1

Assessing Chinese Learners of English: The Language Constructs, Consequences and Conundrums – An Introduction

Guoxing Yu and Yan Jin

1.1   Introduction

In this chapter, we introduce the context and the rationale for the edited volume on assessing Chinese learners of English as a foreign language. In specific, we will discuss the constant challenges and conundrums in understanding the language constructs, the various assessment methods, Chinese learners' preparation for and performance on English language tests, as well as the wide-reaching consequences of assessing Chinese learners of English. This introduction chapter also presents the logic of the sequence of the individual chapters and the overall organisation of the edited volume. The central question that we keep asking ourselves throughout this edited volume – *What have we learned from research on assessing Chinese learners of English?* – helps us to draw together, though very much tentatively, the implications of the findings of the studies reported in this volume which represents our collective endeavours as researchers to contribute to solving part of the conundrums.

1.2   The context and rationale

Understanding how Chinese students are being tested, how they are preparing or being prepared for different purposes, at different educational levels, and for different tests, will lend some insight into not only the validity of the tests per se but also the wider issues in relation to local and global impacts of the tests. English language assessment as a social practice is hugely complex in terms of assessment policies, practices and hence its impacts at different educational levels. The

uses, misuses and abuses of English language assessment transcend the traditional studies focusing exclusively on the reliability and validity of tests. The policies and practices of assessing Chinese learners of English as a foreign language are intertwined with the social, political and educational systems in which the tests operate; as a result, the impacts of English language tests are social, political and educational in nature. As Ross (2008) rightly pointed out: "Language assessments for high-stakes purposes invariably involve policy making at some level. Language assessment policy analysis requires an appreciation of the social, economic, and historical contexts in which assessment policies are introduced, modified, extended, or abandoned" (p. 5).

To understand the current status of English language assessment in China, it is imperative and inevitable that first and foremost we take into account the history of Chinese imperial examinations and the impact of the examinations on the present social, political and education systems. It is widely accepted that China is the origin of large-scale examinations of individuals' abilities for selection purposes (Bowman, 1989; Martin, 1870). Although the system of imperial examinations was abolished in 1905, its influence is still permanently embedded in the present education and assessment systems in China. *Sit for the exam and fight for the rank* – was and still is not only a manifestation of the nature of competitiveness in all aspects and levels of educational assessment in China but also one of the key mechanisms used by the Chinese government to manage resources and social mobility. Issues in educational access, equity and quality (Davey, Lian, & Higgins, 2007; Hannum, An, & Cherng, 2011; Rong & Shi, 2001; Wang, 2008), social justice and political centralism (Feng, 1995) are the main criticisms of the selection purposes of education assessment in China (see Yu & Jin, 2014).

Compared to imperial examinations, English language assessment, which probably started in the 1860s in China (Cheng, 2008; Fu, 1986), is relatively a "small baby" in terms of its history. However, in terms of its size, scope and reach of influence, English language assessment is colossal; it now permeates every aspect and moment of Chinese society. A phenomenal number of Chinese learners of English, from nursery to higher education institutions and beyond, are taking English language tests. English is the compulsory school subject from year three almost everywhere in China, rural and urban. English is one of the three key subjects (the other two being Chinese and mathematics) in Gao Kao – the national university entrance examinations. College English Test (CET) has millions of test takers every year, e.g., in 2012 alone it had 18 million test takers. There has been a substantial increase in

the number of Chinese taking international English language tests. In 2010 there were over 300,000 Chinese who took International English Language Testing System (IELTS), and a similar number of Chinese taking TOEFL iBT (Test of English as a Foreign Language, internet-based test). Educational Testing Service, the owner of TOEFL iBT, reported a 19% increase of Chinese test takers in 2011 from 2010, and a further 32% increase in 2012 from 2011. According to a recent ETS publication (Liu, 2014), Chinese test takers represent about 20% of the TOEFL iBT population. Test preparation courses, especially for TOEFL iBT and IELTS, have been the major income sources of some public listed Chinese companies such as New Oriental at NYSE and Global Education and Technology at NASDAQ which was purchased by Pearson in December 2011. To gain a sense of the scale of English language learning and assessment, this TED video by Jay Walker is particularly telling:

