

Voices in the field: An interview with Nick Saville

By David Allen

Ochanomizu University

<https://doi.org/10.37546/JALTSIG.TEVAL25.1-1>

Bio

Nick Saville is currently the Director of Thought Leadership at Cambridge Assessment. He studied Linguistics at the University of Reading and holds a PhD from CRELLA (Center for Research in English Language Learning and Assessment) at the University of Bedfordshire in language test impact. He is currently Secretary General of the Association of Language Testers in Europe (ALTE). He was a founding associate editor of the journal *Language Assessment Quarterly* and has been a series editor of the *Cambridge Studies in Language Testing* (SiLT) series since 2014. He has published widely in language testing and is the co-author of *Learning Oriented Assessment: A systemic approach* with Neil Jones.

Keywords: Assessment, testing, Cambridge, Japan

Interview

This interview took place on Zoom in early March, 2021. Some light editing has been carried out, and clarifications and references have been added where necessary.

Can you give us a brief description of your career in language education and assessment?

I was reflecting the other day that it's 40 years since I started my first teaching job in Sardinia, in Italy, at the University of Cagliari. I taught there for six years, though a lot of my time there was spent doing assessments – I was having to do oral assessments regularly, and to write and deliver exams of English at different levels of a degree program. This brought home to me very soon that, back in 1981, there wasn't very much information about how to do these things. So, where people looked for such information was the existing exams, and what existed back then, in Italy particularly, were the Cambridge exams – the Proficiency and the First Certificate. As I was interested in staying in academia I went back to the UK to do a Master's degree. There I linked up with a group at Reading University doing language testing that included Don Porter, Arthur Hughes¹, and Cyril Weir². And I decided not only to do the Module on language assessment, but also my dissertation on an assessment theme. So, by 1986-7, I was already going in the direction of testing within applied linguistics. Having finished my Master's degree, I came to Japan for the first time. I stayed a couple of years working for Cambridge based in Tokyo and then joined the first Evaluation Unit which was being set up in 1989 with Mike Milanovic in the new EFL department of UCLES, as it was called back then. Over the last 30 years, I've had several roles there, becoming the Director of Research and Validation 20 years ago, which put me in the senior leadership team of the English department. My title is now Director of Research and Thought Leadership, which I've had for the last six years. It's within a much-expanded organization, going from just a few people in the English department to a large research team these days. And that's about where I am now, in the twilight of that career.

A glance at the webpage³ shows just how many researchers are now working in the validation of Cambridge Assessment English exams...

Yes, it's a massive organization and in a sense the journey of the last 30 years has moved us from a kind of cottage industry with a few experts 'hand-crafting' things to a fully integrated system model. I think one of the successes was to move research into the development phase of exams and the assessment systems and to make it clear that validation is an integrated function – not something that you have done by a few people 'in the shed'! The people that you see on the website who do the research and validation are now involved in designing, developing, and validating over the long term the propositions that we put forward. And that I think has been a great success and where we've shown leadership in Cambridge. Back in the 1980s language testing wasn't really a profession and many exam boards didn't recognize what language testing and evaluation entailed if you wanted to do it to the highest possible standard. Perhaps the Americans were ahead – there was a tradition in psychometrics of setting standards of professional conduct in assessment in the USA. But not in language assessment in the UK, and possibly even less in Europe. I think this has now changed and it was partly precipitated around the time I finished my Master's degree when folk in Cambridge at that time set up the *Cambridge-TOEFL Comparability Study* (see Bachman et al., 1995). That project brought into contrast the different traditions in assessment, and actually flagged up the strengths and weaknesses of both approaches across the Atlantic – a very strong focus on reliability and psychometric principles in the USA, whereas in the UK a very strong focus on what could be called validity, the impact on learning, and the interaction between curriculum and assessment as a design principle dating from the early days of the English exam boards.

Can you tell us about a memorable project you've taken part in?

One of the most interesting projects that I was involved in was the one I joined when I came to Japan in 1987. It put me 'on course' for the rest of my career. I joined a project to tailor the Cambridge exam system, which was a rather underdeveloped set of level-based tests, to meet the needs of Japanese learners. The aim was to introduce two new, level-based tests into what later became the Cambridge Main Suite of exams. This was a multi-partner project funded by the University of Cambridge in collaboration with the British Council and the publisher Kenkyusha that also involved setting up some flagship Cambridge English Schools in Tokyo and Kyoto. I was relatively young back then, starting a new test development project in a new country; I found that extremely interesting and I took away lots of learnings from the experience.

