Assessing for student success: An interview with Dr. Liying Cheng

Sachi Oshima[1] and Jeffrey Martin[2]
oshima.s@mc.cgu.ac.jp
martinjpsla@gmail.com
[1] *Chuo Gakuin University*
[2] *Daito Bunka University*

Dr. Liying Cheng is Professor and Dean of the School of Education at the City University of Macau. Before taking up this role, she served as Professor and Director of the Assessment and Evaluation Group at the Faculty of Education, Queen's University, Ontario, Canada. Dr. Cheng is internationally recognized for her research on washback illustrating the global impact of large-scale testing on instruction, the relationship between assessment and instruction, and research on how to align assessment practices with student success and language development.

The Testing and Evaluation SIG of JALT was honored to help sponsor Dr. Cheng as a plenary speaker at the 50th JALT International Conference, held at the Shizuoka Granship in November 2024. In a joint opening plenary with Dr. Andy Curtis, she reflected on JALT's 50 years of contributions to language education and shared insights on Opportunity, Diversity, and Excellence. Dr. Cheng also conducted a workshop titled *Assessing for Student Success*, where she emphasized the importance of assessment as a process for supporting learning through alignment, fairness, and engaging students in discussions about their learning progress.

The following interview was conducted online and via email correspondence following the conference. Building on the insights that she shared, we explored a range of topics, including her initial interest in language assessment, the CARE framework for academic acculturation in supporting student success (Cheng, 2020), types and purposes of assessments, and the application of these ideas in diverse L2 learning contexts.

Dr. Cheng, thank you for taking the time to speak with us. It was a pleasure to hear your plenary and to be in your workshop at JALT2024. We'd like to know more about the ideas you shared at the conference.

It's my pleasure to do this interview with the Testing and Evaluation SIG of JALT following the 50th JALT International Conference! I'd like to thank you for making the time in your busy schedule to conduct the interview.

At the JALT2024 conference in Shizuoka, Japan, you mentioned differences between your formative education in Beijing, China, and your graduate education in the UK, and that they brought your attention to language assessment. For readers of Shiken, what were some of the formative experiences that helped shape your subsequent work in this area?

It's quite a big question for me to answer. And between the experience I had in China and the experience I had in the UK doing my master's degree, I have to say that most of the experience was really in what differential teaching and learning were like. I came from a very traditional testing background in China, where testing was used for selection, but it was not the case in the UK then. For example, instead of testing as a selection purpose, part of my master's research involved a test called TEEP (Teaching English for Educational Purposes; https://www.reading.ac.uk/isli/english-language-tests/teep). It is designed to show whether a student has a level of English sufficient for degree-level study in UK higher education. The test was designed for international students who wanted to study there and would start what was called the Pre-sessional Course. This was instead of a pass/fail situation but based on a criterion: depending on their mark, they would do a shorter or longer number of weeks in the pre-sessional courses, which were like ESL courses before students actually joined study courses in subject areas.

Studying in the UK was a shock for me. I think the biggest shock was that I had been learning through Chinese before then, but in the UK, I had to learn in English (the medium of instruction). The way of learning there and then was also different. At that time, the program was very tough. I was at the University of Reading, which had one of the top three applied linguistics programs in the UK then. Back at that time, an example of the strictness was that, while we were required to do group presentations, projects, and tests outside the classes, for the final course paper, we had to submit it by 4 o'clock to the departmental secretary. If we didn't submit it on time, we basically failed that course.

That academic testing situation was very tense for me. It reminded me of what I experienced in China around that time. The assessment field has really changed over the past 30 years. There was a major reform in education in 1998 when Black and Wiliam (1998, 2010) published their papers entitled *Inside the Black Box*. Their work was the breakthrough that brought the core value of assessment to support teaching and learning. Before that, testing done almost everywhere was primarily for selection.

In my initial education in China, testing was conducted at the national scale. At the University of Reading, assessment was really focused on measuring how much you had learned. It was based on criterion-referenced standards, meaning if you didn't meet the criteria, you would fail. It was a type of selection within the university's academic program. This was 1994, when fairness was defined differently. For example, nobody had personal computers. We had to go to a certain place to use shared resources. For example, I remember using a dot matrix printer at another department and running back to the other half of campus to meet the 4 o'clock deadline.