http://www.ted.com/talks/jay_walker_on_the_world_s_english_mania.html

English language assessment affects not only millions of people within China but also has far-reaching global effects, academically and financially, on recruitment and education of Chinese students in English-speaking universities. According to the UK Council for International Student Affairs (UKCISA), there were 428,225 international students in UK higher education institutions in the 2010–2011 academic year; they made up of 48% of full-time research degree students, 70% of full-time taught postgraduates, and 14% of full-time first degree students. Several UK universities recruited a substantial percentage of their students from overseas (e.g., LSE 66%, Imperial College 40%, UCL 38%, Cambridge 30%, Warwick 30%, and Edinburgh 28%). In the USA, there were 723,277 international students in colleges and universities in 2010–2011 academic year. In Australia, there were 184,830 international university students enrolled as of July 2012. In New Zealand, there were 22,811 international university students enrolled as of April 2012 (around 13% of university enrolments). China is the leading place of origin for international students enrolled in the aforementioned countries; and the number of Chinese students has been increasing substantially year on year. For example, UK higher education institutions enrolled 17% more Chinese students from mainland China in 2011/12 than 2010/11 (Source: UKCISA). As a well established but highly debatable, global practice, universities use students' English language test results as one of the most important admission criteria (Rea-Dickins, Kiely, & Yu,

2007). As a result, we witness an increasing number of Chinese taking TOEFL iBT and IELTS year on year as we described above. These English language tests shape and are shaped by the globalising higher education sector. The English language abilities of Chinese students have an impact on the extent to which the students can access and benefit from their higher education experiences, and affect their lives as students and the overall quality of higher education.

In addition to Chinese from the mainland, there are similarly a large number of Chinese learners and test takers of English in Hong Kong and Taiwan who share in many aspects the cultural, linguistic and educational traditions and values as their mainland Chinese counterpart. In this edited volume, we use Chinese or China as terms associated with the Chinese language and people, rather than as a political entity, unless otherwise stated explicitly.

Among policy makers, curriculum designers, material writers, English language instructors, and assessment professionals, at all educational levels, there are substantial and sustainable interests in understanding the issues surrounding the assessment of Chinese learners of English. A number of academic publications have recently appeared or are under preparation to address these issues. For example, *Researching Chinese Learners: Skills, Perceptions and Intercultural Adaptations*, (Editors, Jin & Cortazzi, 2011, Palgrave), *English Language Assessment and the Chinese Learner*, (Editors: Cheng & Curtis, 2009, Routledge), *English Language Education and Assessment: Recent Developments in Hong Kong and the Chinese Mainland*, (Editor: Coniam, 2014, Springer Singapore). *Assessment in Education* (Taylor and Francis) published a special issue on the assessment of Chinese learners of English, edited by Yu and Jin (2014). *Language Assessment Quarterly* (Taylor and Francis) published a special issue on English language assessment in Taiwan (Guest Editor, Vongpumivitch, 2012). Another special issue on high-stakes English language testing in China is under preparation by Professors David Qian (a contributor to this edited volume) and Alister Cumming (OISE, University of Toronto), to be published by *Language Assessment Quarterly.* Together, these publications make incremental contributions to understanding the constructs and consequences of assessing Chinese learners of English.

1.3   The chapters

Given the nature and scope of the complexity of the issues in assessing Chinese learners of English, no single volume would be able to capture all. This edited volume is intended to provide some insights into

language constructs of assessment, various assessment methods and innovations, Chinese students' preparation for and performance on a number of English language tests, and consequences of assessment. These chapters are arranged broadly in line with the fundamental questions that have been continuously challenging the field of language assessment: who, what, how and why to assess.

