

## RESEARCH

## Open Access



# Developing common listening ability scales for Chinese learners of English

Lianzhen He\* and Dajian Chen

\* Correspondence: [hlz@zju.edu.cn](mailto:hlz@zju.edu.cn)  
Linguistics Department, Zhejiang University, 866 Yuhangtang Road, Hangzhou, China

## Abstract

**Background:** The Common European Framework of Reference (CEFR) developed by Council of Europe has been gaining popularity since its publication in 2001, and reshaping the practice in language teaching, learning and assessment not only across the European continent but also in many other places such as Japan (CEFR-J project). Some pioneering work was done in China over the last ten years, leading finally to a call for a Chinese framework of reference. Currently, ESL teachers and researchers in China are working on a national English ability scale, China Standards of English (CSE), which hopefully will guide or even regulate the English language teaching and assessment practice in China.

**Methods:** The listening construct is defined herein in a use-oriented approach consisting of cognitive ability, listening strategy, linguistic knowledge and performance in typical listening activities. Descriptors are collected from both teachers' description of students' listening ability (with a sample of 159 teachers and 475 students at different educational stages and 119 professionals) and existing documents, both in English and Chinese, such as proficiency scales, teaching syllabuses, curriculum requirements, test specifications and rating scales.

**Results and discussion:** In the descriptor collecting process, a total of 1403 descriptors for listening ability are collected, with 1240 from documentation and 1263 from the sampling approach, plus 113 listening activities in the Chinese context with sufficient details, and nine levels (1-9) are identified in line with the educational stages in China. Descriptor analyses have been done, and scaling and validation are being done. We have encountered some problems so far in the development of the listening ability scale. For one thing, the cognitive view of English ability is regarded as a distinguishing feature of the language ability framework for CSE, especially for comprehension abilities, but it is hard to operationalize it in the description stage and a more feasible approach should be worked out. For another, listening strategy has turned out to be an unfamiliar concept for many teachers and students when sampling descriptors, which calls for rigorous training and detailed illustration.

**Conclusions:** The CEFR, though not directly applicable in the Chinese context, serves as a good example in the development of CSE. In CSE, language ability is put into the big picture of cognitive ability, thus more attention is given to the language processing stage, although it has proved to be difficult to operationalize. In the descriptor collecting and scaling process, we have taken full advantage of research resources and managed to ensure the coverage and representativeness of the data. The outcome of the project may probably be far from flawless, follow-up research and revision should continue as the scales are put into use.

**Keywords:** Listening ability scales, CEFR, China Standards of English (CSE)

## Background

In China's public education system, English is a compulsory course from the third year onward at primary schools, in secondary schools (middle/high schools) and colleges. Plus numerous part-time English learning/coaching programs, China has millions of English learners and thousands of English teaching institutions, leading to dozens of syllabuses for English teaching and various standards for assessing English ability (see Jin, Wu, Alderson & Song, this issue, for a more detailed review). Without coherent teaching practices and a common ability standard for reference, the English teaching and assessment practice in China is inefficient, resulting in a waste of educational resources. To deal with this situation, with the help of governmental organizations, researchers and scholars in language teaching and assessment are currently working on a national English ability scale, China Standards of English (CSE), which hopefully will guide or even regulate the English language teaching and assessment practice in China.

The Common European Framework of Reference (CEFR) developed by Council of Europe has been gaining popularity since its publication in 2001, and reshaping the practice of language teaching, learning and assessment across the European continent and beyond. A large number of schools in the European Union (for example, the EAQUALS<sup>1</sup> member schools) have made efforts in aligning their curricula and examinations with the CEFR (North 2014). Japan published its first version of CEFR-J in 2012, which is an adaptation of the CEFR for the specific English language teaching context in Japan. Cambridge English Language Assessment has also linked all of its exams to the CEFR levels. As a widely accepted conceptual framework, the CEFR has set a good example for China. However, the CEFR is not directly applicable in the Chinese context. CSE aims to serve as a guiding framework for teaching, learning and assessment, which requires the scales to be comprehensive and carefully balanced rather than illustrative ones such as the CEFR.