Before I arrived, a project had already been set up to bring Brian Heaton⁴ from the UK to work at the University of Tsukuba with Kenji Ohtomo-sensei and his team to revise the Preliminary English Test (or PET) for the Japanese context and also for it to be escalated to-exam status within the Cambridge Exam Suite as a whole. The PET at that time was a minor test, which was developed around 1980 based on the *Threshold Level*, which emerged from the Council of Europe (CoE, 2001) specifications of objectives project, and which was designed more-or-less as a classroom-based formative test.

I came after Brian, and my job was to develop the Pre-PET (what would become the Key English Test, or KET), at the Waystage level from the CoE. In fact, that's when I first started working with the level concepts from the CoE. It was also when I first met the language testers in Japan who later went on to start up and manage the Japan Language Testing Association (JLTA), and who became leaders in the field more generally. I first got to know the STEP exams because the EIKEN model was already a step-by-step, five level system of grades, very much designed for the Japanese system and to cater for what I was told were 'Japanese tastes', but very much unreformed as regards the communicative language teaching shift we'd seen emerging in the 1980s in light of the CoE movement and the *action-oriented approach* to learning, teaching and assessment.

What are some of the exciting things going on at the moment?

In the last decade, what is different is seeing how the promise of technology, both the original Ed-Tech that we've seen in the last 25 years and the new Ed-AI (Ed-Tech + AI), can provide us with the opportunity to develop more transformative, different, and better assessment and learning tasks. A project I'm currently involved in at Cambridge is the University Institute for Automated Language Teaching and Assessment, which brings together my colleagues in the research team with academics in the Computer Lab and the Engineering Department, particularly their speech unit there, together with some other linguists and neuroscientists at the university. We've created a research community in this field, and I think it is quite exciting because we've situated the learning and assessment objectives very clearly at the beginning, rather than working with computer scientists or engineers who come with different perspectives and then try to 'bolt' their ideas onto the constructs and procedures of assessment. I think we are taking a more integrated view, and we can then see how cutting-edge developments in machine learning AI, can be applied to challenges in language learning and assessment from the start. We've created that interface from the beginning, which although not unique, is unusual and is quite exciting. We've got some of the best brains in computer science working with some of the best brains in speech engineering, working with a leading team in learning and assessment. It's very productive intellectually, and likely to lead to solutions which are 'valid' – you could unpack that – designed for the purposes that they are intended from an early stage.

In the last year, I've also been on the advisory board for something called the Institute for Ethical AI in Education, which was set up in the UK under the auspices of Lord Clement-Jones, one of the peers of the realm in the House of Lords. It has brought together three leading thinkers in education in order to come up with recommendations for ethical uses of AI in education. I joined the advisory board about 15 months ago, and the final report is coming out soon⁵. It's intended to provide guidance for educators in general on ethical practices in using AI, but one of the questions I've been asking is, what are the *specific domain-related issues* that crop up in English language learning and assessment? The advantage of assessment is that you capture rich information about people to give them as feedback; the better the information you get about someone, the better the evidence you can provide on what that person can or can't do or is good at in various ways. That is, in a sense, a form of 'surveillance' – you look at people doing things – and the promise of AI is paradoxical because it allows you to surveil people better and capture more data about them. *But do people want to be surveilled? Do they want to be watched in their own homes, for example?* Getting the balance right to take advantage of the AI while at the same time having the checks and balances, the laws, regulations and social practices, is going to be the challenge: it's going to need leadership in our own field of assessment.

What are your thoughts on the current assessment reform proposal in Japan, that is, the use of four-skills tests for university admission?

Over the last 25 years, I've been talking to people working on various grant-funded projects in Japan, including some of the people in the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), concerning the reform of the curriculum. I worked with Professor Koike and his colleagues back in the early 2000s, who were investigating how the learning from the *Common European Framework of Reference* (CEFR; Council of Europe, 2001) could be transferred, adapted, and brought to Japan to introduce a more communicative approach to language education and to benchmark progression through schools from low-level ability to high-level ability to a model like the CEFR. And, as you know, the CEFR itself has been adapted, and the level system used in the revised curriculum. So, we've seen what I would call the *intended uses* of the CEFR, which is a tool for understanding the progression of learning in languages across a school system, as in the Japanese context, with the level system and the learning objectives used as a way of mapping progression across the different cycles.