- What are the characteristics of Chinese learners we are assessing?
- What makes Chinese learners of English different from learners of other first languages?
- To what extent do the social, political and educational systems in China affect the students' learning motivation and test preparation strategies?
- How are Chinese learners being assessed?
- What are the underlying language constructs of assessment?
- What is Chinese learners' performance in English tests, and what affects their performance?
- What are the consequences of assessment?
- What are the policy and pedagogical implications of requiring students to reach a certain English language proficiency level before they are allowed to graduate?
- How do different stakeholders cope with assessment policy changes? For example, how do teachers implement formative assessment in response to government assessment mandate?

These are the main questions that the research studies reported in this edited volume endeavour to address, from different perspectives. The authors of the chapters come from Australia, mainland China, Hong Kong and Taiwan, UK and USA. Some are seasoned researchers who have published widely on language assessment, and some are recent PhD graduates; however, it is our shared experience in assessing and working with Chinese learners of English that brings us together to address collectively a number of perennial issues in assessing Chinese learners of English.

   Below we briefly introduce the focus of each chapter.

   In Chapter 2, Hamp-Lyons, as one of the main architects of School-Based Assessment (SBA) in English in Hong Kong, reflected on the aims and structure of SBA, and the challenges and issues in developing and implementing SBA in this fervently examination-oriented society. SBA is a typical example of Hong Kong government's initiative to address the dominant culture of summative assessment in schools. As a kind of teacher-based assessment, SBA is intended to serve both summative and

formative purposes, in the high-stakes English Language examinations for secondary school students at age 15–16/16+. SBA was introduced by the government, seeking a balance between summative and formative assessment to make a major educational shift in assessment; however, it met with strong resistance from teachers initially. Hamp-Lyons explained that some of the cultural and political influences helped and hindered the effective implementation of SBA in Hong Kong. She argued that a rigorous teacher's professional development programme and a carefully developed and validated set of assessment criteria and standards are two essential components for successful implementation of assessment innovations. Although SBA is increasingly being accepted by teachers and other stakeholders (including researchers), there are a number of issues that have remained problematic in the nearly ten years of this innovation. In this Chapter, Hamp-Lyons highlighted two of these issues. The first issue is related to the planning or preparation time for Group Interaction tasks in SBA English. Thanks to the "test prep" culture that is "ubiquitous" in Hong Kong, variation in planning time for the Group Interaction tasks could potentially pose threats to the validity of the tasks. The second issue has something to do with the different interpretations of "fairness" – fairness often viewed as equivalent to reliability in the examination-oriented societies, and fairness in terms of opportunity for learning, an opportunity for every student to develop and demonstrate their knowledge and ability to the best of their capabilities. This chapter clearly demonstrates what Ross (2008) argued, which is that language assessment policy analysis requires an appreciation of the broader social, cultural and political contexts in which educational assessment policies or innovations operate, but more importantly, Hamp-Lyons presented a very interesting and thought-provoking first-hand, first-person narrative of the challenges and issues that SBA English faced and still faces in the nearly ten years of implementation.

Following on the same topic, but from a more technical perspective of the implementation of SBA English, Lam reported in Chapter 3 a validation study on Group Interaction tasks. Lam observed that there was a considerable variation in the amount of planning or preparation time given to students for Group Interaction tasks – one of the two continuing challenges that SBA English faces as Hamp-Lyons pointed out in Chapter 2. He looked at how the task was implemented in schools and the authenticity of engagement in student interactions. Based on conversation analysis of student interactions and the stimulated recall interviews with students and teachers, Lam reported that the spoken

discourse of the Group Interaction tasks exhibited some superficial features of authentic interactions and that the students' pre-task planning activities revealed the "contrived and pre-scripted nature" of such interactions. The interactions observed were essentially a "staged performance of pre-scripted dialogues", in other words, "the product of students acting out a composed dialogue based on their knowledge and perceptions of what interactional competence is, rather than students' spontaneous performance of the competence that involves moment-by-moment monitoring of and contingent reaction to each other's talk in real time". The findings of this study can have important implications for designing SBA Group Interaction tasks and the assessment criteria. More generally, as group and paired speaking tasks aiming to assess students' interactional competence often have "planning time" as a key task condition, the findings of this study offer further evidences on the effects of planning time on the features of interactions in such tasks. The next three chapters (4–6) continue the same topic on speaking assessment. Chapter 4 reports on the communication strategies used by test takers in computer-based and face-to-face discussion tasks, Chapter 5 on test takers' use of single words and multi-word clusters in a paired speaking test, and Chapter 6 on test takers' use of formulaic sequences (similar to multi-word clusters in Chapter 5) in a monologue story-retelling task.