This article attempts to introduce the approaches we've adopted in developing the common listening ability scales, which is an important part of CSE, by answering the following questions:

- 1) How to define the construct of listening ability with respect to the English teaching and learning context of China?
- 2) How to describe listening ability in a comprehensive manner?
- 3) How to collect descriptors with reasonable representativeness?
- 4) How to scale the descriptors and how to validate the scales?

In the following sections, we are going to review the CEFR's influence on language teaching and assessment in China, followed by a brief discussion on the differences in conditions and goals between China and Europe with regard to the necessity for a common framework. We will then introduce in detail the development of the listening ability scales, including a discussion on the listening construct and the steps involved in developing the listening ability scales. The paper ends with a discussion on some issues encountered in the development of the scales.

## The CEFR in China

Chinese scholars have long been trying to build an efficient foreign language teaching system (Dai 2001), but it was five years after the publication of the CEFR that the value of

language ability scales was realized when Han (2006) did a comprehensive review of some influential language ability scales in the world, with the CEFR as the most representative one. Han ended his review with a call for nationally unified English teaching standards that would help to link different educational stages in China. Based on the CEFR, Yang and Gui (2007) proposed the idea of developing a common Asian framework of reference for English, in light of the fact that English is the working language among Asian countries. Following the mushrooming of the CEFR-related studies, Chinese scholars have tried to follow the CEFR approach and develop a common language ability scale for Chinese learners. Fang et al. (2008) discussed the principles and steps involved in developing a national English ability scale in the Chinese context. In their discussion they covered some key elements of the CEFR approach such as a descriptive scheme based on communicative language ability, “can-do” statement style, action-oriented approach and empirical data scaling method. Following that, there have been two more papers, introducing the theoretical underpinnings of the CEFR. Fang et al. (2011) did an in-depth investigation of the “can-do” description style, and Liu et al. (2012) discussed the CEFR’s successful application of the communicative language ability (CLA) framework within the illustrative scales and its advanced language teaching philosophy.

Along with theoretical discussions, language policies in China are gradually influenced by the CEFR. For example, “can-do” statements were used in *College English Curriculum Requirements* to describe students’ language ability (Cai 2012; Qin & Zou 2011). The need to develop a unified Chinese ability standard was also much discussed among Chinese as Foreign Language (CFL) researchers (Bellassen & Zhang 2008; Fang 2007).

Several empirical CEFR-related studies have also emerged in recent years. Huang and Jia (2012) tried to align College English Test (CET) with the CEFR using the Dutch Grid (Alderson et al. 2006). Yang et al. (2011) did a pilot study to develop the speaking ability scales of English, and explored the ways of collecting, analyzing and scaling descriptors. The most recent attempt was made by Wang (2012) who investigated the ways to develop and validate descriptors for language comprehension, i.e., listening and reading, for Chinese learners of English.

The publication of the CEFR served as an excellent example, and has profoundly influenced language teachers and researchers in China. With the theoretical preparation and empirical studies reviewed above, China is ready to develop its own national English language ability scale. As North (2014) articulated in his recent book *The CEFR in Practice*:

The CEFR was developed with two broad aims: on the one hand to act as a stimulus for reflection on current practice and on the other hand to provide a common reference point for the elaboration of language syllabuses, curriculum guidelines, examinations and textbooks across Europe. And the purpose was to contribute to reform, innovation and networking in order to improve the efficiency of language learning in the school system (p. 9).

Nowadays, China shares the same needs as Europe—the needs for a national reference framework to guide the language teaching practice at different educational stages and in different educational institutions. However, there are still some specific considerations in the Chinese context. Unlike the Council of Europe that has to deal with many different languages, China only needs to deal with a few foreign languages.

English is the most important foreign language and is a compulsory subject in the Chinese school system, so a national English ability scale is more than a conceptual framework of reference. The CEFR's non-prescriptive scheme and illustrative descriptors make it flexible and internationally applicable, but as is mentioned in the introduction part, in order to increase the efficiency of English language teaching and assessment practice in China, a unified English language ability scale would serve as a standard which can be prescriptively applied off the shelf. In addition, Chinese learners may have different language use situations and cognitive styles from the European learners, so the construct should be defined in a different way. Furthermore, China has millions of English learners across the country with diverse learning conditions and proficiency levels, which requires the levels of the scales to be as fine-grained as possible.

