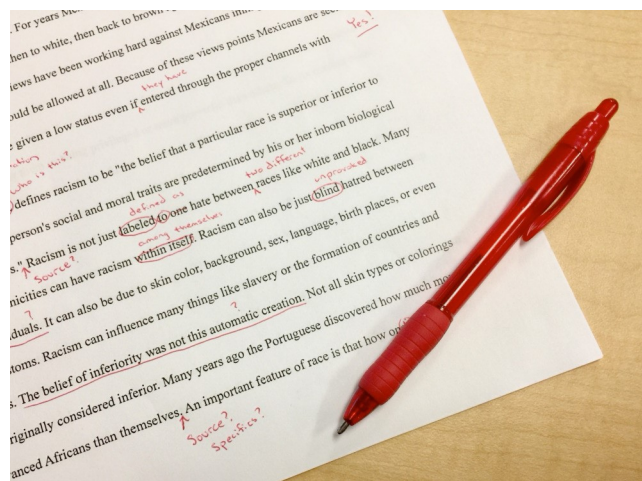


The Teacher Development SIG in Dialogue: An Interview with the Testing and Evaluation SIG's James Sick

Sam Morris

Kanda University of International Studies

This interview is the second of three from our annual forum at JALT's PanSIG conference, with 2019's edition titled - Understanding Teacher Efficacy: SIGs in Dialogue. In this forum, three Teacher Development (TD) SIG officers each conducted short 20-minute interviews with invited representatives from the Mind, Brain, and Education (BRAIN), Extensive Reading (ER), and Testing and Evaluation (TEVAL) SIGs, in order to collaboratively and dialogically explore the topic of teacher efficacy together; one of the central themes of the 2019 JALT national conference. The forum provided the chance, through formal interaction, for a group of SIGs to not only cooperatively construct a richer understanding of teacher efficacy from multiple perspectives, but to strengthen an awareness of each other's group's activities. In this interview, TD SIG treasurer Sam Morris asks questions to James Sick, an officer of the TEVAL SIG about the relationships between teacher efficacy and language assessment. The following exchange has been adapted from the original spoken version, along with additional written communication. It has been edited for readability and clarity purposes.



Sam: Thank you very much for your time today to talk about this important topic, James. I was very interested in talking to your SIG about this issue because assessment is one of the most important parts of our job, but it's also something that many teachers may lack confidence in. I wonder if you could begin by introducing the work that your SIG does?

James: The testing SIG, or TEVAL for short, was founded in 1995 by a group of teachers within JALT who were concerned with the university entrance exams and the negative washback they were perceived to have. The entrance exams at that time primarily focused on discrete grammar points and answering multiple-choice questions about very long and difficult reading passages. There was little in the way of communication that was addressed by these tests.

The SIG quickly widened in scope to address the more challenging question of 'how do we appropriately assess communicative English in the classroom?' Our goals are to support all forms of assessment and evaluation at both the institutional and teacher level. We are interested in all methods of assessing individual language performance, group outcomes, program impact, and washback.

Some of the activities we've carried out in the past few years include providing workshops by experienced assessors at chapter events, sponsoring experts in assessment as featured speakers at JALT conferences, publishing a biannual journal featuring both research and practical articles on assessment, and recently by compiling a full-length book from 20 years of a regular column on testing and statistics by JD Brown. The book, *Statistics Corner*, is sent as a complimentary gift to all new SIG members. We also publish a biannual newsletter and have run workshops at JALT national and other conferences related to assessment. I became the SIG coordinator in 2008 and have held that role since. Besides being the SIG coordinator, I teach courses in testing and research statistics at Temple University Japan and a few other places.

Sam: As part of the TEVAL mission statement, you note that "tests and assessments serve many purposes" could you briefly define the different kinds of assessment and testing that language teachers typically employ during their work.

James: There recently has been some debate within the profession about what exactly constitutes an assessment. One possible response to that question is that an assessment is nothing more than an activity that is assigned a letter or number when it's completed. At PanSIG 2018 in Tokyo, TEVAL sponsored Jerry Talandis, who spoke about how to turn standard speaking activities into speaking assessments.

With that definition in mind, I would say that individual teachers typically create two categories of assessment: traditional tests and performance assessments. Traditional tests target specific skills or knowledge sets, such as reading or listening comprehension, vocabulary, or grammar points. These assessments are usually comprised of a series of test items scored as right or wrong and then summed to create a single score. They are used as entrance exams, placement tests, or course achievement tests, but short quizzes or other ongoing formative assessments can also fall in that category.

The second type, performance assessments, assign a subjective grade to a speaking or writing performance. This category covers a wide range of activities, such as graded speeches and presentations, one-on-one interviews with a teacher, role-plays with other students, and observed group discussions, as well as graded written

work. More and more, these types of performance assessments are being used not only for classroom assessments, but for placement or program exit standards. And as most JALT members are probably aware, there is now an official plan to incorporate speaking and writing performance as part of university entrance exams. Considering the history of getting the listening section into the center test, my prediction is that this is going to happen eventually, but it's going to be delayed year by year. It's going to take some time. Initially, only a few universities will do it, but I predict that this is not only going to have an enormous impact on teaching but also on teachers learning how to construct and grade these kinds of performances in unbiased and fair ways.

