLs, items need to be constructed very carefully in order to avoid confusion. One of my studies (Butler, 2018) indicated that children's interpretations of the meanings of some of the commonly used words in SA – such as 'understand' or 'comprehend' – vary greatly. In addition, teachers often ask 'do you like English classes?', but these kinds of items do not really give you meaningful information because children tend to please adults. Moreover, how people respond to SA items is not really well understood, even among adults, not to mention among children. There are many questions unanswered regarding SA.

Teachers often ask how reliable SA is and whether children can be responsible for assessing their own abilities. There are some developmental and experiential effects – their age and the amount of experience with SA are very influential factors. So, if you just show up and say 'here's a SA, just fill this in'; children would not be able to handle it very well. It's very important to contextualize SA. It is better to ask about children's performance in a particular task immediately after they complete the task. Feedback is also very important. But currently as a profession, we do not know what kind of feedback is the most useful for children after SA.

Some researchers invite learners to create SA items. That's not something that Japanese schools have explored much – it's a really new idea for most Japanese teachers. But when I started asking children about this possibility in my study, they seemed to be very excited about it. It turned out that the children already had a very good idea about what an assessment should look like and what kinds of assessment they wanted. Children's insights can be incorporated as meaningful input when teachers construct assessment items. Inviting students as part of the assessment development process is something that teachers can explore in the future. That would help children to be more responsible for their own learning as well.

Perhaps at this point some readers may be wondering about the reliability of SA. Is it fair to assume that SA should only be used as low stakes formative assessment, not medium/high stakes summative assessment?

It can be used for both purposes. For adult learners, we know that the reliability is reasonably high in general so that my university, for example, uses SA for placement purposes for foreign language classes. The correlation with other measures of proficiency is high enough, so SA can be used reliably.

When it comes to YLs, the reliability of SA can vary depending on the children's age and their experience with SA. If a SA is not reliable for YLs, it is usually not because YLs do not have the ability to self-assess their performance, but because we, as researchers or teachers, do not administer SA in such a way that is appropriate for their developmental level. As long as we can administer SA appropriately, I think it can work well with children even for summative purposes. But, in my view, the real benefit of SA is to use it as a pedagogical tool rather than for a measurement purpose. That way of using SA is probably more powerful for learners than using it for summative purposes.

What can young learners tell us about tests?

My colleagues and I have been working on a project looking at YLs' language assessment literacy (Butler, Peng, & Lee, under review). Of course, there's substantial individual differences in to what extent they know about language assessment. But we found that the Chinese students with middle class backgrounds that we worked with knew so much about assessment based on their experiences. They could tell us much about what kinds of topics, formats, and procedures that they felt comfortable with. One of the children expressed her frustration with the current assessment practice at school saying that 'we want to have humanized assessment'! These children wanted to have more meaning-focused and communicative-based assessment. They wanted the assessment content to be practical and fun, often in a story-based format, and cognitively challenging. They also wanted the assessment format to be more accessible to children. Some of the children could articulate construct irrelevant factors in the tests that they experienced.

It was very surprising for us to find out how much they knew about assessment based on their experiences; they actually drew on their experiences of assessments from other subjects such as Chinese and science. This finding got me thinking that we – language assessment professionals – do not have much conversation with assessment professionals in other subject areas, but that might be very fruitful. We also found that the children had actually taken many tests outside of the classroom, including international proficiency tests such as Cambridge and

TOEFL tests for YLs. These tests tend to be based on more recent communicative-based approaches. The children who had been exposed to such tests can be very critical of some of the traditional assessment practices at school. It's very important for teachers to listen to YLs; they know so much more than what teachers expect.

The children also told us that the kinds of topics that we often dealt with in assessment did not seem to be very exciting for them. We must better understand what kinds of topics are most suitable for reading and listening assessment tasks for YLs. The children are looking for something exciting, meaningful, or new to them so that they can learn something new by taking a test.

What do you think is in store for the future of assessing young learners?

Technology will play a greater role in assessment for sure. YLs are growing up with digital technology from very early stages in their lives, and they increasingly use language through technology. Technology is very important as a means to assess language performance and also as a target of language use. We don't yet have a very clear idea of whether their cognitive processes have changed due to digital technology – we still need to have basic research on that – but I wouldn't be surprised if YLs' cognitive processing and preferred learning strategies are somewhat different from previous generations. Because the target language use is changing in the era of digital technology, construct definitions need to be changed as well.