In Japan's national course of study, MEXT's English curriculum documentation makes it very clear that it is based on a communicative model, that a certain number of hours are required for classroom teaching, that specific materials are to be used, and that specific skills are to be used in classroom activities. But, in the model of alignment that the CEFR promotes, the assessment is required to be in keeping with what's taught and how it's taught. Therefore, if you have moved from a knowledge-based approach to an action-oriented or more communicative approach, then the assessment needs to be based on the same thing as the teaching. The alignment between the curriculum objectives and the outcomes requires the test providers to take that on board. So, when MEXT decided to move to a four-skills approach for university entrance exams, it seemed to be in keeping with the move towards the four-skills curricula in schools (or what now would be multi-modal or integrated skills, covering the productive skills as well as the receptive ones). It was quite exciting to see because it looked like the alignment would be complete. And we all know that without the alignment you get what we would call negative washback, or negative impact, from a test which is out of alignment with the learning. That is, if you teach conversation but only test grammar, increasingly the teaching focuses on the grammar and forgets about the conversation. Unfortunately, the rowing back on the decision to implement the four skills assessments will ultimately slow down the implementation of the curriculum goals because you will not persuade parents, the community at large, teachers and learners, to move to a curriculum based on communication if the exams are based on something else. It's going to be out of alignment for sure.

In Jones and Saville (2016), Learning Oriented Assessment: A systemic approach, you describe how both classroom and large-scale assessment can work together to support learning. What insights does this approach provide for language educators, curriculum developers, and researchers in Japan?

I was looking at the book Neil and I put together and I was thinking about what we called a 'systemic approach', and one of the things we have always been worried about is the lack of alignment, or an exam system which prevents you from achieving your goals. And it says here at the end of Chapter 7,

'Depending on context there may be several ways of achieving an ecological solution – one where no element of the assessment framework is allowed to subvert the goals of learning'. (Jones & Saville, 2016, p.92)

What I try to point out, both in my thesis and in this book, is that people often only look at one bit of the system and wonder why their reform program doesn't work. They tend to ignore real people, the actual influencers, one or two key people that actually change the way things happen, not 'the government' or 'the ministry'. Or they ignore other aspects of the system for all kinds of reasons, perhaps because it's below their dignity to think about it, but it's those things that are impacting. In the UK, the impact that we were worried about which led us to thinking of things being subversive, were the league tables in schools, where exams are used to judge two things, the learner but also the effectiveness of the teaching and therefore the effectiveness of the school. In the league table model, in order to get a high rating of your school, you need to get high exam results. That meant that schools choose subjects that are easier to get good results in: chemistry is difficult, so 'let's not do chemistry'. That's not an educational goal. So, an educational goal has already been subverted by the league table. Then the teachers say to the students, 'here's what the exam looks like', not three weeks before the exam, but on day one of the course – as students go in to their GCSE course they look at the final exam paper. Well, that's not the intention. What students should look at is what they've got to learn, which is hopefully embedded in the exam and what they'll be tested on, and if they've done the learning it'll be a breeze to pass the exam. What we would call revision or exam prep is perfectly acceptable – I would call that '*seasonal washback*' – we all know when the exam season arrives, there'll be

intensive exam prep – that seems to be normal and acceptable because people shouldn't go into an assessment without knowing what's going to happen. But from day one of the course, possibly two years before students are going to take the exam, they're thinking about it – that kind of '*extensive washback*', way back into the curriculum and the pedagogy and the learning model, is subversive, right? And so, unless we can change it ecologically, not just changing one bit of it, the fundamental changes won't happen.

But it's hard to change people's minds. We started our book with a quote from John Dewey (1933, p. 29-30) to emphasize this:

‘It requires troublesome work to undertake the alternation of old beliefs’

The systemic model isn't just about having a great bunch of people in the curriculum department of MEXT, it's still troublesome to alter the beliefs of the people who implement the systems that you are trying to impact. The systemic model is an evolution of the system to bring it into alignment. And old beliefs and new beliefs have to be brought into alignment. It's a change process. The leadership, the top-down needs to impact, but the grass roots also need to be brought on board, the parents and teachers, and the learners, who need to know that what they are taught will actually be assessed. The exams tend to be the authority: if you fail to pass an exam, you might end up on a different path. Everyone knows it and parents behave accordingly: whatever the policy of government, they will try to help their kids be successful. For example, you get negative washback of this kind in China where 'tiger parents' ask teachers in primary schools if their kids can take IELTS⁶. It's a distortion of reality but of the kind that we're talking about.

