

Mapping the SMEEA Gaokao tests to the CEFR

Jane Lloyd, Cambridge University Press and Assessment

Graham Seed, Cambridge University Press and Assessment

Xu Wen, Shanghai Municipal Educational Examinations Authority

Hu Yue, Shanghai Municipal Educational Examinations Authority

<https://doi.org/10.37546/JALTSIG.CEFR6-4>

This article is open access and licensed under an Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence.

This paper reports on a project where the Shanghai Municipal Educational Examinations Authority (SMEEA) engaged Cambridge University Press and Assessment (Cambridge) to begin the alignment process of their six foreign language tests to the Common European Framework of Reference (CEFR) (Council of Europe, 2001, 2020). These six language tests are developed by SMEEA and form part of China's National College Entrance Examination system. They are referred to as China's National College Entrance Examination (Shanghai Paper), or the Shanghai Gaokao. This project was undertaken to enable a comparison of the relative difficulty of each language version of the Shanghai Gaokao, in terms of the CEFR. Additional project aims were to carry out training in the CEFR and in CEFR mapping procedures, using a blend of online delivery and a cascaded training model. This project took place between September 2021 and January 2022, involving a series of linked training and mapping activities and workshops. In this paper the focus is on the practical aspects of a computer mediated CEFR mapping project, on our reflections and recommendations, and on the participant feedback. Some outcomes are included which are directly related to the CEFR, but not any outcomes which are confidential.

Keywords: CEFR, alignment, examination, Shanghai Gaokao, mapping, online, cascaded training, China

1 Project aims

This paper reports on a project carried out by Cambridge University Press and Assessment (Cambridge) and the Shanghai Municipal Educational Examinations Authority (SMEEA) on six foreign language versions of China's National College Entrance Examination, commonly known as the Shanghai Gaokao, between September 2021 and January 2022.

The paper begins by stating the aims of the project; the use of the CEFR in language assessment, and a brief overview of the SMEEA Gaokao language tests. It then goes on to describe an overview of the method for the activities. The report continues with an evaluation of the mapping process, including findings from a closing survey of participants in the project, acknowledges limitations, and reflects on how this CEFR mapping project could be further refined. The report concludes with suggestions for other language test professionals involved in this kind of mapping work, and the future development related to CEFR mapping exercises in general.

- The project had the following aims:
 - to familiarise participants with the CEFR by:
 - introducing the CEFR's core conception of language learning
 - introducing the CEFR's level framework of language proficiency
 - inducting participants into the characteristics of input (listening and reading) and output

(speaking and writing), as well as general communicative competence (e.g. vocabulary and grammar) for relevant CEFR levels

- to encourage participants to reflect on how the specific parts of the tests relate to the CEFR and therefore map the six Shanghai Gaokao language tests to the CEFR, enabling the following:
 - an overview of how well each test aligns to the CEFR's theory of an 'action-oriented approach' to language assessment
 - which areas of language use (scales) within the CEFR the items within the test map to
 - which proficiency levels within the CEFR the items within the test map to
- to observe the similarities and differences of the different language versions of the tests by comparing the six language test versions to each other, especially in terms of CEFR language proficiency levels;
- to train up key people within SMEEA's staff and consultants in the use and application of the CEFR according to their context, so that they can carry out the mapping activity by cascading information to others; and also be able to plan and consider how to use the mapping tools and activities in the future.

It should be noted that a full-scale alignment of the tests to the CEFR, as set out in the *Manual for Relating Language Examinations to the CEFR* (Council of Europe, 2009), was not the aim. Nevertheless, the mapping activity that was carried out was based on the principles within the manual, and the process is described in greater detail in Section 4 below.

2 The CEFR

The CEFR, first published in 2001 and then later updated and expanded in its Companion Volume of 2020, is a reference framework of language proficiency for the purposes of learning, teaching and assessing second or subsequent languages. Although originating in Europe, it may be applied to any language, and is now commonly used in many places around the world, not just Europe. It is used as a tool of reference to illustrate the levels of language proficiency; or a means of comparison across different learners, or different groups of learners (e.g. schools, regions, education systems, countries), or one learner's proficiency in different languages or skills within a language. It can be used to compare the appropriacy of different learning and assessment materials, often by its level system.