In Chapter 4, Jin and Zhang reported a small-scale exploratory study investigating the comparability in test takers' use of communication strategies in two different modes of speaking tasks. Data were collected from six pairs of test takers who sat both the computer-based and the face-to-face College English Test – Spoken English Test (CET-SET). Like Lam in Chapter 3, Jin and Zhang conducted conversation analysis of test takers performance in the two discussion tasks, and found a high level of similarity in both the quantity and variety of communication strategies used by the test takers. They also reported that test takers were generally capable of making effective turn-taking decisions in the computer-based discussion task. Furthermore, in both computer-based and face-to-face discussion tasks, test takers who were awarded a high score on communicative effectiveness made more frequent use of interaction strategies while low performers made more frequent use of production strategies. This small-scale study provided some supporting evidences for the implementation of computer-based CET-SET discussion tasks. Given the number of students taking CET annually, these are particularly welcoming evidences to support the on-going reform and improvement of the delivery of the test.

In Chapter 5, Xu compared the basic spoken vocabulary used in face-to-face interactions by Chinese learners of English and English native speakers. Xu analysed the high-frequency single words and multi-word clusters in the College Learners' Spoken English Corpus and the broadcast conversation and discussion component of British National Corpus. Xu reported that Chinese university students tended to underuse lexical items of interactive functions (e.g., interactive words, interjection in discourse markers) and clusters of vagueness and approximation function; but they tended to overuse conjunction and hesitation in discourse markers. The analysis of the learner corpus also revealed that Chinese students used only a limited number of multi-word clusters in interactions and that they often used them repeatedly, in a sharp contrast to the diverse use of multi-word clusters by English native speakers in similar contexts or genres. Xu argued that the considerable differences between Chinese learners and English native speakers in their use of single as well as multi-word clusters might be attributable to the lack of emphasis or opportunity to learn these aspects of language in the English curricula in Chinese schools. She suggested that "interactive words, discourse markers and clusters of politeness and vagueness functions that enhance communicative competence should be introduced at an early stage of language learning" as the key implications of the findings of her study.

In Chapter 6, Wang and Chen examined the features of formulaic sequences used by test takers in a story-retelling task of Spoken Test for English Majors – Band 4 (STEM4). Test takers listened to a story (about 300 words) twice, taking notes while listening, and then retold the story within three minutes, without any preparation time after listening. To some extent, the story-retelling was a listening/speaking-integrated task, as test takers had to understand the source before being able to retell the story. The extent to which test takers used formulaic sequences directly from the source text or modified them could provide some glimpses of (a) the role that short-term memory might have played for successful completion of the task and (b) the validity of story-retelling task as a measure of speaking ability. The use of formulaic sequences from the source was found to be helpful for test takers to construct fluent texts with less effort. However, the formulaic sequences in the source text were not readily useable unless test takers made a full use of language knowledge and their ability to memorize (though short-term) formulaic sequences to reproduce meaningful and grammatically correct sentences in English. In other words, memorization of formulaic sequences alone did not guarantee successful completion of the

story-retelling task. Story-retelling may be an old-fashioned method for assessing speaking ability, but useful as a pedagogical task for developing speaking ability.