There is also a problem in the language curriculum as it is currently set out. It is impossible for MOST young people to reach communicative competence beyond a very limited level (A2) just by attending the class hours in the curriculum and doing the associated homework. In other words, treating English as a normal school subject won't work. For a school subject, you have x number of hours, say three or four hours a week, plus some homework, alongside all the homework you get for the other subjects. The reason there is a *shadow education system* for language education around the world is because we know you have to connect the school learning with *learning out of school*. This creates more time, more focus, and more consistency in the learning endeavors in order to reach the higher proficiency levels. Young people who become successful learners typically have the opportunity to take advantage of this.

The work of the CEFR-J⁷ flagged up where the education system currently is with regard to international levels. And it's low, right, in Japan? It's A1, A2 level. Actually, that's not unusual around the world. The Japan system has exactly the same kind of model of putting English, or whatever language is the target, into a curriculum with x numbers of hours a week, with multiple reforms over many years, which lead to minor increments not system-wide change. And I think one of the important things we see coming out of the *Companion Volume to the CEFR* (Council of Europe, 2018) is a focus on learners as social agents, that is they use the language successfully for communication, and it's built into their vision from the start. The old-style model of knowledge-based learning, '*just another subject in the curriculum*', needs to change. That's an attitudinal change towards how languages are learned – and particularly if the language you want to learn is the *lingua franca* of the world that thrives in 'language learning friendly societies'.

Personally, it seems to me you pervert the main educational goal of language learning in the idea that learning is knowledge based. It asks people to waste time and effort on things that ultimately are relatively pointless educationally. It favors that view of the world that you can get on and do it 'in a box'. My abiding early recollections of being in Japan are of how some people are reluctant to do certain things which reveal them in ways that they feel uncomfortable with. And, of course, the concept of the CEFR is all about imperfect progression. Imperfect progression is good. The fact that you speak with all the problems is good. It's not to be shamed; but in some ways in Japan I feel there's a strongly embedded sort of cultural feeling that to reveal your weaknesses is shameful, particularly in ways that could lead to losing face or losing reputation. At the heart of this system-based approach, therefore, are the underlying *cultural norms and mores* of a society – and sometimes they run quite contrary to the learning model for speaking that we're promoting.

So, at the heart of alignment is the ecosystem of a particular context. If you've got this macro-level context, what can you do to succeed at the micro-level? This comes down to the learner interpreting and doing the things she's supposed to do to make progress and be successful. And that's embedded both within a learning model like the CEFR and within the cultural model, and the family model, which influence it. And it's this disconnect or non-convergence that comes with mixed messages or mixed influences which makes it difficult to learn languages in many countries. Because these attitudinal and cultural aspects are not only about the target language, but they also reflect the other languages that are being learned

whether it's the home language, the language of schooling, a wider regional language, or a lingua franca. They bring with them all these cultural and social practices.

In sociolinguistics, I think we're moving towards an understanding that the goal is *social practice*. What we are increasingly trying to do is ask learners to adopt the social practices for communication and not to see the thing they're doing as an academic subject. Of course, if you want to, you can study English literature, and can even do that without really knowing the language. I've met professors in Italy and Japan who don't speak English at all but who know a lot about the literature. That doesn't seem to be ideal, but it can happen. What I'm talking about with the systemic model is to see language as social practice. What you want for Japanese learners is for them to be proficient in their own language and other languages that they need in their lives – and one of them is English. It has been revealed by many studies that the Japanese need English language skills in the current phase of the evolution of this century as much as the Chinese and the Koreans do. So how can you achieve the educational goal of communicative competence without distorting it with the cultural and educational influences, which are from a past era really, about understanding how languages are learned? I would say as a coda to this, get your exam system sorted and a lot of the rest will follow. You have to win the argument about alignment of exams first. The naysayers, the conservative views, need to be put on one side so that this is given a chance, otherwise what you're doing is writing a fiction about language learning in school. You can have a wonderful policy document that shows everything is aligned, but actually the hidden (implicit) curriculum, the pedagogical practices, and the outcomes, will be based on something else. In other words, the implicit curriculum and the social practices override the stated policy.

Concerning the reform, a very recent and somewhat controversial argument is that speaking should not be assessed in high-stakes English exams for university admissions purposes in Japan because students from higher socio-economic status (SES) backgrounds will have an advantage over those from lower SES backgrounds (Butler & Iino, 2021). What are your thoughts on this issue?