The CEFR has several identifiable limitations (Alderson 2007; Hulstijn 2007; Little 2007), the most obvious one being the incompleteness of the illustrative scales which are acknowledged as the most influential parts of the CEFR. There are no illustrative scales for mediation abilities and limited scales on strategy use and receptive skills, i.e. listening and reading comprehension. Moreover, the descriptors in the CEFR are often considered vague, imprecise and not reader-friendly. Another limitation is the CEFR's failure to take account of the cognitive aspect of language proficiency, which plays an important role in language processing. Fortunately, with the funding and administrative support from the government organizations, China has the advantage of involving more researchers and institutions at the research stage to take full account of the English learning context and overcome many of those limitations.

### **The listening ability scales**

The China Standards of English include eight language ability components which are currently in practice in the Chinese context of teaching, learning and assessment, namely listening, speaking, reading, writing, translation, interpretation, grammar and pragmatics. In this section, we are going to introduce the approaches we've adopted in developing the listening ability scales.

### **Defining the listening construct**

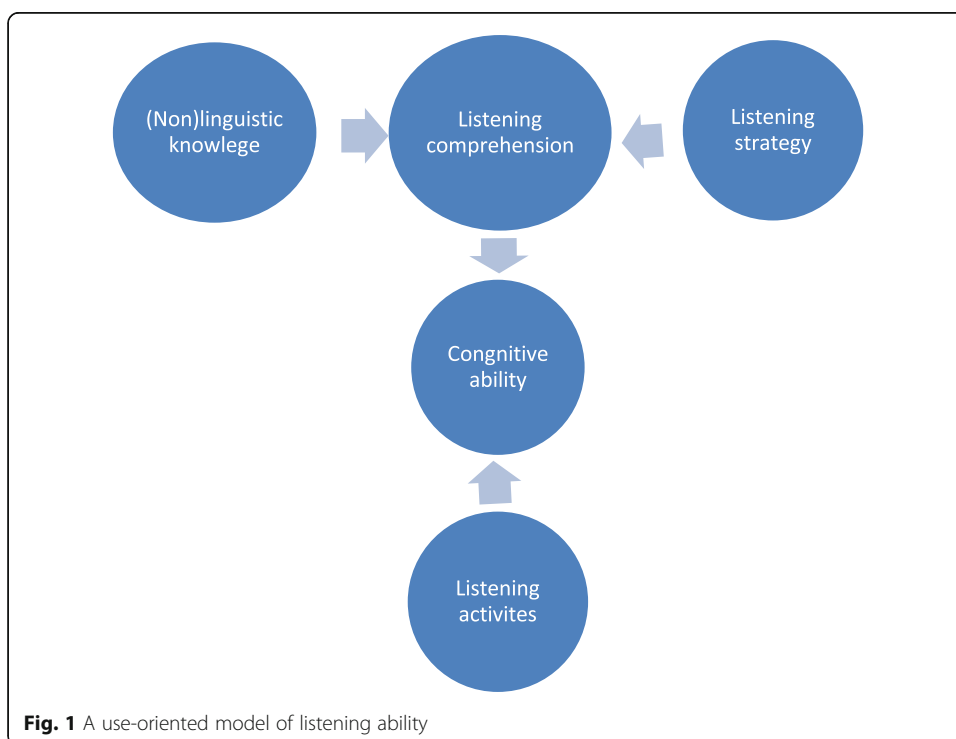
To meet the practical needs of different stakeholders (learners, teachers, policy makers, testers, employers, etc.), language ability is defined in this project as the ability to use language to create and interpret meaning. It is the practical ability of language comprehension and expression presented in specific language activities about certain topics in certain circumstances, with the help of knowledge (both linguistic and non-linguistic) and various strategies. This definition takes a use-oriented approach based on Bachman's (1990) communicative language ability (CLA) model. For years, listening has been the most under-researched skill in language teaching, learning and assessment literature, and major discussions of listening ability tend to focus on the cognitive process of listening comprehension (Anderson 1995; Buck 2001; Field 2002; Rost 2002; Vandergrift 2004). Drawing on existing literature and the general framework of the CSE project, we attempted to define listening comprehension as the ability "to use different sources of knowledge (both linguistic and non-linguistic) and strategies to create a mental representation out of the spoken text; with this representation, a listener

can accomplish certain cognitive tasks such as retrieving (detailed) information, analyzing relationships or making inferences.”

