Interview

Language Testing as a Profession: An Interview with Yan Jin

Lin Zhang
Shanghai Jiao Tong University, China

Abstract

This is an interview with Dr. Yan Jin, professor of linguistics and applied linguistics at Shanghai Jiao Tong University and Chair of the National College English Testing Committee in China. In this interview, Dr. Yan Jin discusses how she started her career as a language tester and why she has dedicated herself to the language testing profession. She looks back on her 30 or so years of experience working with the large-scale, high-stakes College English Test (CET) developed by language testers in China, and introduces the important measures that have been taken to maintain the high quality of the CET and to adapt the test to meet changing social needs. She stresses the need for language testing professionals to consider their social responsibilities and take care of the social dimension of language testing. She also comments on the recent development of the national standards of English proficiency in China and the relationship between the language standards and English curricula and assessments. She further describes her active involvement in community services in the field of language testing and assessment which she believes has contributed to her professional career as a language tester. At the end of the interview, she provides valuable advice, based on her testing research experience, to young testing researchers and prospective researchers in the field, which can be very helpful on their path to pursuing language testing as a profession.

Keywords

Language testing practitioner, large-scale language testing, College English Test, fairness and social justice, social responsibility

Dr. Yan Jin is a professor of linguistics and applied linguistics at the School of Foreign Languages, Shanghai Jiao Tong University. Her research interests include the development and validation of large-scale, high-stakes language assessments. She is currently Chair of the National College English Testing Committee in China and President of the Asian Association for Language Assessment. She is co-editor-in-chief of the Springer open-access journal *Language Testing in Asia* and is also on the editorial boards of *Language Testing*, *Language Assessment Quarterly*, *International Journal of Computer-Assisted Language Learning*, and a number of academic journals published in mainland China.

Address: 800 Dongchuan Road, Minhang District, Shanghai, 200240, China
E-mail: zhang_lin@sjtu.edu.cn

Thank you, Professor Jin, for accepting my invitation to do this interview. As your MA and doctoral student and your colleague for about 20 years, I have been particularly impressed by your passion for and dedication to your work in the field of language testing and assessment over the past three decades. And it seems to me that your passion in language assessment has never waned. So, in this interview, I hope you will share with us your experiences as a language tester. Could you please first tell us how you developed an interest in language testing?

Thank you, Dr. Zhang. Let me answer your question about my early interest in language testing from my undergraduate studies. I started my undergraduate work in 1984 at Shanghai Jiao Tong University. The major I chose at that time was English for science and technology. In our curriculum we had not only courses in foreign languages, linguistics, and literature, but also courses in advanced mathematics, physics, and so on. Mathematics and physics were taught in English by professors who received education overseas or graduated from Shanghai Saint John's University. They spoke beautiful English and taught those content courses in English. So, it was a kind of what we now call CBI or CLIL, that is, content-based instruction or content and language integrated learning. Shanghai Jiao Tong University is known for its science and engineering courses, so I also took advantage of this and chose some optional courses. I remember that I studied discrete mathematics in the mathematics department and PASCAL language in the computer science department. This kind of blended learning helped me to develop a better understanding of the interdisciplinary nature of language testing and to prepare myself for a long career in language testing.

The other important reason for me to study and work in language testing was the development of the CET (CET is short for College English Test, a nationwide large-scale language test for tertiary-level English language learners in China). As we know, the idea of creating a national test to promote the then newly released National College English Teaching Syllabus was initiated by a group of professors in the mid-1980s. In 1987, the CET4 was started (The CET is a test series consisting of written and spoken English tests at two levels: Band 4 [CET4 and CET-SET4] for learners at a lower level of English proficiency and Band 6 [CET6 and CET-SET6] for those at a higher level). When I graduated in 1988, I decided to continue with my MA and I had a great interest in the CET program. Therefore, I conducted some research on the item types of the CET. The topic of my MA thesis was on the comparison of subjective and objective items for testing reading comprehension. Professor Hongzhang Liu, my MA supervisor, was the director of the CET Test Center II, in charge of test administration in the East China area. At that time, the British Council sent experts to help us train CET item writers. Through participating in these activities, I was intrigued by the techniques of producing "good" items, items that were at an appropriate difficulty level and would be answered correctly by higher-level test takers but incorrectly by those at a lower level. These early experiences of studying and working for the CET influenced my decision to continue my career in the field of language testing and assessment.

As a professor of linguistics and applied linguistics at Shanghai Jiao Tong University, you have been a teacher of English language and applied linguistics. Many students have completed their MA or PhD study under your supervision. As I mentioned, I am very honored to be one of them. In what aspects do you find your teaching career, particularly your work as a supervisor, rewarding?