Chapters 5 and 6 investigated Chinese university students' lexical knowledge as demonstrated in their speaking test performance. In Chapter 7, Tang and Treffers-Daller reported the effects on incidental vocabulary acquisition of six reading tasks in a secondary vocational school. The six reading tasks had different levels of involvement ("need", "search" and "evaluation") according to the Involvement Load Hypothesis (ILH) proposed by Laufer and Hulstijn (2001). As Tang and Treffers-Daller pointed out, learning English vocabulary is particularly challenging for Chinese students because the typological distance between the two languages means that there are hardly any cognates between Chinese and English. The results of the experiments showed that students learned more words in reading tasks with a higher involvement load and they also retained more words as shown in a delayed and unexpected post-test. In terms of the contribution of the three different components of involvement – "need", "search" and "evaluation", Tang and Treffers-Daller found that "evaluation" was the most important and "search" the least important of the three. When involvement load was the same, students who carried out output-oriented reading tasks did not outperform those who did input-oriented reading tasks.

Unlike chapters 2–7 that report Chinese learners' performances in English speaking or vocabulary tests, the studies reported in the next five chapters (8–12) focused on the perceptions of the two key stakeholder groups, learners and teachers. Stakeholders' perceptions are essential for test validation purposes. In specific, chapters 8–11 report Chinese learners' attitudes and reactions to assessment innovations and assessment policy changes; and Chapter 12 reports teachers' attitudes towards the use of Standard English and Chinese English in assessing Chinese learners.

SpeechRater[SM] is an automated scoring system which is used to provide quick score feedback on the speaking section of the TOEFL® Practice On-line (TPO) test. Xi, Schmidgall and Wang (Chapter 8) investigated the perceptions of 227 prospective TOEFL iBT test takers from China about automated speech scoring and the impact of the use of SpeechRater[SM] on their test taking strategies. They also looked at the participants' perceptions, interpretations and uses of SpeechRater scores. The research team administered an online survey to Chinese TPO users of various background characteristics and interviewed 35 of them after the survey. The data suggested that the majority of the

participants considered human scoring more accurate than computer scoring and would prefer human scoring. The combination of human scoring with computer scoring was considered more favourably than computer scoring alone for high-stakes decisions. If only computer scoring were used for high-stakes decisions, the participants indicated that they would try to trick the system. However, there was a good level of acceptance among the participants of using SpeechRater for low-stakes purposes, i.e., for test preparation or practice online. The use of SpeechRater did not change the way they responded to TPO speaking section when preparing for TOEFL iBT. As the authors rightly pointed out, users' perception of automated scoring was much under-researched, but critical for understanding the impacts of automated scoring on test takers' strategies for language learning, test preparation and test taking. In terms of test taking strategies, how would test takers, knowing that they are being assessed by an automated scoring system, respond or interact with various assessment tasks, and to what extent would the outcome and the quality of their speaking and writing performances be affected? The authors suggested a number of other interesting research topics, from users or stakeholders' perspectives, to complement the current studies on automated scoring, which often focus on the technical quality of the systems.

In Chapter 9, Qian reported a study on students' attitudes towards the implementation of project-based group assessment (PBGA) in a Hong Kong university. From the perspectives of assessment for learning, PBGA is widely used for evaluating student works. However, for high-stakes purposes, PBGA as a formal assessment of student academic achievement is less accepted. In this study, Qian administered a semi-structured questionnaire with a sample of 62 English major students (42 first-year and 20 senior-year students) in the English Department of the Hong Kong university. The purpose of the survey was to understand the students' views of using PBGA for assessing their performances in the English language classroom. The data suggested that the senior-year students were more positive than the first-year students. The first-year students tended to focus on negative aspects of PBGA and considered "free-riding" as the main drawback of the assessment method. The senior-year students were generally more positive, although they felt that there were some persistent issues with PBGA, especially the issues of fairness in assessment. As indicated in the data, the majority of the students preferred to be assessed through individual rather than group projects. Other issues such as how to put students into groups, the effects of students' personality, availability, learning motivation and

commitment on group dynamics, the development and use of transparent and fair assessment criteria, were among those discussed by Qian in this chapter. Qian offered some suggestions on how to improve the implementation of PBGA as a formal assessment method, while stressing the importance of taking into consideration the nature of competition at individual level in Hong Kong's education and assessment (see also Chapter 2).

