

INTERVIEW WITH THE EXPERT, DR STEFAN O'GRADY, CONDUCTED BY DOROTA LUBIŃSKA-PYLA

Dorota Lubińska-Pyla (D.L.P.): At our Centre, we teach languages to undergraduate students in their first or, more often, second year. They enrol in our groups after years of exposure to a very traditional, if not fossilized, state education system obsessed with testing, which can be summed up by the Polish phrase ‘the three Z’s rule’ (‘Zakuć, Zdać, Zapomnieć’) – ‘Cram it, pass it, forget it!’. And even if we refer to it half-jokingly, it still reflects the sad reality that most students have to endure, and most teachers have to labour in. So, my first question is: How can we, university language teachers, approach assessment and testing so students are encouraged to take responsibility for their learning?

Stefan O’Grady (S.O.): This is quite common in the culture of high stakes testing. Essentially, the process is preparation for a summative snapshot. The results of the test represent an image of a person at a particular time. The problem is the snapshot also largely reflects the personal circumstances of the test taker at that specific point in time. So yes, the test may generate information about an individual’s language competence, but it also reflects the fact that they didn’t sleep well the night before (because they were worried about the exam), perhaps the fact that their tooth had been hurting for the past few days, maybe a bad break up, etc.

These limitations are especially concerning when the learning focus is dictated by people that do not know or interact with students. There’s no opportunity to personalise the content and make it relevant to the students’ own needs and interests. So, the answer is to look for the alternatives, which would seem to be frequent assessment rather than singular, low stakes rather than high stakes, and developing assessments where students have the opportunity to personalise the content and make it relevant to themselves. This seems to be the key here.

(D.L.P.): Formative assessment and summative assessment are two concepts that are always mentioned in the context of giving feedback and testing.

Could you elaborate on them from the practical perspective of a university language teacher? And how do you balance the two in your courses?

(S.O.): Summative assessment refers to the assessment that takes place at the end of a course of study and is often formal and high stake. In contrast, formative assessment refers to the techniques and strategies teachers apply to obtain evidence about their students' progress during a course.

Feedback provided on summative assessments is often not acted upon. And that's because there is seldom an opportunity to do so. The course has finished, and the students have moved on. There are also often strict time constraints involved in summative assessment, which prevent teachers from providing rich and detailed feedback. This is one of the reasons people recognise the importance of formative assessment. Formative assessment allows students to receive feedback on their progress which they can later apply. So, for feedback to be helpful it needs to encourage learning that can be directly applied in the future.

Carless (2007) talks about feedback as feedforward in learning-oriented assessment, and I think this is a helpful distinction for people getting involved in assessment. Feedforward needs to be made with the goal of improving the student's completion of the next task. Feedforward should be frequent and continuous and directly related to the learning goals. It also needs to be specific to target the skill and not the person (e.g. "try harder" or "you've done well on this assessment"). I think this is a trap lots of teachers fall into. While these kinds of comments may help boost the student's self-esteem, they don't help them next time they have to complete a similar task. So they don't effectively feedforward.

(D.L.P.): Could you share any experiences or best practices related to providing effective formative assessment to students? And at this point, I would also like to ask about correction. What are the most common 'sins' of correction?

(S.O.): One of the courses I am involved in, students work on a long essay and submit a series of drafts for one-on-one tutorials. During the tutorials, the tutor focusses on language use and writing mechanics as well as content. The essay progresses and develops throughout the course, and we keep a detailed record of this development as well as encouraging the student to keep a record of the feedforward they receive in the tutorials. We use the virtual learning environment which stores everything for us and can be consulted later to see the development the student has made. This is a writing task but could easily translate to speaking, featuring e.g., recordings in preparation for a seminar or a presentation. The regular tutorials encourage personalised and continuous feedforward ultimately leading up to the completion of a relevant task.

Regarding the sins, as I mentioned before, I think a common one is to focus on the student instead of on the task. While feedforward can help to motivate, the student's motivation is likely to reflect their personal circumstances at the time so comments like "try harder" don't really help them understand their strengths and weaknesses in the skill. Instead, the feedforward should be applicable next time they attempt the task or one similar. It needs to target processes that evidently aren't as smooth or advanced as they could be. Another related sin is to not actually give an opportunity to try the task again after the correction. The student needs to be given a chance to apply the feedforward. Obviously, excessive corrections can overload the student, and they don't know what to prioritise. Asking students what they think they should receive feedback on can be helpful in this regard e.g., please prioritise my pronunciation. So, in essence, the comments should target a specific aspect of the task completion, and this can be negotiated with the student.