After I finished my MA, I started to work at the CET Administration Office. I was responsible for

taking notes during item review sessions, editing test papers, and so on. Meanwhile, I taught what was called "public English" at that time to non-English majors. I had the opportunity to gain some firsthand experience in testing and assessment. For example, I got to know how students viewed language testing and how they coped with language tests. I also found that students from different disciplines had different learning styles. For example, students from humanities departments and students majoring in science and engineering had different styles and strategies of English learning. After I finished my doctoral studies, I began to teach English majors and supervise MA and doctoral students.

The most rewarding part of being a supervisor is that we have to keep learning new things. Each student has their own research interest, background, and capabilities, so we should find a suitable research focus for each of them. To achieve this, we must have sufficient breadth and depth of the knowledge of the area we are researching. Being a supervisor therefore gives us not only responsibilities but also opportunities to keep abreast of developments in the field. One lesson I learnt from supervising students is the importance of mutual trust: students devote their most precious years to going through a short journey with us and we should try our best to help them enjoy every moment of the journey. They should feel not only the pressure of finishing their dissertations and publishing in academic journals, but also the pleasure of doing research and the joy of growing up. They may not necessarily feel this way while working on their publications and dissertations, but I hope the experience of graduate study can be recalled with warmth and joy.

Aside from being a language teacher, you are also known as a language tester and are currently Chair of the National College English Testing Committee (henceforth CET Committee). As we know, the CET is a test of a super large scale and has a current annual test population of nearly 20 million. Can you tell us something about your work experience relevant to this test? And what are the major challenges of working for a large-scale language test?

As I said before, one of the main reasons for my interest in the field of language testing and assessment is my early experience of studying and working for the CET. I should say that my career development has been mainly motivated by the on-going development and reform of the CET. The testing program was proposed and designed to meet social needs. When China started to open to the outside world in the late 1970s, there was a higher requirement for university graduates' foreign language abilities. So, a group of professors and scholars, led by Professor Huizhong Yang from Shanghai Jiao Tong University, my doctoral supervisor, designed and started the CET in the mid-1980s. Since I finished my MA, I have been working for the CET, first as an admin office member, then director of the admin office, and Chair of the CET Committee in 2004. So I consider myself more of a practitioner than a researcher, or you could say a practisearcher, of language testing and assessment.

The biggest challenge of working for a large-scale, high-stakes language test, in my view, is being able to work under great pressure—pressure from various stakeholders. First, it is the pressure from test developers ourselves. We hope to develop and validate the test to meet technical quality requirements. Language testing is a very fast-growing area, so we need to ensure the test's validity and fairness on an ongoing basis. The second source of pressure comes from teachers and students. A major concern of the developer of a large-scale test is whether the test will bring about positive or negative washback on teaching and learning. And there is pressure from society—whether the test has ensured fairness and/or social justice. We are also under great pressure from test management departments. With a scale of about 20 million test takers a year, it is no exaggeration to say that the test's security and quality could have an impact on social stability.

The CET has been around for over 30 years and it is seldom criticized for its design or the quality of its items or tasks. Also, very few test users have doubts about the test results. Could you tell us what measures you take to maintain the quality of the test?

The top priority for language testers is to ensure the quality of the test. In the early 1990s, after the CET had been administered for several years, the CET Committee started a validation study in collaboration with the British Council. Our understanding of test validity during that period of time was based on early validity theories which viewed validity as a characteristic of a test: validity is the extent to which a test measures what it is supposed to measure. And reliability was seen as distinct from and an unnecessary condition for validity. So our task at that time was to look for evidence supporting the three main types of validities, called the earlier trio of validities: content validity, criterion-related validity, and construct validity. Establishing a test's validity was considered the sole responsibility of test developers and researchers. This was the first comprehensive validation study for a large-scale language test in China. Based on the results of this study in the 1990s, the design of the CET was revised, and a lot of measures were taken to improve the quality of the CET.

As a test for educational purposes, the assessment criteria of the CET have always been closely aligned to the College English teaching syllabus or curricular requirements. Every time when the curricular requirements were revised, the test would be reformed to meet the changing needs of the curriculum. For example, in the mid-1990s, there was a need to improve college students' English speaking ability to meet the new curricular requirement of English speaking, so the CET Committee designed and launched the CET Spoken English Test (CET-SET) in 1999. The test took a face-to-face oral interview format, in which three or four test takers and two examiners formed a test group to complete a number of monologic and interactive tasks.