I think you're asking the wrong question about SES and speaking. The question is SES full-stop and its impact on learning more generally. The digital divide has proven to be the thing that the Organization for Economic Co-operation and Development (OECD) and other big influencers have flagged up almost from day one of homeschooling during the pandemic (e.g., OECD, 2020). In April last year, when the majority of jurisdictions in the world had moved to emergency remote teaching, for some that meant Zoom all the time, and for others it meant, 'how on earth do we do this because we have no technology'? Some kids can't access anything. If you haven't got the tablet and the broadband, you're completely stuffed. The issue about hybrid models of learning have brought into focus the digital divide, and with it the engineering challenge to ensure that everyone can have broadband, and the economic challenge to make sure everyone can afford it and have a device, or devices if there's more than one child in a household. In some families you've got four kids trying to access one device – so you need four devices not one. Society needs to own this problem and speaking, that is, learning to communicate where speaking is one of a range of skills that you need, should be top of your agenda when it comes to communication, not at the bottom. You're solving the wrong problem if you say let's remove it so that it doesn't become an issue. You should be saying solve the socio-economic access problems and ensure that you can deliver the learning goals for languages as a result.

For language learners who are successful, for example the Scandinavians or the Dutch, research shows that they learn English well in school but what makes it successful is the whole of their society is *language learning friendly*. English exists in society so when they go *out of school* they have opportunities to connect what they've learned in school to something that is really useful in their society. And we need to build on this in Japan. The world has changed for English, especially in the pandemic era, because people have access to English in ways they didn't before through technology. Therefore, you can connect up what goes on in school with what goes on out of school, for conversation classes, for listening comprehension, for authentic task interaction, for game-based learning, for hobby-based learning in a wider sense, for every child. You just have to conquer the digital divide, and to ensure that every child in your country has access to right technology. That would be my answer to the systemic challenge, is to make sure that everyone is digitally connected.

Where does impact research fit in to all of this?

To understand impact, you need to understand what happens in the context where the learning and the use of assessments take place. Impact studies can't be done outside of the context; they need participation from within. You need an understanding of the situational features of the context, and what hypotheses you've got for achieving the intended impact, or the *impact by design*. For instance, if you say you want learners to reach B2 level by the end of high school by 2030, that could be an *impact hypothesis*. If you've designed the system to enable that, (i.e., you've got the CEFR levels in place and over ten years you can raise the level of achievement up one band, so instead of being at B1, a decade later it's B2), now that might be an *impact by design feature*. But how do you know what's happening? This is where washback studies, and

wider impact research come in. Such studies need to feed back into the policy making and pedagogical practice. These studies may well reveal that the intended impact is not being achieved because, for example, the implicit curriculum is still dominating. Why? Well, guess what the pinch point is? The non-aligned exam still exists – the same test that we had before the new curriculum is impacting and preventing the innovation or the evolution of the practices we had in mind. If the assessed objectives are out of alignment, the teacher will have a difficult job because they've got mixed messages: Which voice do they listen to – the voice of the curriculum planner or the voice of the exam? And that becomes the pinch point. Like this, impact by design is a systemic concept and it must be investigated through ongoing impact research.

I think impact research needs to be based on a *theory of impact* and a *theory of action* around how you can achieve the desired impact. How you find out about it is through impact research, and that's not research that's done through the odd impact study, but a program of finding out what happens over time. Therefore, impact is a *longitudinal concept*, it's an iterative concept, and it's a multimodal and mixed-methods concept in terms of research. So, it's like doing a jigsaw puzzle – you find all the pieces and you put them together, but it doesn't fall out of the box and construct itself, you have to piece it together and that takes a long time if you've got a thousand pieces. However, the longitudinal model is not what you normally get in academic studies, in which a researcher conducts a study, publishes the results and moves on to something else. It's more akin to what we in assessment call validation, which is an ongoing requirement to continually build up evidence about what happens. I think many assessment practitioners do not yet hold the view that impact is something we do, just like we do validation. The origins of washback studies can be found largely in academic research but I've tried to put it back into language assessment by saying it's equivalent to validation. Just like working out the reliability of your test every time you administer it. Impact is not something you do once at the time of writing the test, or something you give to an academic to investigate for you once and then you say, well, 'my test has positive washback', which is like saying 'my test is 0.9 reliable'. You can do that, and people do; and that's fine, but it's not job done.

Finally, is there anything exciting happening in 2021 that you wish to let us in on?