The reference level system introduced by the CEFR comprises three major bands of the description of language proficiency: the 'A' levels for basic users, the 'B' levels for intermediate users who are able to start using the language independently, and the 'C' levels for proficient users. There are two levels in each, as well as a pre-A1 level for those just beginning their learning journey of a particular language. This results in a seven-level system going from Pre-A1 for the least proficient, then A1, A2, B1, B2, C1 and finally C2 for the most proficient user. The CEFR also includes sub-levels of A2+, B1+ and B2+ for high performance within the A2, B1 and B2 levels only. As these 'plus-levels' do not exist at other levels, and to reduce the cognitive load, this project does not report in terms of the plus-levels, though their existence and relevance were pointed out and discussed with the participants. It should be recognised that C2 denotes high proficiency, rather than native speaker competence.

According to the CEFR, the goal of a language learner is not to achieve the proficiency level of a 'native-speaker'. Rather, the CEFR's level system provides stepping stones on the path to proficiency, with any goal in terms of proficiency level being determined by the context in which learning takes place. It therefore means that partial competence can be recognised and awarded, as well as proficiency whereby a learner is stronger at one skill than another, for example better at oral comprehension than written production. The CEFR categorises language according to its underlying theory that the purpose of learning, teaching and assessing languages is for *communication*, to be used in the real-world in what

is called an *Action-oriented Approach* (AoA). This means that language users will ultimately use language to perform communicative acts in the real world, rather than simply as a subject to study or to pass an exam. The categorisation of language is therefore according to language activities and strategies, foregrounded by four *modes of communication* as summarised in Table 1.

Table 1. Four modes of communication

Reception	These are language activities and strategies used when <i>receiving</i> language: usually by listening, reading or watching.
Production	These are language activities and strategies used when simply <i>producing</i> language: usually by speaking or writing.
Interaction	These are language activities and strategies used when producing language (usually by speaking or writing) in response to having received language (usually by listening, reading or watching). The language produced contains different content to that received, but is directly affected by it.
Mediation	These are language activities and strategies used when reproducing language (usually by speaking or writing) in a different format suitable to the context. The general ideas and content in the language stay the same, but the exact words and manner in which it is reproduced will be different, depending on the context. This may be a change in formality (e.g. formal to informal language), skill (e.g. summarising in speech what has been read), language (e.g. interpreting or translating). Other forms of mediation involve facilitating the appropriate environments for communication to take place.

The ability to engage in language activities has to be supported by *communicative language competences*, categorised in the CEFR as shown in Table 2.

Table 2. Communicative Language Competences

Linguistic	This area comprises the vocabulary, grammar, phonology and orthography needed to communicate.
Sociolinguistic	This area relates to the appropriateness of sociolinguistic and sociocultural use within the context of communication.
Pragmatic	This relates to the knowledge of how language is used, and includes elements such as turn-taking, fluency, coherence and cohesion, and others.

The CEFR also describes plurilingual, pluricultural and sign language competences, which are less relevant for this study and therefore not described in this document. These areas (reception, production, interaction, mediation, linguistic competence, sociolinguistic competence, pragmatic competence) are illustrated in a number of *scales*. As an example, scales within the ‘interaction’ section include oral interaction scales such as conversation, formal discussion and others, as well as written interaction scales such as correspondence and writing notes, messages and forms.

Each scale contains a number of ‘Can Do’ illustrative descriptors, which give a descriptive statement of what language learners are typically able to do. These are ordered in terms of proficiency level, according to the Pre-A1 to C2 spectrum that was described earlier. As an example, in the ‘Conversation’ scale, a pre-A1 descriptor states that a learner at this level can ‘greet people, state their name and take leave in a simple way’, whereas a C2 level learner can ‘converse comfortably and appropriately, unhampered

by any linguistic limitations in conducting a full social and personal life'. It should be noted that these descriptors are indeed *illustrative*, and not prescriptive, in that they give an illustration of what a learner typically can do in a particular area at a particular level. It therefore means that these descriptors can be amended or added to, depending on the context of use.

Assessment bodies frequently use the reference level system to ensure their stakeholders have a common understanding of what level the assessment is targeting. Items within that assessment may be at the target level, as well as just above and just below the target level, in order to provide test items that are useful in providing an accurate assessment of different learners' ability levels.