The other major aspect or theme of the reform of the CET is the use of technology. The CET Committee has been trying their best to develop and implement technology-enhanced language testing systems. For example, in the early 2000s, to cope with the increasing test population, an online CET scoring system was developed and has been used ever since, which has greatly improved scoring efficiency and scoring reliability. When the test population of the CET-SET increased, the computer-based CET-SET was developed in 2013 and replaced the face-to-face format in 2015. So technology-enhanced testing has always been a major focus of the CET test development and reform.

Apart from your work for the CET in China, you are also active in serving the language testing community in various ways. You were member-at-large of the International Language Testing Association (ILTA) and have chaired or participated in various award committees. Now you are President of the Asian Association for Language Assessment (AALA). In what ways do you think these various strands of work have contributed to your professional career as a language tester?

Having worked for the CET for three decades, I have gained a better understanding of what it means to be a language tester. The top priority is certainly to ensure the technical quality of a test. In recent years, I began to pay more attention to the social dimension of large-scale testing programs. In China, learners of all educational levels are required to take English tests for a variety of high-stakes decisions, including, for example, admission to junior or senior high schools, admission to colleges or universities, job

applications, and promotions. Decisions based on these tests can have especially important consequences for both individuals and society. As language testers, we need to constantly keep in mind our responsibility for ensuring test fairness and social justice and to collaborate with test users to make sure test uses are appropriate and misuses are avoided. A language tester, in my view, should not only enjoy the beauty of the science and art of language testing, but also be fully aware of our social responsibilities.

I started to get involved in various community services in the hope of sharing our experiences of developing and using high-stakes language tests with language testers from other parts of the world. In the past few years, I was a member of the ILTA Public Engagement Committee (PEC). Together with three other professors, Cathie Elder from Australia, Caroline Turner from Canada, and Nick Saville from the UK, we proposed various measures to engage more people in language assessment activities. One of the proposals was to set up the ILTA Advocacy and Public Engagement Award, which was approved by the ILTA Executive Board. I chaired the first Award Committee in 2020. In the process, I got a much deeper understanding of the social role of language testers and the ways in which language testers can help build a better world.

The recent outbreak of COVID-19 has changed tremendously the way language is taught, learnt, and tested. Language testing communities have been working hard to help people around the world cope with the increasing demand for online assessment. ILTA launched its first series of webinars and I gave a talk on online assessment of speaking, using the computer-based CET-SET as an example.

In Asia, the CET Committee started the first academic forum on English language testing in 1998, which was named the Academic Forum on English Language Testing in Asia (AFELTA). Through this forum, major language testing organizations in Asia meet every year to discuss the new developments and challenges facing language testing in the Asian context, where the number of English language learners is the biggest compared with other parts of the world. The Asian Association for Language Assessment (AALA) was started in 2014. At the first AALA conference hosted by Zhejiang University, there were 89 participants. In recent years, 200-300 language testing researchers participate in the annual conference. I hope the AALA will continue to grow and help promote the professionalism of language testing in Asia.

As you have just mentioned, the recent pandemic has had a tremendous impact on the way language is tested. We know that while almost all paper-based tests were called off, online testing is becoming more prevalent. How do you view this trend? Has the CET made any attempts in this respect?

Since the outbreak of the pandemic, all the big tests, both international and national ones, have been suspended. The May CET-SET tests and the June CET tests have been rescheduled. The CET-SET and TOEFL iBT are online tests, but test takers need to go to a test center to take the tests, so they were also suspended. Overseas universities are in desperate need of tests for admission purposes. There are a few at-home online commercial testing programs, but at-home tests are not normally trusted for making high-stakes decisions due to users' concern with test security.

As a test of a super large scale, the use of technology is essential for the CET. I mentioned previously that, since the early 2000s, the CET has been making full use of ICT (information and communications technology) to improve the efficiency of test administration, delivery, and scoring. In 2007, the internet-based CET was developed by the CET Committee in collaboration with an IT company. Speaking was one component of the internet-based CET, instead of a separate test. Although the test has not been put into operational use, the team has accumulated a lot of experience in developing and delivering online tests.

Going online, however, is only the first step for large-scale language testing. As I have mentioned, the computer-based CET-SET has replaced the face-to-face test since 2015. The speaking test now has about 1 million test takers. The biggest obstacle to accommodating the increasing number of test takers is scoring. Online scoring surely improves efficiency for the CET writing and translation tests, but for speaking, scoring of several million test takers will be a mission impossible. There are issues with qualified raters and the issue of time it takes to score the performances. So the CET Committee has been working on developing automated scoring systems for its constructed-response items, such as writing, translation, and speaking. The aim is to use automated scoring systems as a second rater or a check rater. Validation studies are in progress and have yielded some preliminary findings.