In this respect, listening ability is an integrated cognitive ability presented in specific listening activities such as identifying, retrieving, summarizing, analyzing, critiquing, and evaluating (adapted from Anderson et al. 2001). In the comprehension process, linguistic knowledge serves as the basic resource for understanding, while world knowledge fills meaning gaps. Listening strategy is the method used by a listener to make the understanding less stressful and more successful. Listening comprehension is temporal and the result is invisible, thus we have to turn to actual listening activities when describing this ability. See Fig. 1 for the relationships between the listening ability components.

### Development of the descriptive scheme

Following the defining of the construct is the most critical issue of mental ability scale construction, i.e., how to describe it? This would involve the style and scheme of description. In this project, we decided to follow the “can-do” statement style, the approach adopted in the design of the CEFR, since “can-do” descriptions can help to “establish a link between real world tasks and the language points and other competences that are necessary to perform those tasks effectively” (North 2014, p. 11), thus giving teachers and learners the learning objectives, and also giving other stakeholders (testers, policy makers, employers, etc.) the convenience to use the descriptors directly. “Can-do” statements should positively describe what kind of language activity the user can accomplish and how good the performance is. As the



**Fig. 1** A use-oriented model of listening ability

scheme of description determines the scale structure and the arrangement of descriptors, it should reflect the predefined listening construct.

An operational descriptive scheme (or categories) covering all the essential aspects of listening ability was developed from the model in Fig. 1. Listening-related cognitive ability, listening strategy and linguistic knowledge are identified as three core attributes of one’s listening ability, and will be described intensively. The illustrative descriptors of listening activities in the CEFR have proved to be useful for various users, and our scheme intends to select several typical listening activities in the Chinese context and describe them in detail. There are four subsets of scales (cognitive ability, listening strategy, typical listening activities and use of linguistic knowledge) under the overall listening ability. See Fig. 2 for a complete picture of the scheme.

The cognitive ability subset is the key component of the listening ability and has the largest number of descriptors. Cognitive performance at different levels (identifying, retrieving, summarizing, analyzing, critiquing, and evaluating) regarding a specific listening activity under certain conditions are identified to describe listener’s listening ability, e.g. “*can follow complicated stories narrated at normal speed, and analyze the motivation or intention of a certain character*”. Descriptors are divided into six categories according to the function of target listening text (or activity), namely narration, description, exposition, argumentation, instruction and interaction, and each category constitutes a subscale. For example, descriptors relating to conversation listening are put into the interaction scale, while listening to lectures are included in the exposition scale.

Similar to the cognitive ability subset, the listening strategy subset consists of three scales – planning, execution, evaluation and repair, corresponding to the cognitive processing stages of listening comprehension. The strategy scales include both metacognitive strategy (referring to mental awareness, e.g. *can anticipate the goal of listening in familiar settings and pay attention to key information*), and cognitive listening strategy (referring to specific approaches taken to assist comprehension, e.g. *can use topical knowledge and context clues to figure out the meaning of unfamiliar words*).