In Chapter 10, Chen and May reported on Chinese university students' reactions to the government's initiative to include formative assessment in College English aiming to promote students' learning and engagement. It was a case study involving interviewing one College English teacher and four of her students, observations of her six lessons consecutively and a survey with 100 students of the College English teacher. In this chapter, Chen and May presented the profiles of the four students – two considered as active in learning and assessment and the other two inactive. The analysis of the profiles of the four students indicated that their responses to the change of assessment policy, especially the inclusion of their performance and participation in classroom towards the final grade that they would receive for College English, were influenced by a number of sociocultural factors. The imbalanced economic development in different regions in China did not seem to have a direct impact on the students' behaviour in classroom or their attitudes towards the change of assessment policy; however, it was evident that students from the more developed areas had higher English language proficiency, especially in speaking, than those from disadvantaged areas. Their previous experience in English language learning and assessment, especially their learning styles, and the degree of their willingness and motivation to play the assessment game, in other words whether they were test or learning-oriented, seemed to be the key factors that influenced these students' responses to the assessment policy change implemented by the university authority. However, it is important to bear in mind that only one student was really responsive to the assessment policy change. The other students would probably be more active in classroom participation if there had been a greater weighting of classroom participation in the calculation of the final grade that they would receive for College English.

Student motivation to learn English is also one of the focuses of Chapter 11. In this chapter, Zheng reported Chinese university students' views on what affected their English language learning and their performance in College English Test Band 4 (CET-4). As part of the larger study, Zheng selected 12 from over 800 students, who responded

to a survey, to conduct in-depth interviews with them. The interview data demonstrated that the students had a variety of motivations to learn English, from contributing to the country's globalization and economic development to more personal reasons. Overall, due to the fact that English is used as a main *lingua franca* in the world, the students attached great importance to learning English; however, they were also mindful that they should keep a balance between learning English and maintaining their Chinese culture and identity. In addition, some students also reported that their motivation for learning English changed over time, particularly from the time when they took the national university entrance examination to the time when they took CET-4. Differences between male and female students with regard to their commitment to learning English were also observed; so were differences between high-proficiency and low-proficiency students in their resources to learn English. The influences from the society, teachers, parents and peers were all instrumental for the students to put efforts into learning English. Passing CET-4 was considered one of the major external forces that motivated the students to learn English since the test provided them with some kind of direction for them to learn the language. The students also made a number of suggestions for improving English language assessment in China with reference to CET-4. For example, they suggested that it would be wise for the test provider to learn from international English language tests to include integrated assessment tasks in CET-4, and that it should make available the speaking test for every test taker of CET-4 rather than just for those who can manage to achieve a certain score in the written test.

In Chapter 12, Zhang investigated teachers' attitudes towards Standard English and Chinese English. The sample of her survey included 20 native English-speaking teachers (ten teaching English as a second language or a language-related subject in Australia; and the other ten working in Chinese universities but originally coming from USA, New Zealand or Canada) and 20 Chinese teachers of English in universities. All the teachers in the study had exposure to Chinese English. It was found that the two groups of teachers were generally in good agreement about the ownership of English and the definitions of the native speakers of English. However, the native English speakers were more open-minded about the use of different varieties of English than the Chinese teachers of English who preferred Standard English (i.e., American/British English) as the norms for learning, teaching and assessment. The teachers working in Chinese universities, regardless of their first language (English or Chinese), were reluctant to accord

Chinese English as a variety of English in its own right, in contrast to the Australia-based native English speaking teachers who considered Chinese English as a language. Teachers' different attitudes towards varieties of English can have a number of implications for teaching and assessing Chinese learners of English. In terms of assessment, teachers who have different attitudes towards varieties of English may operate different constructs of language from teachers who think Standard English should be the norm for assessment; and as a result, their assessment criteria may differ.