In the past few years, as one of the main contributors, you have participated in the development of the China's Standards of English Language Ability (CSE), and you were also in charge of the development of the speaking scales of the CSE. We know that as the first full-range English proficiency scale designed for Chinese learners and users of English, the CSE is intended to provide a set of national standards of English proficiency to promote English teaching, learning, and assessment in China. Why do you think the Chinese government initiated and supported the CSE project? And how do you view the relationship between language standards, assessment goals, and curricular requirements?

In 2017, Professor Guoxing Yu from Bristol University and I co-edited a special issue for the Springer open-access journal *Language Testing in Asia*. The special issue addresses various issues involved in the development of the CSE. One of the articles in that special issue co-authored by Professor Zunmin Wu from Beijing Normal University, Professor Charles Alderson from Lancaster University, Ms. Weiwei Song from Foreign Language Teaching and Research Press, and myself, discusses the macro- and micro-political challenges facing the development and implementation of the national standards.

In our view, the main purpose for the Chinese government to initiate and support the development of the CSE is to improve the internal consistency and coherence of curricular requirements at various educational stages. Externally, it is hoped that the development of the CSE will improve the transparency of China's policies and practices of English language education and better prepare Chinese people to live and work in an increasingly globalized world.

Talking about the challenges, in a book chapter I wrote for a new book titled *English Language Proficiency Testing in Asia: A New Paradigm Bridging Global and Local Contexts*, published in 2019 by Routledge, I called the CSE "a new kid on the block" because of the close and long-standing relationship between curricular requirements and English language tests in China's educational system. I pointed out in the chapter that the implementation of the CSE may experience challenges of coordinating and negotiating the views of different stakeholders and challenges from practitioners who may pay lip service to the new standards.

To effectively implement the CSE, reform of English language education is inevitable, and this will meet resistance, both active and passive. At the macro-political level, the CSE is expected to introduce a degree of commonality in terms of the national-level educational policy in English teaching, learning, and assessment. The "consistency" or "commonality" intended by the CSE, however, is likely to meet resistance from policymakers in charge of English language education at different educational levels. The relationship between the CSE and English curricula and assessments will not be established if the governmental departments and curricula or assessment developers are unwilling to align their curricular requirements and assessment criteria to the new standards. At the micro-political level, practitioners of

English language education in China may find it difficult to understand and apply the proficiency levels and can-do descriptors in their routine practices.

The CSE is certainly an important achievement in English language education in China. However, there is a long way to go before the CSE could have a significant impact on English language education in China.

In some people's eyes, developing a language test is as easy as pie and does not deserve to be treated as a scientific discipline. How would you respond to this view? In your view, what are the most effective ways for improving stakeholders' assessment literacy?

In the view of many lay people, language testers are nothing but item writers. In the case of selected-response items, such as multiple-choice questions, the job of an item writer is nothing but producing one correct answer and several "tricky" distractors. This view is widely shared among stakeholders including educational administrators, teachers, and end-users of language tests such as admission officers. I do not think we should blame these people for being naive. In my view, it is the responsibility of language testing professionals to educate stakeholders and make it known that language testing is a highly professional area that needs professional training.

In 2010, I published an article in *Language Testing* reporting the results of a survey I conducted for investigating the types of courses provided in universities for pre-service language testers. I received answers such as preparing students for taking TOEFL, IELTS, CET, etc. Test preparation was considered as a course related to language testing. Of course, test preparation is an area of research in language testing, but a language tester's job is to measure learners' language ability not their test-taking strategy. Test-taking strategy use in most cases is construct-irrelevant and should be controlled or avoided.

Language teachers are not born language testers. Hands-on training is, in my view, the most effective way to equip them with the necessary skills for producing good language tests. For example, workshops can be organized on such topics as developing test specifications, writing selected-response items, writing constructed-response items, scoring and score reporting, test data analysis, and so on. In these workshops, teachers should be given chances to develop test specs, write test items, and analyze test data. Lectures on the principles of and ethical concerns in language testing would also be useful for promoting teachers' awareness of the social role of a language tester.