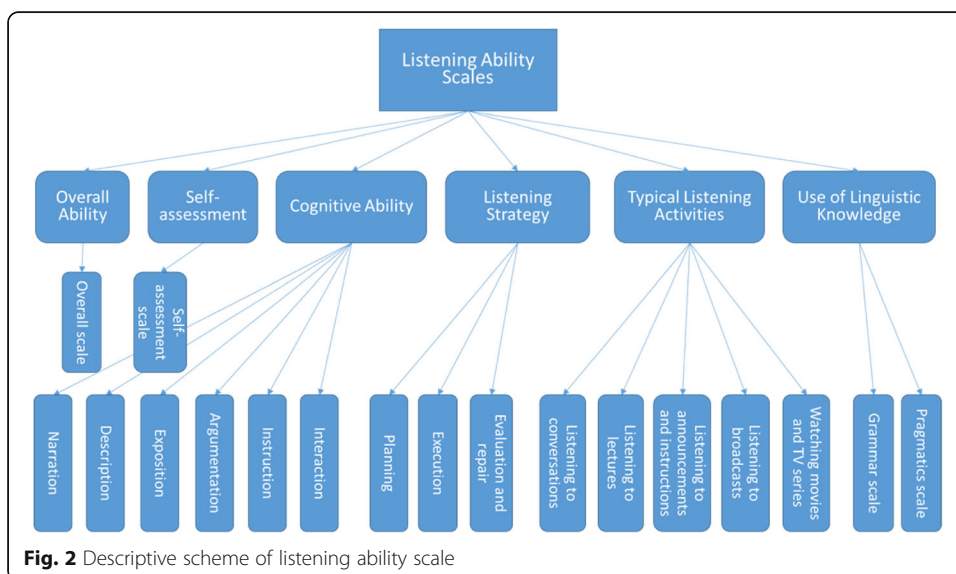


Fig. 2 Descriptive scheme of listening ability scale

The typical listening activity subset consists of scales focusing on a few most frequently attended listening activities for the Chinese learners, such as listening to conversations, listening to lectures, listening to news broadcastings, etc. These scales can be very useful for teachers when designing curricula and exams. The use of linguistic knowledge subset<sup>2</sup> includes descriptors about the quality of linguistic knowledge application in the listening process, especially phonological knowledge and vocabulary knowledge. The global scale and self-assessment scales are developed based on the four subsets of scales, with a broader purpose of serving different users of the listening ability scales, such as policy makers and self-learners.

### Steps involved in developing the listening scale

#### 1. Descriptor collection

In order to collect descriptors with a large enough representative sample of the Chinese learners, two methods were used in the descriptor collecting process: documentation and sampling. Through documentation, we collected a wide range of existing documents such as English proficiency scales (e.g. CEFR, Canadian Language Benchmark), teaching syllabuses, curriculum requirements (e.g. College English Curriculum Requirements), test specifications and rating scales (totally 42 documents, both in English and Chinese) which may contain descriptors of listening ability. In the sampling process, a total of 159 teachers and 475 students sampled from various educational stages and 119 professionals from different fields (see Appendix for the sampling distribution) were invited to write descriptors. After rigorous training regarding the requirements for writing the “can-do” statements and the descriptive scheme, teachers were asked to describe the actual listening ability of their students in aspects of listening-related cognitive ability, listening strategy and linguistic knowledge; students and professionals were asked to describe their own listening ability as to what listening activities and tasks they could accomplish. In total, we collected 1240 listening ability descriptors from documentation and 1263 descriptors from the sampling approach.

The descriptors collected are supposed to cover the whole listening ability continuum. Therefore, in order to make the scaling process smoother, each descriptor needs to be assigned an initial level. Nine levels<sup>3</sup> (1–9) were identified in the descriptor collecting process in line with the educational stages in China. Thus the sampling descriptors fit naturally into these levels initially. For the descriptors collected from the existing documents, we assigned them to the corresponding levels based on their original ability levels (target students’ educational stage or CEFR-related level of the documents).

To develop the typical listening activity scales, all the listening activities that Chinese learners may be involved in need to be collected. During this stage, listening activities were collected from the participating teachers and students, listening textbooks used by students at different educational stages, and from people in different professional fields. In total, we collected 113 listening activities in the Chinese context with sufficient details, such as business negotiations, in-service training classes, job interviews, TED talks, etc.

## 2. Descriptor analysis

During this stage, the first step was selecting and editing raw descriptors, including deletion of redundant and overlapping ones, and editing of those failing to meet the requirements for descriptors. In our study, a descriptor must have three essential elements: the cognitive verb, the target listening text (activity) and defining words. The defining words are those that modify either the cognitive action or the listening text, thus pinpointing the level of a certain descriptor. For example, in the descriptor “*can understand the main points of radio news broadcasts on familiar topics delivered slowly and clearly*”, “understand the main points” indicates the cognitive action of summarizing, “radio news broadcasts” specifies the target listening text (narration), and “on familiar topics delivered slowly and clearly” are defining words which detail the listening text.