In order to determine the level of an item, assessment bodies may use the Can Do descriptors found in the CEFR as a basis for their test items. This therefore gives a starting point to ensure a particular item is at a particular CEFR level, and also checks that the test item complies with the CEFR's underlying theory of the communicative language approach. While learning material creators may also take a similar approach, an assessment body is likely to want to carry out additional validation steps to ensure that their material is at a particular level. These steps may include CEFR mapping and alignment studies such as standard setting and performing item response theory (IRT) (Rasch) analysis on pre-test or live administrations to improve accuracy of scoring and results delivery.

In some situations, assessment bodies have a pre-existing test that they want to check whether it aligns to the CEFR in general, and to which CEFR levels in particular. Their starting point may not have been the CEFR, but they wish to carry out the additional validation steps to determine alignment to the CEFR. This is the case of the SMEEA Gaokao tests. The project used the scales and 'Can Do' descriptors found in the 2020 Companion Volume in the mapping activities. These are freely available from the Council of Europe in a downloadable version in both English and French, supplied to SMEEA by Cambridge. The 2021 Chinese translation of the CEFR/CV, coordinated by Xiangdong Gu and Zehan Chen of Chongqing University, China, with the support of the Council of Europe, was also supplied by Cambridge to SMEEA. It is believed that this project was the very first major use of the Chinese translation of the CEFR Companion Volume.

The Can Do descriptors in English, French, German and Spanish are also freely available in an Excel spreadsheet format from the Council of Europe's website, and these were also supplied to SMEEA. Although the full German version of the CEFR/CV is available to purchase in a hard copy format, published by Klett, this version was not supplied due to the project's online format. It is understood that other translations with the support of the Council of Europe are underway, but at the time of the project, any full translations of the Companion Volume in either Russian or Japanese were not available.¹ The 'Global Scale', an overall scale describing language proficiency as a whole, which was published in the 2001 CEFR document and again on p. 175 of the Companion Volume (English version), was made available to the project participants in all six of the languages.

Project participants were free to use whichever language version of the CEFR documentation they felt most comfortable accessing. The project leaders from Cambridge primarily made use of the English version, though attempted to cross-reference to the Chinese version where appropriate.

3 The SMEEA tests

The six language versions of the Shanghai Gaokao are English, German, French, Spanish, Russian and Japanese. Each of the six language versions of the test are broadly similar in terms of structure, although minor differences occur in terms of item numbering, and in some subsections. Generally speaking the written test paper contains several sections: Listening, Linguistic Knowledge, Reading, Summary Writing, Guided Writing and Translation. There is a Speaking test which contains a combined Speaking and Listening section. The structure of the English, German, French and Spanish tests is almost identical;

1. By the time of writing this report, full versions of some other languages, including Spanish, had since been made available on the Council of Europe website.

the Russian test does not have a Summary Writing question and there are some subtle differences in the Speaking test; while the Japanese test is perhaps the most divergent as the number of items in the Listening, Grammar and Vocabulary, Reading and Translation parts are different. The Japanese test also does not include Summary Writing; and has both Japanese to Chinese and Chinese to Japanese sections in the Translation section. An overview of the sections of the test is shown in Table 3 and Table 4.

Table 3. Written Paper

Section	English, French, German and Spanish versions	Russian version	Japanese version
<i>Written paper</i>			
Listening	Questions 1-20. Usually 20 x 4-option multiple choice questions. Section A is questions 1-10; Section B is questions 11-20.		15 questions only: Questions 1-15
Linguistic Knowledge (Grammar and Vocabulary)	Questions 21-40. Usually 10 open cloze or 4-option multiple choice; 10 matching.		27 questions: questions 16-43
Reading	Questions 41-70. Section A is usually 15 x 4-option multiple choice cloze and focuses on vocabulary (q.41-55); Section B contains 11x 4-option multiple choice on three texts (q.56-66); Section C is 4 x matching (q.67-70).		20 questions: questions 44-63. Does not contain the Section A vocabulary-focused cloze.
Summary Writing	Question 71. Grades awarded for Content (A-F) and Language (A-F).	Not in Russian version	Not in Japanese version
Translation	Questions 72-75; four sentences to translate Chinese to target language.	Questions 71-75 (one extra sentence).	Questions 64-71. Four sentences Chinese to Japanese, and four sentences Japanese to Chinese
Guided Writing	Question 76. Writing task according to instructions given in Chinese. Grades awarded for Content (A-F), Language (A-F), Organisational Structure (A-F).		Question 72. Otherwise same as other versions.