To improve test users' assessment literacy, we need to make our work more transparent. For example, we should describe in layman's terms what knowledge, skills, and abilities are measured in a language test, state explicitly the purpose of a language test, and explain how a language test is designed and why different types of items/tasks are used. More importantly, to promote appropriate test use, we need to make assessment criteria and score interpretations known to test users, so that their decisions on the use of a language test can be made based on a good understanding of its intended purposes and the meaning of its scores.

You have "stuck" with language testing for over three decades, ever since your postgraduate studies. I am deeply impressed by your "obsession" with language testing. What do you think is the most fascinating thing about language testing?

You are right that I have been working in the field of language testing for over three decades and

that it is a fascinating area that encompasses disciplines as diverse as applied linguistics, educational measurement, psychology, and sociology. Compared to other areas in applied linguistics, language testing is a young field. When I started my MA studies in the late 1980s, language testing just began to be recognized as an independent disciplinary area in China. Over the past decades, the field has gained traction thanks mainly to an increasing demand for language testing services since China adopted an open-door policy in the late 1970s. I seem to have witnessed the growth of the field. In the late 1980s, very few MA programs offered training in this research area. Now many universities have master or doctoral programs in language testing. And it is now easier for my students to find a position which is somewhat related to developing, delivering, or researching language tests.

However, being a language tester is actually not such a fascinating thing. Working for a large-scale, high-stakes test in China is like walking a tight rope across the Yangtze River. To maintain balance on the rope, we need to be highly professional so as to maintain the quality of the testing service, and we must also be sufficiently sensitive to all the social dimensions of language testing, e.g., the relationship between testing and teaching, test misuses or overuses, test fairness, and social justice. Test security, for example, has always been our top priority in our role as professionals working for a large-scale test. When compromises must be made in test design due to concerns with test security, we have to convince ourselves that fairness is much more important than the beauty of the test format and the precision of measurement.

Understanding and coping with the complexity involved in the development and validation of the CET actually constitutes an important part of my career as a language tester. In 2000, the CET sent a delegate to attend the 22nd LTRC, the annual conference of ILTA, in Vancouver, Canada. I remember that the theme of the conference was "Interdisciplinary Interfaces with Language Testing". We were given two hours to present on the CET validation project. This was my first LTRC, and the first time for the field of language testing to have a deeper knowledge of the CET in China. At the AILA 2005 (The 14th World Congress of Applied Linguistics) in Madison, Wisconsin, US, representatives of TOEFL iBT, IELTS, and CET gave a symposium on the topic "Big Tests". I presented on behalf of the CET Committee. Professor Liz Hamp-Lyons from The University of Nottingham chaired the symposium. Professor Liying Cheng from Queen's University was a discussant and made a presentation on the impact, washback, and consequences of the "big tests". Professor Alan Davies from The University of Edinburgh was also a discussant and talked about "Ethics and the big tests".

Since then I have become particularly interested in knowing the experiences of other large-scale testing programs and sharing the experiences of the CET with professionals from other parts of the world. One question that I find very intriguing is what makes a language test "local" or "global". Why are local tests necessary when a range of global tests are available and competing for markets? This will probably continue to be a theme to explore during the remaining years of my profession as a language tester.

Thank you very much, Professor Jin, for taking time to share with us your life as a practitioner or, as you put it, a practisearcher in the field of language testing and assessment. Before the interview ends, I would like to invite you to give some advice to young language testing researchers and those who are going to enter the field of language testing.

The most important thing for a young researcher, no matter what field they are interested in, is to have a good understanding of the history of the field and its current status and to try to find out in what ways their work may contribute to the field. For example, the field of language testing has developed around conceptualizations of the construct of communicative competence, definitions of test validity, and

approaches to test validation. With a good knowledge of the scope of the field, they are more likely to find meaningful topics of research and ways to contribute to the field, in other words, setting a long-term goal for their careers.

I would also like to suggest that young researchers should develop good reading and writing habits. They should find time to read classics as well as articles on the most up-to-date research, and they can write about their reflections on what they read, or write research proposals. Young researchers should apply for research projects and make the best use of research grants. A project will help them set clearer short-term goals, and they are more likely to make visible progress towards the goals.

A useful way for young researchers to start their careers is to develop a sense of community by participating in the activities of professional organizations, including, for example, conferences, SIGs (Special Interest Groups), webinars, and workshops. Through these activities, young researchers will have chances to learn from senior members of the community and chances to collaborate with researchers who are more experienced in research and share a similar interest.

Dr. Lin Zhang is a lecturer at the School of Foreign Languages, Shanghai Jiao Tong University. She is also a member of the Administration Office of the National College English Testing Committee in China. Her research interests include oral assessment, language test development and validation.