(D.L.P.): What innovative assessment methods, going beyond the tired end-of-semester test, have you found effective in promoting student learning and engagement?

(S.O.): I think methods like projects and portfolios, applications of task-based approaches like task-based language assessment linked to needs analysis can help to promote positive washback and motivate the students. These are helpful methods but obviously introduce subjectivity in scoring, so they require rather training, examiner standardisation and occasionally depending on the context, statistical adjustment of scores. So, you need the resources in place to be able to meet all of these demands, or the innovation will backfire. This is one of the major reasons for developing assessment literacy among teachers. There's often a lot at stake for the students so innovation needs to be approached carefully.

(D.L.P.): Can you recommend any resources or books that might be useful for those of us who would like to start working with portfolios, myself included?

(S.O.): For this, a first step I would recommend would be getting familiar with the Common European Framework of Reference for Languages, the CEFR. The Council of Europe webpages have lots of information about the European Language Portfolio. There are also links to case studies and activities. These are linked to the reference level descriptors and so they can be a very helpful resource for language instructors looking at ways to sequence activities and tasks to lead to a series of language objectives.¹

¹ <https://www.coe.int/en/web/portfolio/overview-of-cefr-related-scales> [accessed: 17 April 2024].

(D.L.P.): What are some of the misconceptions about assessment at the tertiary level that you have come up against in your everyday practice as a lecturer?

(S.O.): I think people often regard assessment as an obligatory element at the end of a period of study. During course design, assessment is often an afterthought and that's a major weakness. It should be the first consideration. Following principles of constructive alignment, course design should start with the identification of important goals and objectives, skills that will be valuable to the students, then proceed to the development of assessment tasks that engage these goals, and then finally learning activities. This way, the objectives, the evaluation of the objectives and the daily activities will present more of a coherent whole. Development often goes in the opposite direction and that is a problem from my perspective.

(D.L.P.): To date peer assessment hasn't been very popular in our teacher-centred academic culture. Besides, from what I have seen, students do not feel very comfortable doing it and they are either too harsh, much more than I would ever be, or they are too lenient and influenced by peer pressure or friendships. Is it worth attempting? And if so, how do we set it up without wasting precious class time?

(S.O.): If we apply some ideas from Vygotsky, for it to succeed, peer assessment requires one person to be more informed than the other on a particular aspect of the task. So, for peer assessment to work, tasks need to be developed to create that dynamic. In other words, the task developer needs to set out to design a task to facilitate peer feedback rather than attempt to use a task that was originally intended for teacher feedback. For example, tasks can be designed where each student naturally has a unique perspective on the content; their personal experiences of learning are all different, they have different interests and different experiences to draw from, so the teacher should work with this.

Any feedback can be helpful, even if it's something like one student checking the spelling for another in a dictionary and in the process reading their work, this can lead to discussion. But the students need training too. So, these principles about feedforward like focus on the task not the person, understanding of learning goals, etc. can help the students develop their ability to provide peer feedback/feedforward.

(D.L.P.): Have you any practical tips for us, language teachers, on how to prepare our students for peer assessment? Are there any guidebooks/resources/websites that you recommend using so we feel more prepared for this task?

(S.O.): Again, I would recommend utilising the CEFR for this purpose. I think the tendency with this document among many teachers is to regard it as a series of proficiency levels that help with benchmarking tests, but there's much more to it. Using the CEFR self-assessment grids, for example, learners can complete tasks that teachers design to engage the relevant skills described in the descriptors and reflect in groups on how they completed the tasks, what they found challenging, areas to improve upon, etc. These are available in a range of languages, including Polish, so even if the learners are beginners, this is something that can be set up relatively easily.

So, for example, a teacher could identify a relevant can-do statement e.g., "Can follow in outline short, simple social exchanges, conducted very slowly and clearly" (p. 56). They would then develop a task that engages this ability, perhaps using a smart phone to record themselves with a friend or colleague talking in the target language and having a social exchange that is both slow and clear (perhaps the friend takes on the role of a learner, so the exchange would naturally be slow). The students would then listen and complete some kind of task, a summary task, or comprehension questions, gap fill, etc. Then students would use the completed task and the grid to discuss, possibly in the L1, their experiences of completing the task and how well their peers have completed the activity.

(D.L.P.): I have tried some self-reflection questionnaires with my online students during the pandemic. The results were mixed; some of them took those questionnaires very seriously whereas some almost totally ignored them. What role does self-assessment play in your teaching approach, and how do you encourage students to reflect on their own learning?