Table 4. *Speaking paper*

Section	English, French, German and Spanish versions	Russian version	Japanese version
Speaking paper			
Speaking	2 x read aloud sentences marked at 1.0 or 0.5 each (Section A); 1 x read aloud paragraph marked at 2.0, 1.5, 1.0 or 0.5 (Section B); 2 x ask questions marked at 2.0, 1.5, 1.0 or 0.5 (Section C); 1 x talk about pictures marked at 3.0, 2.5, 2.0, 1.5, 1.0 or 0.5 (Section D).	2 x read aloud sentences marked at 1.0 or 0.5 each (Section I); 1 x read aloud paragraph marked at 2.0, 1.5, 1.0 or 0.5 (Section II); talk about pictures marked at 5, 4, 3, 2, 1 (Section III).	As English, French, German, Spanish
Listening-Speaking	4 x respond to short sentences marked at 1.0 or 0.5 each (Section A); 2 x questions based on a longer passage heard (first one marked at 2.0 or 1.0; second one marked at 3.0, 2.5, 2.0, 1.5, 1.0 or 0.5) (Section B).	5 x respond to short sentences marked at 1.0 or 0.5 each (Section IV); questions based on a longer passage (first one marked at 2.0 or 1.0; second one marked at 3.0, 2.0 or 1.0) (Section V).	As English, French, German, Spanish

Cambridge was provided with publicly available sample tests from the SMEEA battery of Gaokao language tests in order to gain an initial oversight into how the tests are organised. As a result, Cambridge staff (the co-authors of this article) were able to make an initial judgement of which of the CEFR's modes of communication each part of the SMEEA test could fit into. To avoid cognitive overload, the different parts of the SMEEA tests were loosely divided between the CEFR categorisations of reception, production, interaction, mediation and communicative competence. This enabled the project participants to take each element in turn by examining how the CEFR approaches each element, and how those parts of the SMEEA test can be mapped to the CEFR. This division was as summarised in Table 5.

Table 5. *Mapping the CEFR and SMEEA*

CEFR area	Parts of SMEEA test
Reception	Listening Reading (Sections B and C)
Production	Guided Writing Speaking
Interaction	Listening-Speaking
Mediation	Summary Writing Translation
Communicative Competence	Linguistic Knowledge (Grammar and Vocabulary) Reading (Section A)

Note that Section A of Reading (multiple-choice cloze) was included in the Communicative Competence section due to its focus on vocabulary items. Also note that some of these areas overlap; in particular, communicative competence elements of grammar and vocabulary come into all areas, and the communicative competence element of phonology comes into the Speaking and Listening-Speaking sections. The actual material used in the mapping workshops was chosen by SMEEA and not disclosed to Cambridge at any time. This is because this material is confidential, potentially live test material. One version of each language version of the test was chosen to work with, after being considered as an appropriate representative sample by SMEEA.

4 Overview of method

The project to map the SMEEA Gaokao language tests to the CEFR took place in autumn 2021. The original design of the study, developed in 2019-20, was adapted to consider the impossibility of international travel by Cambridge staff to Shanghai to lead the mapping activity due to the ongoing Covid-19 pandemic. Therefore, most of the interaction in the project was done virtually, through online workshops and remote communication using email and a Microsoft Teams space. The space hosted relevant documentation such as the CEFR booklets, and instructions for the activities.

4.1 People involved

The people involved in the activities in the project are listed in Table 6.

Table 6. *People involved*

Cambridge trainers	Two trainers from Cambridge developed the material for the activities and led the sessions with the lead mappers. They were responsible for collecting and analysing the data, and writing the final report.
SMEEA staff	Staff from SMEEA helped with the logistics for running the project and liaising with the lead mappers.
Lead mappers	Two lead mappers from each language were appointed from the staff and consultants of SMEEA based in China. They were experts in the Gaokao tests of their own language, as well as having a reasonably good level (B2+) of English in order to interact with the sessions from the Cambridge trainers. The lead mappers for each language group completed the familiarisation materials and attended the training sessions with the Cambridge trainers. In the sessions the lead mappers became familiar with the content and received help on how to deliver the same familiarisation materials and training content to the other people in their language group as a workshop. This kind of training is called cascaded training.
Language teams	Each of the six language teams, led by two lead mappers, had others in the language team who were SMEEA consultants based in China. These people received the cascaded training from the lead mappers and took part in the cascaded workshops to do the mapping.
Cambridge Partnership for Education (CPE) support staff	These members of staff, based both in the UK and China, supported the Cambridge trainers and SMEEA staff, as well as arranging contracts and other logistical issues. They also liaised with translators to make sure that key documentation was translated into Chinese for the benefit of the language teams.