We conducted several rounds of revisions to ensure that all the descriptors follow the above-mentioned three-element structure.

### ***The 1st round***

Each descriptor was independently revised by members of the listening team (one of the eight teams of the national CSE project).

### ***The 2nd round***

Cross checking was done within the listening team so that each descriptor was reviewed by more than three team members.

### ***The 3rd round***

Listening comprehension and reading comprehension, both receptive skills, have some similarities in the cognitive process, so the descriptors were further cross checked between the listening team and the reading team. About seven hundred listening ability descriptors remained after the above-mentioned three rounds.

### ***The 4th round***

More than 5000 descriptors covering listening, speaking, reading, writing, translation, interpretation, grammar and pragmatics were then arranged into more than 200 questionnaires to collect feedbacks on the quality of the descriptors, including explicitness and clarity of wording, completeness of structure, and appropriateness of classification. All the researchers of the CSE project, over 100 in total, were involved in this reviewing process.

### ***The 5th round***

The listening team further revised the descriptors based on the feedbacks from the 4th round. To ensure the readability and quality of the descriptors, feedbacks from the stakeholders’ (teachers, students, experts, professionals, etc.) were also collected through questionnaire surveys and interviews.

For the listening activities, we conducted a questionnaire survey among teachers, students, professionals, and experts (totally 1756 respondents) to decide on several typical ones for further description. Five typical listening activities in the Chinese



context were determined, namely listening to conversations, listening to lectures, listening to announcements and instructions, listening to broadcastings and watching movies and TV series.

So far, every descriptor has found its place in the descriptive scheme and has been assigned an initial level. It will take one more step to become a verified ability scale: scaling with empirical data.

### 3. Scaling and validation

The scaling and validation process is realized by nationwide large scale questionnaire surveys and in-depth interviews. All the descriptors in the scale including listening, speaking, reading, writing, translation, interpretation, and linguistic knowledge (grammar and pragmatics) are put together and about 50 questionnaires are designed. Each questionnaire contains descriptors of all aspects (listening, speaking, reading, etc.) of English ability with a target level and subject group. Anchor descriptors will be used for scaling purposes. According to the overall design of the development of CSE, questionnaires will be distributed to around 10000 teachers and 100000 students in over 1000 schools which are stratified through random sampling from the whole nation. In the questionnaire survey, teachers will be required to rate the descriptors against their students' actual English language abilities on a five-point scale, while students will be required to conduct self-evaluation with the descriptors. Multi-facets Rasch Modeling (MFRM) will be used to estimate the difficulty parameter of the descriptors and learners' English ability parameter. Vertical scaling of the descriptors could be achieved through parameter estimates of anchor descriptors (North, 2000). Once appropriate cutoff points of levels considering both the statistical logit locations and practical needs are determined, the scaling process is finished.

Before the final publication of CSE, we will invite potential users to try out the scales and gather feedbacks through in-depth interviews. Handbooks for its application in teaching, learning and assessment will also be published.

The CSE project has been going on for two years, and we've finished the stages of descriptor collection and analysis. Scaling and validation will be finished by the end of 2016, and CSE is expected to be published in 2017. In the next section, we are going to discuss some problems we've encountered so far.

### **Discussion and conclusion**

The cognitive view of English ability is regarded as a distinguishing feature of the language ability framework for CSE, especially for comprehension abilities, but it is hard to operationalize it in the description stage. Cognitive activity is the verbal phrase part of a descriptor, and is supposed to be categorized into different cognitive ability levels (i.e., identifying, retrieving, summarizing, analyzing, critiquing and evaluating). Sometimes, we find it difficult to categorize certain descriptors, because their focus is on the textual features of listening materials, while the verbal part is simply "understand" or "follow". The cognitive ability subset is the main part of the scale and includes most of the descriptors, so a more feasible approach should be worked out to structure them. After several meetings and discussions among the project team members, the main function of listening text (or activity) has been chosen as the criterion for

categorization. In the end, six functions have been identified among those descriptors, namely narration, description, exposition, argumentation, instruction and interaction.