Sam: To jump off from that final point, one issue I think that many teachers struggle with is the inherent subjectivity of grading, for example, written papers. Do you have any advice on how teachers might grade their students in an unbiased way?

James: I think the key here is the two Cs: collaborate and calibrate. Work with other teachers to create rubrics that have clear, detailed descriptors of the language or performance aspects you will grade in your assessment. If you're in a program that has a standardized program-wide assessment, that might be easier, but I'd say even if you don't, find a teacher teaching a similar course with a similar kind of assessment and share your rubrics. Talk about them, and even look for some opportunities to see how well you are calibrated. Can these rubrics be used to give reasonably consistent scores to observe performances? Even if you're working independently and are not officially required to have uniform assessments, sharing ideas, and working together to design your rubrics and then testing whether they are sufficient to give consistent grades to similar work can take you a long way toward relieving some of that stress and creating less biased assessment.

I also think there's a whole lot to learn about rubric design. I've done several workshops on rubrics and found that many people use what I would call a quasi-rubric. It has criteria like pronunciation, body language, and content, and it has levels like excellent, very good, good, and needs work, but these cells are just empty. Those are not genuine rubrics. The empty cells need to be filled in with detailed descriptors of what can be observed. And those are then used to calibrate teachers. These kinds of rubrics can be very tricky to write, and really, we need two versions. We need one for the teachers and another simplified version for the students, because they need to get meaningful feedback.

I might also add that subjective assessments will never be completely free of bias or error. You're better off to just accept that and take steps to minimize it, not eliminate it.

Sam: I think creating fair assessments can require statistical knowledge. What are some easy ways for

teachers to develop their understanding of statistical techniques?

James: Firstly, I'd like to say that statistical knowledge is definitely helpful, but it's not the only kind of knowledge that is necessary to be competent at assessment. For most people, I would say basic item writing and rubric design techniques come first and are more important than statistical analysis.

Second, I'd say that if you want to analyze your multiple-choice tests (and you should because you can learn an enormous amount about your test making skills, your students, and whether your goals have been achieved), the statistics are not that complicated. The two most important test statistics, which are item facility and item discrimination, are not that difficult to compute or understand. Item facility is simply the percentage of your test takers who answered an item or a set of items correctly. Instead of asking, "How many test problems did individual A answer correctly?" you are asking, "How many individuals answered test problem 1 correctly?" And when test items are tied to curricular goals, that is how you make sound judgments about the degree to which your objectives are being achieved.



Sam: What kind of knowledge and skills do you think a teacher who is 'confident in testing and evaluation' needs to have, and how do you think teachers can develop or achieve those skills?

James: To me, the clearest way to identify a teacher who has become strong and confident in testing and evaluation from a teacher who has not, is that the former will be able to quickly and easily identify the connection between specific curriculum goals and specific assessments. Expert teachers enter a course with explicit objectives that they hope to accomplish. These objectives may be general in nature, but an experienced teacher will usually be able to state them in terms of observable outcomes. An experienced assessor will additionally be able to state which assessments will be used to determine the extent to which an outcome has been achieved.

In other words, self-efficacy in testing entails competent syllabus design as well as testing skills. A good teacher begins by saying, "These are five specific goals I hope to achieve in the next 15 weeks, and this is how I am going to determine whether I and my students have been successful." A good assessor will answer the questions, "How am I going to know if these goals have been achieved?" and "What kind of principled observation or assessment or test am I going to make to confirm that they learnt what we set out to learn?"

Sam: Finally, to bring the discussion back to the TEVAL SIG, how would you say that your SIG can support teachers who lack confidence in assessment? What resources or events do you offer that you feel would be of use?

James: One of the most important recommendations I can recommend is the book I mentioned earlier that we have published called *Statistics Corner*. For 20 years, JD Brown has been answering questions about testing and statistics in non-technical, teacher-friendly language in SHIKEN, the TEVAL SIG Newsletter. In 2017, we compiled all those questions and answers into the book. You can purchase that book on Amazon for US\$30, or if you join the TEVAL SIG, you'll receive a complimentary copy. Short of taking a testing course in a master's program, I think reading that book is probably the best

way to get yourself up to speed on both the practical and statistical side of testing.

We have also conducted workshops cosponsored by chapters in the past. We hope to organize more of those so watch for them on the TEVAL website, ELT Calendar, or other JALT information sources. We often do workshops or forums at the PanSIG or the JALT national conference. In the past we've had workshops on basic item writing skills, easy item analysis using Excel, using and interpreting Rasch analysis, and other topics. If these interest you, drop by the events, or drop by the SIG table to talk to us.

Sam: Thank you very much for your time, it's been very interesting and useful to speak with you.

Bios

Sam Morris is the current treasurer of the TD SIG, a senior lecturer in the English Language Institute at Kanda University of International Studies, and a PhD candidate at the University of Leicester. He is interested broadly in the role that emotions play in second language teaching and acquisition, as well as the affective dimension of teacher psychology. His principal focus is on the situated emotion regulation that teachers employ during their work.

James Sick is the coordinator of the JALT Testing and Evaluation (TEVAL) SIG.



Follow us on Twitter

Follow the conversation (and participate) on Twitter by following our handle. We share information on upcoming SIG events, reports of past workshops and sessions, JALT activities, language teaching conferences throughout Asia, special member requests, and more...

[@jalt@_TD_SIG](https://twitter.com/jalt_TD_SIG)