4.2 Theoretical basis for method

The format of the activities in the project followed the broad lines as set out in the Council of Europe's *Manual for Relating Examinations to the CEFR* (2009) and the recent publication by the British Council, UKALTA, EALTA and ALTE – *Aligning Language Education with the CEFR: A Handbook* (2022). These documents set out the following phases as being important for any type of mapping or alignment of tests to the CEFR, and briefly how these stages were followed in the SMEEA project is presented in Table 7.

Table 7. *Stages of the project*

Stage	Description of stage (from p. 12 of the <i>Handbook</i>)	How the stage was carried out in the SMEEA Project
Familiarisation	Ensuring that all participants in the alignment process have a sufficient knowledge of the CEFR, its levels and descriptors.	Pre-workshop activities and training set by Cambridge trainers.
Specification	Describing the content of the test in relation to the categories of the CEFR.	Preparatory work by Cambridge as described in Section 4 of this report; decisions made by lead mappers as to which CEFR scales are relevant for each section of the test
Standardisation	Ensuring, through training, a common understanding of the CEFR levels.	Cascaded workshops aimed at a consensus among participants.
Standard setting	Determining judgements for assessment purposes.	The cascaded mapping workshops (an abbreviated form of standard setting)
Validation	Collecting and presenting appropriate evidence in support of alignment claims.	The analysis and explanation described in this report, written by Cambridge, to be used by SMEEA as appropriate.

4.3 Activities in the project

The actual procedure of the activities in the project followed the order of the CEFR areas given in Table 5, resulting in four main areas:

- reception
- production
- interaction and mediation
- communicative competence

Each of these areas comprised the following main stages:

- pre-workshop activities for the lead mappers
- a live online session led by Cambridge trainers with the lead mappers, providing training about:
 - the CEFR itself – its status, approach, and function;
 - discussion about which CEFR scales may be useful for the mapping process;
 - discussion about what the items and tasks in the SMEEA tests are targeting for testing;
 - how to map relevant sections of the SMEEA tests to the CEFR
- a cascaded session led by the lead mappers to the rest of their language team

- additional sessions as appropriate for the language team to map the test sections to the CEFR
- an informal Q&A live online session for lead mappers to raise any questions or issues with the Cambridge trainers
- the chance to revisit mapping judgements at a later date

Coupled with introductory and consolidation sessions, the activities in the project contained ten main sessions, which were each given separate channels in the online communication platform. All the relevant documentation was stored in the 'Files' sections of the platform. The major presentations, as well as accompanying notes for lead mappers to cascade the presentations with their language teams, was provided in both English and Chinese. More details on the activities in these stages are shown in Table 8. The consolidation workshop was the last formal opportunity to meet with the Cambridge trainers, and then language teams were asked to review all the mapping they had done with the aim of making a final group CEFR level judgement for each decision. After submitting their final judgements, lead mappers were asked to complete a closing survey to gauge their reactions to the activities in the project.

Table 8. Workshops and activities

<p>01 Welcome tasks</p>	<p>Lead mappers were asked to:</p> <ul style="list-style-type: none"> • introduce themselves in a post on the Teams area • watch a short video to get a basic understanding of the CEFR and post any comments • answer some questions about the CEFR • review the CEFR's 'Global Scale' and answer questions about it <p>Lead mappers attended a short introductory session live online with the Cambridge trainers which included:</p> <ul style="list-style-type: none"> • the opportunity to introduce themselves in person • the aims of the project • the role of the lead mapper • where to find the different language versions of the CEFR • how to use Microsoft Teams
<p>02 Introduction to CEFR workshop</p>	<p>Lead mappers attended a live online session with the Cambridge trainers introducing them to the CEFR. The session covered:</p> <ul style="list-style-type: none"> • the CEFR and how it sees language learning and assessment • the CEFR framework of levels, scales and descriptors • how the SMEEA language tests fit into the CEFR <p>Lead mappers then cascaded the session to their language team. The presentation and accompanying notes were provided.</p>
<p>03 Pre-workshop activities for receptive skills</p>	<p>Lead mappers were asked to:</p> <ul style="list-style-type: none"> • take the Listening and Reading sections of their test as if they were a candidate and make notes on what they notice • determine which CEFR scales would be relevant • complete a matching activity to familiarise themselves with potentially useful CEFR scales