Listening strategy has turned out to be an unfamiliar concept for many teachers and students when sampling descriptors. The reason for this may be related to the English teaching and learning traditions in the Chinese context. Listening and speaking ability has been overlooked or little emphasized for a long time in English language teaching until recent years. There has hardly been any systematic syllabus focusing on the teaching of listening comprehension, let alone the teaching of listening strategy, and few existing documents include listening strategies apart from the CEFR. Through rigorous training and detailed illustration, the teachers finally got the idea and came up with some useful strategies often used by their students in daily life or in the classroom. When dealing with strategy descriptors, abstract ones concerning planning, monitoring, and evaluating are grouped into metacognitive strategies, while specific and concrete ones involving inferencing, elaborating, summarizing, repeating and note-taking are grouped into cognitive strategies; the ones related with emotion or social interaction are grouped into socio-affective strategies (Vandergrift 1997).

Another issue with the descriptive scheme is the overlapping of descriptors between the cognitive ability subset and the typical listening activity subset. One may find similar descriptors in those subscales. But there is a difference in the design, because the cognitive descriptors are meant to cover the core part of the listening ability scales and have the most comprehensive and detailed descriptors to represent a learner's ability; while the typical activity subscales focus on a few real-life listening settings and are of practical use in the work place, in daily life, etc. Nevertheless, it may seem a bit confusing at first glance.

In particular, the scales are designed to be prescriptively used in the Chinese education system, which requires the descriptors to be explicit and internally consistent. For the listening ability descriptors, we usually use criteria such as delivery speed, length of the speech, topic and lexical complexity to modify the difficulty level of a piece of listening material. However, in the case of delivery speed, what's the difference between "delivered slowly" and "delivered at normal speed"? In our study, we manage to distinguish delivery speed at four levels – very slow, slow, normal and fast, then attach a specification for each level, for instance, normal speed means 140–160 words per minute. This level-defining approach can also be applied to other criteria, whereas decisions should be carefully made, with sufficient empirical data support and theoretical grounding, because the development sequence of language proficiency is still unclear (Alderson 2007). As with the CEFR, revision and validation of the descriptors should be an ongoing process along with the research progress in second language acquisition.

In conclusion, this paper, following a brief introduction of how the CEFR has influenced the teaching practices and language policies in China since its publication, gives an overview of the Chinese context which calls for a featured and practical framework for language teaching, learning and assessment. The article then discusses in detail the principles we have followed and the steps involved in developing the listening ability scales. As with the CEFR, the language ability model is based on CLA, but there is some difference. Language ability is put into the big picture of cognitive ability in our model, thus more attention is given to the language processing stage, although it has proved to be difficult to operationalize this model. In the descriptor collecting and

scaling process, we have taken full advantage of research resources and managed to ensure the coverage and representativeness of the data. In the last part, some issues or problems we've encountered so far are identified and discussed. The outcome of the project may probably be far from flawless, follow-up research and revision should continue as the scales are put into use.

**Endnotes**

<sup>1</sup>Evaluation & Accreditation of Quality in Language Services, an international association of institutions and organizations involved in language education.

<sup>2</sup>Scales of this subset are developed in the sub-project of grammar ability scale, and will not be discussed in detail in this article.

<sup>3</sup>In Chinese public education system, English teaching starts from the 3rd year of elementary school, and continues onto middle school, high school and college. Students are assumed to have higher English proficiency level along the continuum of educational stages. So, level 1 corresponds to grade three of elementary school, level 2 grade six of elementary school, level 3 grade three of middle school (3 years), level 4 grade three of high school (3 years), level 5 year three of junior college, sophomore year of non-English majors and freshmen year of English majors of college or university, level 6 senior year of non-English majors and sophomore year of English majors, level 7 senior year of English majors, level 8 MA students majoring in English language and literature, and level 9 the most advanced level for English language users such as senior translators/interpreters.