(S.O.): Developing this ability is crucial. Students won't always have a teacher with them to help and correct their mistakes. The ability to self-assess successfully is really the ability in the skill of interest. Understanding the criteria for successful performance is also understanding how to perform successfully. I think for this reason, evaluative judgment should be a priority in course design, prioritising growing independence and socialisation into a community of practice or discourse community. Again, this requires training. It requires understanding the goals and objectives of the course and why these are in fact valuable. Clearly, assessment criteria and the way these are communicated to the students plays a big role in this process.

(D.L.P.): And finally, what are some recommendable online tools, perhaps even AI-aided, which you have used and can vouch for?

(S.O.): From a teacher's perspective the growth of generative AI is creating a lot of opportunities for automatic test design and development, or at least AI

supported design and development. I've had success using Chat GPT and Gemini to develop texts for assessment tasks e.g., input flooding tasks and comprehension questions featuring a particular language form can be created in seconds. I've also used a chatbot as a Socratic opponent in assessment tasks e.g., students argue for the best teaching methodology for a particular context against the AI.

We're currently looking at ways for AI to support the provision of feedback/feedforward on assessments. The trouble here is determining whether the AI can interpret the assessment purpose and the construct in the same way the course designers do, because if it doesn't, any feedback it provides is likely to be flawed and potentially misleading.

I think overall, AI is really changing the way we think about assessment. The academic essay, for example, has been a cornerstone of university assessment for centuries but it's so easy for students to completely hand this over to AI now. Clearly, people found ways to avoid writing essays historically too, but it would likely be expensive or involve some kind of conspiracy or coercion. It's very easy now to generate plausible writing immediately for free. The problem is, this is often discussed as wrong for moral reasons rather than for reasons of validity. Thinking of this in terms of validity, a GenAI produced essay simply doesn't give you sufficient grounds to make a decision about the extent to which a student has met the learning objectives. It prevents us from making informed decisions; that's the problem it causes. More broadly, we probably need a new vehicle for the demonstration of learning objectives that promotes the kind of thinking and application required in essay writing. It will be interesting to see how this develops in the future.

(D.L.P.): Do you think AI tools like Chat GPT can be safely used for creating criteria or rubrics needed to assess students while doing more task-based projects with them for example Academic Reading Circles?

(S.O.): Yes, I do think AI tools can be used to help create effective criteria and rubrics for language tasks, but the awareness and ability to recognise whether the AI produced instrument is likely to be beneficial or not needs to be developed first. Rubrics can be designed for different purposes and for different audiences, and it's the developer's responsibility to understand why they are developing the instrument and who they are developing it for. This involves understanding the context, the requirements and expectations, the characteristics of the users as well as the characteristics of the learners, their purposes for learning the language, and how they demonstrate language proficiency. Performance descriptors in rubrics should be realistic and reflect the ways in which the learners use the language to meet task demands. Clearly, this is much more achievable when you are familiar with the individuals you are developing the instrument to describe. This

means that experienced teachers are often perfectly placed to develop criteria, with appropriate training. AI can support this process, for example by suggesting ways of organising the criteria to increase common application and accessibility, but identifying and generating relevant descriptors can't be completely handed over to the AI, because it's likely to write descriptors that are irrelevant or even unachievable for this specific group of learners. This may lead to negative washback.

(D.L.P.): To sum up, what are the main principles of good feedback?

(S.O.): The first principle is that feedback should feedforward. It should help the students complete the same or a similar task in the future. It should be frequent and continuous and there should be a clear opportunity to apply the feedforward. It also needs to raise student's awareness about their own ability to self-assess so it should target the student's own judgments about their work e.g., why did I decide to use a past tense here? This is so that they can formulate means that might help them self-assess in the future, independently. Because that's the ultimate goal; autonomous language learners.

Bibliography

- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332> [accessed: 17 April 2024].
- Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge.
- Overview of CEFR-related scales. <https://www.coe.int/en/web/portfolio/overview-of-cefr-related-scales> [accessed: 17 April 2024].



Stefan O'Grady is a lecturer in TESOL and academic English at the University of St Andrews. His research interests are in test development and validation with a specific focus on university admissions assessment. He has published in leading journals and is the recipient of the 2024 TOEFL New Scholar Award.

Stefan O'Grady jest wykładowcą TESOL i akademickiego języka angielskiego na Uniwersytecie w St Andrews w Wielkiej Brytanii. Jego zainteresowania badawcze koncentrują się na tworzeniu i walidacji testów, ze szczególnym uwzględnieniem oceniania w procesie rekrutacji na studia. Stefan O'Grady publikował w czołowych czasopismach naukowych i jest laureatem nagrody TOEFL New Scholar Award 2024.