The final two chapters (13–14) of this edited volume report the impact of the requirement of English language proficiency for graduation in Taiwanese universities, and the impact of Cambridge English tests in schools in mainland China respectively. A number of Taiwanese universities set an exit requirement for English language proficiency before a student can graduate, aiming to improve students' English language proficiency. The universities have the autonomy to decide whether or not to implement this policy and which tests are acceptable for this purpose. In Chapter 13, Lin reported part of her PhD study that investigated the impacts of the requirement of English language proficiency for graduation. In specific, she examined the impacts of the requirement on the English for Academic Purposes curriculum for non-English majors, in relation to teaching and learning within the English language classroom and students' learning outside classroom. Lin collected data from two universities – one with the requirement for English language proficiency and the other without. Her main data included relevant policy documents from the universities, lesson observations and interviews with teachers and students. One of the findings of her study, reported in this chapter, showed that the locally developed General English Proficiency Test (GEPT) had only limited influence on teaching, compared to international tests such as IELTS and TOEFL iBT. However, the influence of GEPT in local universities was much reinforced by the implementation of the exit requirement and the importance that the general public attached to GEPT because it is arguably the best-known test in the society.

In Chapter 14, Gu and Saville briefly reviewed Cambridge English examinations in China in the last two decades, and looked at the notion of impact in the Chinese context and the effects and consequences that Cambridge English examinations exerted on English language teaching and learning in China. They reported in this chapter two studies on the impact of "Cambridge English: Key for Schools", "Cambridge English: Preliminary for Schools" and "Cambridge English

Business Certificates". In both studies, they used a structured question-naire and interview as the main instruments to collect data with on a number of important aspects of the impact of the three tests. The main data reported in this chapter included test takers' characteristics, their perceptions of the tests, motivations for and anxiety about taking the tests, and preparation for the tests. In Study One, which examined the impact of the two tests on schools, they found that Chinese students taking the two tests were younger than the targeted groups in the rest of the world. They attributed this to the prevailing culture of test taking in China. Even at a young age, the students were very positive towards the tests and highly motivated, for a variety of reasons, to take the tests. The data also evidenced some negative impacts of the tests, especially in relation to anxiety and extra workload that the students experienced. In Study Two, Gu and Saville looked into the impact of Cambridge English Business Certificates – Vantage and Higher – on university students. Similar findings to those in Study One were noted, with regard to test takers' highly positive views about the tests and high motivations for taking the tests. However, it was also found that the Chinese test takers were less familiar with certain aspects of the tests, e.g., the rating scales for speaking and writing, than similar cohorts of test takers in the rest of the world. They attributed the test takers' lower familiarity or aware-ness of the rating scales to the fact that Chinese university students are typically not assessed in their day-to-day learning in the same way as in the Cambridge examinations.

In summary, the studies reported in this edited volume covered a diverse, but still focused range of issues that we encounter when assess-ing Chinese learners of English. Chapter 2 reflected on the issues and challenges in the development and implementation of School-Based Assessment in English in Hong Kong. Chapters 3–6 reported on Chinese students' performances in speaking tests, notably in relation to their communication strategies, interaction and lexical knowledge. Chapter 7 reported on the assessment of secondary school students' vocabulary acquisition through reading. Chapters 8–12 looked at Chinese learners' and their teachers' attitudes and reactions to assessment innovations and assessment policy changes. These chapters covered a range of assessment issues, from automated speech scoring, project-based group assessment, formative assessment, motivation to learn English, and the use of Standard English and Chinese English in teaching and assessing Chinese learners of English. Chapters 13–14 presented studies on the impact of local and international English language tests on teaching, learning and specific test preparation efforts.

1.4   Conclusion

Across the studies, it is evident that we must take into account the social, political, educational contexts in which English language tests and assessment innovations and policies operate in China in order to better understand the impacts of the tests and policy initiatives, and how well Chinese students perform in various tests and why. At the individual level, it is equally important for test validation purposes to investigate Chinese students' attitudes towards assessment innovations, their language learning motivations as well as other personal characteristics which are arguably shaped by the wider social, political and educational contexts.