**Appendix**

**Table 1** Descriptor collecting sources

Methods	Objects	Counts
Documentation	Existing English language ability scales	6
	Curriculum requirements/Syllabuses for English language teaching	12
	Syllabuses for English proficiency tests	24
	Total	42
Sampling	Elementary school teachers	10
	Middle school teachers	20
	High school teachers	24
	Junior college English teachers	60
	Teachers of non-English major undergraduates	36
	Teachers of English major undergraduates	9
	Elementary school students	100
	Middle school students	135
	High school students	81
	Non-English major undergraduates	66
	English major undergraduates	51
	Graduate students majoring in English language and literature	42
	Professionals	119
Total	753	

**Acknowledgements**

We would like to thank Professor JIN Yan and Professor YU Guoxing for their valuable suggestions on earlier drafts of the manuscript.

**Funding**

This paper was based on the project "The Applicability of the CEFR for English Language Education in China" (GZ20140100) funded by Foreign Language Teaching and Research Press and the Key Project of Philosophy and Social Sciences "The Development of China Standards of English" (15JZD049) funded by the Ministry of Education, P. R. China.

**Authors' contributions**

Both authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

Received: 22 September 2016 Accepted: 28 January 2017

Published online: 15 February 2017

**References**

- Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the common European framework of reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly: An International Journal*, 3(1), 3–30.
- Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal*, 91(4), 659–663.
- Anderson, J. R. (1995). *Cognitive psychology and its implications*. New York: W.H. Freeman.
- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston: Allyn & Bacon.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bellassen, J., & Zhang, L. (2008). *Chinese Teaching in the World*, 03, 58–73+3
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Cai, J. (2012). CEFR's influence on Chinese foreign language teaching. *China University Education*, 06, 6–10
- Dai, W. (2001). The construction of the streamline ELT system in China. *Foreign Language Teaching and Research*, 05, 322–327
- Fang, X. (2007). CEFR's enlightenment on Chinese proficiency test development. *Chinese Teaching in the World*, 02, 136–143
- Fang, X., Yang, H., & Zhu, Z. (2008). Creating a unified scale of language ability in China. *Modern Foreign Languages*, 04, 380–387+437
- Fang X., Yang, H., & Zhu, Z. (2011). The background of and approach to can-do description of language ability: taking the CEFR as an example. *Chinese Teaching in the World*, 02, 246–257
- Field, J. (2002). The changing face of listening. *Methodology in language teaching: An anthology of current practice*, 242–247
- Han, B. (2006). A review of current language proficiency scales. *Foreign Language Teaching and Research*, (06), 443-450+480
- Huang, T., & Jia, G. (2012). Feasibility study of the link between the CEFR and Chinese language tests: taking College English Tests as an example. *Foreign Language Testing and Teaching*, 01, 38–49
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language Proficiency. *The Modern Language Journal*, 91(4), 663–667.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91(4), 645–655.
- Liu, Z., Han, B., & Yan, T. (2012). The communicative language ability framework and foreign language teaching philosophy of the CEFR. *Foreign Language Teaching and Research*, 04, 616–623.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- North, B. (2014). *The CEFR in practice* (Vol. 4). Cambridge: Cambridge University Press
- Qin, H., & Zou, W. (2011). An Impact Study of the CEFR on Chinese College English Education. *Foreign Languages in China*, 04, 31–38.
- Rost, M. (2002). *Teaching and Researching Listening*. Harlow: Longman.
- Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, 30(3), 387–409.
- Vandergrift, L. (2004). Learning to listen or listening to learn? *Annual Review of Applied Linguistics*, 24, 3–25.
- Wang, S. (2012). *Developing and Validating Descriptors of Language Comprehension Ability for Chinese Learners of English*. Beijing: Intellectual Property Publishing House.
- Yang, H., & Gui, S. (2007). On establishing a unified Asian level framework of English language proficiency. *Foreign Languages in China*, 02, 34–37+64.
- Yang, H., Zhu, Z., & Fang, X. (2011). Factor analysis of English speaking ability descriptors and their scaling: an empirical study. *Modern Foreign Languages*, 02, 151–161+219.