I think the most exciting thing emerging from 2021 is a reflection on how the pandemic experience can empower us to think more transformatively about the opportunities we have in front of us to improve multilingual education. The challenge of educational technology and AI is the big one that we're reflecting on. For example, I'm the secretary general of ALTE, and this year we're having our first digital symposium⁸. We will, of course, have more online and hybrid events from now on – as part of the new normal. But our digital symposium is using a virtual reality platform; so instead of taking part in the conference via Zoom and breakout rooms, you'll join a virtual reality conference venue. And you'll take part in a conference that includes sessions, plenaries, networking and coffee breaks. Needless to say, coming to the event will offer challenges. But it will certainly be the first VR conference event I, and possibly many others, will've been to. It only costs 20 euros to sign up, so virtually anyone around the world can join for the first time. So, the impact of the event, its reach, can be enhanced. But it's about balancing the benefits of this technology with other things so that the experience is a rewarding and memorable one for all. That's the challenge we have in front of us.

Thank you, Professor Saville!

Notes

¹ See the classic text, *Testing for Language Teachers*, Hughes & Hughes (2020).

² See the recent collection of essays in honor and memory of Professor Weir, *Lessons and Legacy: A Tribute to Professor Cyril J Weir (1950–2018)*, edited by Taylor and Saville (2020).

³ See <https://www.cambridgeenglish.org/research-and-validation/meet-the-team/>

⁴ See *Writing English Tests*, Heaton (1975)

⁵ See <https://www.buckingham.ac.uk/research-the-institute-for-ethical-ai-in-education/>

⁶ International English Language Testing System (IELTS) is a test of English for academic purposes aimed that assesses the language ability of candidates who wish to study in an English-medium tertiary institution. Therefore, for obvious reasons it is not suitable as an assessment for primary/elementary school children.

⁷ The CEFR-J is a modified version of the CEFR that is adapted for the Japanese context (see Negishi & Tono, 2016; Tono, 2013). Its development has led directly to revisions to the CEFR itself (see Council of Europe, 2018), particularly regarding lower levels of ability and the distinction between upper and lower abilities within specific levels (e.g., A2.1 and A2.2).

⁸ See <https://www.alte.org/Digital-Symposium-2021>. Editor's note: I attended this conference over three days and thought it was a huge success. Not only was the format a completely new experience, but also the presentations were excellent. A conference to remember!

References

- Bachman, L., Davidson, F., Ryan, K., & Choi, I. C. (1995). *An investigation into the comparability of two tests of English as a Foreign Language*. Studies in Language Testing, 1. Cambridge University Press.
- Butler, Y. G., & Iino, M. (2021). Fairness in College Entrance Exams in Japan and the Planned Use of External Tests in English, in B. Lateigne, C. Coombe, & J. D. Brown (Eds.), *Challenges in Language Testing Around the World*. Springer.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/16802fc1bf>
- Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment* (Companion volume with new descriptors). <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Dewey, J. (1933). *How we think*. <https://www.archive.org/details/howwethink000838mbp>
- Heaton, J. B. (1975). *Writing English language tests: A practical guide for teachers of English as a second or foreign language*. Longman.
- Hughes, A., & Hughes, J. (2020). *Testing for Language Teachers*. Cambridge University Press.
- Jones, N., & Saville, N. (2016). *Learning Oriented Assessment: A systemic approach*. Studies in Language Testing, 45. Cambridge University Press.
- Negishi, M., & Tono, Y. (2016). An update on the CEFR-J project and its impact on English language education in Japan. In C. Docherty & F. Barker (Eds.), *Language Assessment for Multilingualism: Proceedings of the ALTE Paris Conference, April 2014*. Studies in Language Testing, 44. (pp. 113-133). Cambridge University Press.
- OECD. (2020). *OECD policy responses to coronavirus (COVID-19): Combatting COVID-19's effect on children*. <http://www.oecd.org/coronavirus/policy-responses/combating-covid-19-s-effect-on-children-2e1f3b2f/>
- Taylor, L., & Saville, N. (Eds.) (2020). *Lessons and Legacy: A Tribute to Professor Cyril J Weir (1950–2018)*. Studies in Language Testing, 50. Cambridge University Press. <https://www.cambridgeenglish.org/Images/582822-silt-volume-50.pdf>
- Tono, Y. (Ed.). (2013). *Eigo tōtatsu-do shihyō CEFR-J gaidobukku* [The CEFR-J handbook]. Taishukan Shoten.