This volume, as our collective efforts to address a number of perennial issues in assessing Chinese learners of English, makes an important contribution to better understanding the complexity and dynamics of assessing Chinese learners of English in different educational contexts and levels. This volume provides some insights into a number of selected key challenges and issues in English language assessment, e.g., language constructs of assessment, assessment innovations and methods, test preparation and performance, and consequences of assessment. However, they only represent the tip of an iceberg, the enormous challenges that we face in assessing Chinese learners of English; not only because of the sheer number of test takers but also because of the wide-reaching influences that the use of test results exerts within China and globally. To some extent, we agree with what Bachman (2009) wrote in the foreword to the volume edited by Cheng and Curtis (2009) – *English Language Assessment and the Chinese Learner*, "the language testing issues discussed … are not unique to the assessment of Chinese Learners' English at all. Rather, the enormity of the enterprise in this context magnifies kinds of problems that are faced by language testers everywhere, and makes it more difficult to find justifiable solutions" (p. x). In this sense, the studies in the present volume also make essential contributions to the global knowledgebase of English language assessment.

References

Bachman, L. F. (2009). Foreword. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. x–xii). New York: Routledge.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist, 44*(3), 576–578. doi: 10.1037/0003-066X.44.3.576.b

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15–37. doi: 10.1177/0265532207083743

Cheng, L., & Curtis, A. (2009). *English language assessment and the Chinese learner*. New York: Routledge.

Coniam, D. (2014). *English language education and assessment: Recent developments in Hong Kong and the Chinese mainland*. Singapore: Springer.

Davey, G., Lian, C. D., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education, 31*(4), 385–396. doi: 10.1080/03098770701625761

Feng, Y. (1995). From the imperial examination to the national college entrance examination: The dynamics of political centralism in china's educational enterprise. *Journal of Contemporary China, 4*(8), 28–56. doi: 10.1080/10670569508724213

Fu, K. (1986). *A history of foreign language education in China (中国外语教育史)*. Shanghai: Shanghai Foreign Language Education Press.

Hannum, E., An, X., & Cherng, H. Y. S. (2011). Examinations and educational opportunity in China: Mobility and bottlenecks for the rural poor. *Oxford Review of Education, 37*(2), 267–305. doi: 10.1080/03054985.2011.559387

Jin, L., & Cortazzi, M. (Eds.). (2011). *Researching Chinese learners: Skills, perceptions and intercultural adaptations*. Basingstoke: Palgrave Macmillan.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: the construct of task induced involvement. *Applied Linguistics, 22(1),* 1–26.

Liu, O. L. (2014). Investigating the relationship between test preparation and toefl iBT performance. *ETS Research Report Series*, n/a-n/a. doi: 10.1002/ets2.12016

Martin, W. A. P. (1870). Competitive examinations in China. *The North American Review, 111*(228), 62–77. doi: 10.2307/25109555

Rea-Dickins, P. M., Kiely, R., & Yu, G. (2007). Student identity, learning and progression: The affective and academic impact of IELTS on 'successful' candidates. In P. McGovern & S. Walsh (Eds.), *IELTS research reports volume 7* (pp. 59–136). Canberra: IELTS Australia and British Council.

Rong, X. L., & Shi, T. (2001). Inequality in Chinese education. *Journal of Contemporary China, 10*(26), 107–124. doi: 10.1080/10670560124330

Ross, S. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing, 25*(1), 5–13. doi: 10.1177/0265532207083741

Vongpumivitch, V. (2012). English-as-a-foreign-language assessment in Taiwan. *Language Assessment Quarterly, 9*(1), 1–10.

Wang, L. (2008). Education inequality in China: Problems of policies on access to higher education. *Journal of Asian Public Policy, 1*(1), 115–123. doi: 10.1080/17516230801900444

Yu, G., & Jin, Y. (2014). English language assessment in China: Policies, practices and impacts. *Assessment in Education: Principles, Policy & Practice, 21*(3), 245–250. doi: 10.1080/0969594x.2014.937936