You mentioned your framework "CARE: Key to Academic Acculturation" in your plenary speech. Can you first elaborate on the components of this framework? C: Compassion; A: Acquisition; R: Respect; E: Evaluation. Then, could you describe differences in how this framework might be applicable in ESL contexts and EFL contexts? We thought of asking you about potential differences in applying CARE since you have taught both in Canada and in Macau. Many Shiken readers work in the EFL contexts of Japan.

Compassion is the ability to put ourselves into others' shoes. And so that means having the ability to feel what other people feel. This is crucial when it comes to the differences in teaching and learning in a diverse higher education context. During my time in Canada, for example, I was teaching a course which had both master's students and PhD students. We had domestic students who were born and grew up in Canada, international students, and short-term exchange students. They represent a range of diverse views regarding teaching and learning. In addition, Canada's situation is complex as it was not obvious from the outlooking surface as to who is a domestic or international student in a classroom due to immigration, so the differences in teaching and learning were embedded.

I use the term *Acquisition* to refer to both intentional and unintentional learning. I adopt this as a second language acquisition term, as used by Stephen Krashen (1981). *Respect* is really about how we deal with differences among others, particularly given the different experiences we all have. The *Evaluation* piece involves making a judgment, which I think is closely related to critical thinking skills. Critical thinking, to me, is the ability to be able to solve new problems, especially in the higher education context, where students come from many different cultural and educational backgrounds. The presence of diverse groups of people coming from different backgrounds brought a great deal of tension to the Canadian higher education classroom. So, I developed *CARE: Key to Academic Acculturation* (Cheng, 2020; Cheng & Fox, 2008).

You asked about the application of the four components of my CARE model by teachers in EFL and ESL contexts. I don't think there's a difference when we use the model between an EFL class or an ESL class. The idea of Compassion should work for all human beings. Compassion, Acquisition, Respect, and Evaluation should happen in and applied to any learning and teaching context. We need to support our learners in developing those four aspects of competences and skills, because it is our foundation to be a better human being.

About *Acculturation*, I think it is the context where I developed CARE. Acculturation is a psychological term. In Canada, acculturation is used to refer to the acculturation on both sides, i.e., I understand you as the host, and you understand me as the newcomer. We all have our own culture—ways of thinking and doing. So, acculturation is about understanding each other. When we talk about academic study, acculturation is also important. If I'm an international or immigrant student and want to succeed in Canada, I need to learn what good teaching or good learning looks like, because that's how my teachers and instructors assess me.

But at the same time, as instructors, we need to understand the backgrounds of our students. For instance, I ran a workshop in Canada on how to pronounce Chinese names, which is a big deal because Chinese names are often mispronounced. It's very difficult for English speakers. I think we had a chance to really help each other. And the more we understood each other, the better we could appreciate our shared humanity. Respecting every culture is important. We need to celebrate the differences, yet we need to recognize how challenging and hard it is to teach and learn in a diverse context.

Acculturation is the larger context where CARE can be used. I actually talk about CARE in Macau, even when all my students are Chinese students. There are still differences. Acculturation in this context means that students and teachers may have different perspectives but can come to a shared understanding.

In your workshop titled "Assessment for Student Success," you introduced three types of assessment. Could you describe them and their differences? For those interested in distinguishing and implementing these ideas, what research or teaching resources would you suggest?

Assessment *for* Learning refers to the process of seeking and interpreting evidence for use by students and their teachers to decide where the students are in their learning process, where they need to go, and how best to get there. Assessment *of* Learning refers to the assessment that happens after learning has occurred and aims to determine whether or not learning has taken place. These assessments are used to make statements about students' learning status at a particular point in time. Assessment *as* Learning occurs when students reflect on and monitor their progress to inform their future learning goals. It is regularly occurring, both formal and informal, and helps students take responsibility for their past, current, and future learning (Cheng & Fox, 2017).

You also asked whether self-assessment, peer-assessment, and peer feedback can be included in Assessment *as* Learning. Yes, but peer-assessment can also fall under Assessment *for* Learning, because that's what we do in the classroom on a daily basis. We do peer-assessments; we do self-assessments as well. This is also true in your teaching context. For example, in your own article (Martin & Oshima, 2024), you discussed how students can engage well in course objectives and enhance their sense of agency when they can help lead the learning of their peers. The practical tips you gave for productive and receptive L2 coursework also demonstrated adaptability in different teaching contexts.