Game-based assessment is also a possibility. At this point, I don't have a very clear idea about what game-based assessment should look like, but I think the elements of gamification can be incorporated in assessment. I conducted a game study with primary school students here in Japan (Butler, 2017). One of the fascinating findings was that the children incorporated unexpected elements in their game plans. In a car racing game, for example, your car suddenly breaks down towards the end of the race – this is an unexpected element. We, adults, tend to think that there should be a linear relationship between learning and scores; the more you learn, the higher score you get. That's the ground rule. But the children do not necessarily seem to subscribe to the rule when it comes to gamification. By having unexpected elements in games, students who learned most don't necessarily get the highest score. You wonder why their game plans included these unexpected elements. The children had a good reason for it. It's because such elements make games more exciting and more motivating; and so even someone who is not so good at English can still win the game. Isn't that fascinating? It sounds a little bit radical, but when it comes to formative assessment, it may be possible to incorporate some ideas from gamification which can excite and motivate children. YLs are looking for something exciting, even in assessment.

Do you have any advice for researchers in Japan who wish to conduct studies with young learners?

Conducting any research concerning children in Japan is very challenging because schools are generally not very open. But we cannot blame schools – this is partly our fault as researchers because we tend to conduct research for the sake of research. In the child developmental field, research *with* children as opposed to research *on* children, has been advocated and is very popular. Instead of just treating children as an object or subject of research, we should try to give them more agency. Some researchers even talk about having a child as a co-researcher, although admittedly it is not always easy and it's very complicated both theoretically and logistically. But the general spirit of respecting children's agency is an important concept. Researchers working with children need to keep in mind that children are not merely the subject of research. Researchers need to listen to their voices more seriously.

Related to that, I think we need to have more pedagogically oriented research for YLs, simply because we don't have much research supporting teachers' pedagogical concerns. Rod Ellis and Natsuko Shintani (Ellis & Shintani, 2013) have talked about the teachers having a very difficult time finding pedagogical solutions in SLA research. This was because researchers tend to follow their theoretical interests when conducting research and only include a paragraph of practical implications in the very last section of their paper – as if it is a peripheral issue for them. Researchers need to be more active in inviting teachers when they plan and execute research. That, in the long run, will probably help to facilitate more cooperation from the schools.

In essence, there are two types of research that need to be promoted more: More child-centered research and more pedagogically-oriented research that is accessible to teachers.

Thank you, Professor Butler!

Selected Publications

- Butler, Y. G. (2007). Factors associated with the notion that native speakers are the ideal language teachers: An examination of elementary school teachers in Japan. *JALT Journal*, 29(1), 7-40.
- Butler, Y. G. (2015). Task-based assessment for young learners: Old meets new cultures, in M. Thomas & R. Hayo (Eds.) *Contemporary task-based language teaching in Asia* (pp. 348–365). London: Bloomsbury.
- Butler, Y. G. (2016). Self-assessment of and for young learners' foreign language learning, in M. Nikolov (Ed.) *Assessing young learners of English: Global and local perspectives* (pp 291–31). New York: Springer.
- Butler, Y. G. (2017). Motivational elements of digital instructional games: A study of young L2 learners' game designs. *Language Teaching Research*, 21(6), 735-750.
- Butler, Y. G. (2018). The role of context in young learners' processes for responding to self- assessment items. *Modern Language Journal*, 102(1), 242–261.
- Butler, Y. G. (2019). Assessment of young English learners in instructional settings, in X. Gao (Ed.), *Second Handbook of English Language Teaching* (pp. 477-496). Switzerland: Springer.
- Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessment among Korean elementary school students studying English. *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.
- Butler, Y.G., Peng, X., & Lee, J. (under review). “I want humanized assessment!’: Young learners’ voices for a learner-centered approach to language assessment literacy.
- Butler, Y. G., & Zeng, W. (2014). Young foreign language learners' interactions during task-based paired assessment. *Language Assessment Quarterly*, 11, 45–75.
- Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. London: Routledge.