These three types of assessment happen in our classrooms all the time, yet at different stages of the instruction. Assessment *of* Learning tends to happen at a specific period of time toward the end of an instruction period. So, it's not something that always occurs. Assessment *of* Learning tends to involve summative tasks, combination of assignments, a quiz, a test, a project, a presentation, or maybe a book report—something that evaluates cumulative learning. As we wrote in our book (Cheng & Fox, 2017), the idea is that students themselves understand their learning situation, their starting point, and how to improve themselves. These three assessment types overlap. In practice, they are all part of teaching and learning.

The definition of these three assessment types of assessment are not my own terms. The field of assessment has been working on these ideas for some time. The initial two terms were *summative* assessment and *formative* assessment (Scriven, 1967). These have been used in our field for a while, and now we've started to use these three terms: Assessment *for* Learning, Assessment *of* Learning, and Assessment *as* Learning.

In your workshop, you also mentioned aspects of the purposes and uses of instruction, diagnosis, and grading in assessment. How do you advise classroom teachers to use their limited time most effectively to best achieve these important aspects of language assessment?

Most teachers engage with these three purposes—instruction, diagnosis, and grading. Among these, instruction is the broader category, as we do that in our classrooms all the time. Diagnosis and grading, however, often require more attention, especially in terms of teacher professional development.

Diagnosis comes with feedback, which I think is very important. We are all teachers, and we know that providing effective feedback can be a challenge. Sometimes, we default to say, "Good job, you've done a good job." But most of the research, especially in the past 10 years on feedback, shows that the key for effective feedback is to focus on the task rather than the person. For example, if the task is essay writing, we should focus on specific objectives of the writing. Let's say the goal is to teach students to use linking words or improve transitions between paragraphs and sentences. You can center your feedback around that. Feedback is a technique, and if you work with a class of students, you don't have to give feedback on everything at once. For instance, in one session, you might focus on coherence—how ideas are linked and flow together. Next time, you might focus on the use of tense. This is especially challenging for Chinese students because, in Chinese, verbs do not change to reflect tense. In terms of diagnosis, for me, it's about knowing the students and identifying how to support them in moving forward in their learning.

Grading is also very important. I've been working on grading issues. I received a major SSHRC grant (the Social Sciences and Humanities Research Council of Canada). We spent five years investigating grading issues across countries (see Cheng et al., 2020). Grading is different from feedback; while feedback is part of the process, grading represents the end of an assessment period. That's when you give a mark. Often, this mark is a combination of various factors. For example, teachers often incorporate multiple components into grades. It's crucial that teachers ask themselves: "Does this mark—let's say, 76—accurately reflect the student's learning? Or does it also reflect something else, like their classroom behavior or participation?" Sometimes, the school or departmental policies require teachers to factor in attendance or participation when assigning grades. But grading is a much broader instructional issue, as schools and universities often have guidelines dictating how to grade. Grading is also a high-stakes process, as it directly affects students' wellbeing. Your decision can

cause students to feel rewarded for their hard work over the semester or, conversely, feel frustrated, upset, or demoralized. That said, teachers cannot give everyone an "A" mark, because that would defeat the purpose of grading. Grading is ultimately about decision-making, which Fox and I detailed in our 2017 book as a process that teachers need to approach carefully.

I believe all teachers need to learn more about feedback and grading. If we have the opportunity to offer more workshops, I would dedicate an hour to feedback and another to grading. More research has been published on grading than on feedback. Together with colleagues, we wrote a paper (Cheng et al., 2020) specifically about the grading differences in Canada and China. This was part of our comparative study that provided a detailed examination at grading practices across countries. This study was not specific to language classrooms but general education classrooms. It included data from teachers, students and parents. The data were very rich, though messy, as they reflected how teachers (and other stakeholders) have approached grading in various contexts.

For our last question, we were curious about how your concepts apply, or have been applied, with learners of other languages. Chinese, for example, is a foreign language studied by many people in other cultures in the world. What are your thoughts about how researchers, teachers, and language programs of other languages are implementing these ideas?

I don't see any differences in how to apply the concepts across different language teaching and learning contexts. I think what we've talked about mostly in terms of assessment applies to all classrooms. I work with teachers on the basic principles of how to use assessment to support students' learning and teachers' instructional practices. If we look at the field of education, we ask teachers to think about their unique classroom contexts and how they're going to apply these theories to their classrooms (Cheng, 2023; Cheng & Fox, 2017). I always start by explaining my own teaching and learning context when I talk about assessment. Then, I move on to provide guiding theories and practices for teachers to consider. I emphasize that teachers need to think about how they can adapt these ideas to their own classrooms, because every classroom is different. I always say, "Teaching is about context." Context is really important, and we need to trust teachers to make their own decisions based on their individual classrooms. They might say, "Okay, this concept might not work in my classroom, so I need to adapt it," or "I need a different pedagogy because my students or my subject matter are different."

To give an example of the importance of the context, respect is universal, but the way we express it can differ across cultural contexts. I'll give you one example. I spend a lot of time with my Chinese students talking about the difference between humility and humbleness. This humbleness is a huge concept, but humility is not talked about much in that context. I couldn't really find the corresponding word in Chinese to represent humility. So, I had conversations with my students, and we debated academically about which Chinese words might represent humility in a Chinese context. These debates were fascinating.

Another point that needs to be highlighted for all teachers is that our prior experiences with assessment are deeply influential and often shape our teaching. Research shows that our past learning experiences are one of the strongest predictors of how we teach. That's why one of the first things we emphasize in Canadian teacher education is reflective practice. We encourage teacher candidates to reflect critically—not just on what worked or didn't work in the past, but also on why something worked and whether it can be improved. Colleagues and I have conducted many qualitative studies about this. An example is Wu et al. (2014), where two of my students and I wrote about the relationship between assessment and motivation. The relationship is foundational in the teaching profession, not only in Canada, where I worked for 20 years, but also in other countries like China.

About the teaching of languages around the world other than English, we also have ethical guidelines for assessment and testing practices such as those from the International Language Testing Association (ILTA; https://www.iltaonline.com/page/CodeofEthics). At the end of the day, we have to remember that education systems are products of their societies. The world is far more internationalized than before, but each education system reflects the society it operates within. That's why compassion is so important. Compassion is about embracing differences. Whether we are teaching in Canada, China, or Japan, compassion allows us to see beyond cultural differences and focus on the shared human values embedded in education.

Thank you for sharing your valuable insights in this interview. We appreciate your contributions to the JALT2024 conference and we hope that this conversation adds meaningfully to the ongoing dialogue on assessment and language teaching.

Declaration of competing interests:

No conflicts of interest were reported.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Cheng, L. (2020). *Compassion, acquisition, respect, evaluation (CARE): Key to academic acculturation*. Human Rights and Equality Office, Queen's University. https://www.queensu.ca/hreo/together-we-are/compassion-acquisition-respect-evaluation-care-key-academic-acculturation/

Cheng, L. (2023). *Language classroom assessment* (2nd ed.). TESOL Press.

Cheng, L., DeLuca, C., Braund, H., Yan, W., & Rasooli, A. (2020). Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools. *Studies in Educational Evaluation*, *67*, 100928. https://doi.org/10.1016/j.stueduc.2020.100928

Cheng, L., & Fox, J. (2008). Towards a better understanding of academic acculturation: Second language students in Canadian universities. *The Canadian Modern Language Review*, *65*(2), 307–333. https://doi.org/10.3138/cmlr.65.2.307

Cheng, L., & Fox, J. (2017). *Assessment in the language classroom: Teachers supporting student learning*. Palgrave McMillan.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Pergamon.

Martin, J., & Oshima, S. (2024). Using peer-led evaluation in productive and receptive L2 coursework. In B. Lacy, R. P. Lege, & P. Ferguson (Eds.), *Growth mindset in language education* (pp. 54–64). JALT. https://doi.org/10.37546/JALTPCP2023-07

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Rand McNally.

Wu, Y., Cheng, L., & Bettney, E. (2014). Assessment and motivation: Perspectives from teacher candidates. In S. V. Nuland (Ed.), *Conference Proceedings of the 58th International Council on Education for Teaching (ICET) World Assembly: Moving forward in curriculum, pedagogy and leadership* (pp. 358–371). Oshawa, Canada: University of Ontario Institute of Technology. https://www.icet4u.org/docs/Canada_2014.pdf