04 Receptive skills workshop	<p>Lead mappers attended a live online session with the Cambridge trainers about the receptive skills. This session covered:</p> <ul style="list-style-type: none"> • the CEFR and receptive skills • applying the CEFR to the Listening and Reading sections • information on how to lead the language teams to map to the CEFR <p>Lead mappers then cascaded the session (and pre-workshop activities) to their language team. The presentation and accompanying notes were provided.</p> <p>Lead mappers facilitated the mapping of the Listening and Reading sections to the CEFR with their language team.</p> <p>The Cambridge trainers ran an informal Q&A session for any questions or comments the lead mappers had during the process.</p>
05 Pre-workshop activities for productive skills	<p>Lead mappers were asked to:</p> <ul style="list-style-type: none"> • take the Guided Writing and Speaking sections of their test as if they were a candidate and make notes on what they notice • review the assessment criteria for these sections • find examples of candidate performances of these tasks • determine which CEFR scales would be relevant • complete a matching activity to familiarise themselves with potentially useful CEFR scales
06 Productive skills workshop	<p>Lead mappers attended a live online session with the Cambridge trainers about the productive skills. This session covered:</p> <ul style="list-style-type: none"> • the CEFR and productive skills • applying the CEFR to the Guided Writing and Speaking sections • information on how to lead the language teams to map to the CEFR <p>Lead mappers then cascaded the session (and pre-workshop activities) to their language team. The presentation and accompanying notes were provided.</p> <p>Lead mappers facilitated the mapping of the Guided Writing and Speaking sections to the CEFR with their language team.</p> <p>The Cambridge trainers ran an informal Q&A session for any questions or comments the lead mappers had during the process.</p>
07 Pre-workshop activities for interaction and mediation Skills	<p>Lead mappers were asked to:</p> <ul style="list-style-type: none"> • watch a short video about interaction and mediation; • write down examples of interaction and mediation tasks they do personally in their everyday lives • take the Summary Writing and Listening-Speaking sections of their test as if they were a candidate and make notes on what they notice • review the assessment criteria for these sections • determine which CEFR scales would be relevant • complete a matching activity to familiarise themselves with potentially useful CEFR scales

<p>08 Interaction and mediation skills workshop</p>	<p>Lead mappers attended a live online session with the Cambridge trainers about the receptive skills. This session covered:</p> <ul style="list-style-type: none"> • the CEFR and interaction and mediation skills • applying the CEFR to the Summary Writing and Listening-Speaking sections • information on how to lead the language teams to map to the CEFR <p>Lead mappers then cascaded the session (and pre-workshop activities) to their language team. The presentation and accompanying notes were provided.</p> <p>Lead mappers facilitated the mapping of the Summary Writing and Listening-Speaking sections to the CEFR with their language team.</p> <p>The Cambridge trainers ran an informal Q&A session for any questions or comments the lead mappers had during the process.</p>
<p>09 Pre-workshop activities for communicative competence</p>	<p>Lead mappers were asked to:</p> <ul style="list-style-type: none"> • think about how communicative competence differs from communicative activities in the CEFR • take the Linguistic Knowledge, Reading Section A and Translation⁵ sections of their test as if they were a candidate and make notes on what they notice • determine which CEFR scales would be relevant; • complete a matching activity to familiarise themselves with potentially useful CEFR scales • look at other tools that would help vocabulary and grammar mapping: for English, the EVP and EGP; for French and Spanish, the CEFRLex
<p>10 Communicative competence workshop</p>	<p>Lead mappers attended a live online session with the Cambridge trainers about communicative competence. This session covered:</p> <ul style="list-style-type: none"> • the CEFR and communicative competence • applying the CEFR to the Linguistic Knowledge, Reading (Vocab) and Translation sections of the test • information on how to lead the language teams to map to the CEFR <p>Lead mappers then cascaded the session (and pre-workshop activities) to their language team. The presentation and accompanying notes were provided.</p> <p>Lead mappers facilitated the mapping of these sections to the CEFR with their language team.</p> <p>The Cambridge trainers ran an informal Q&A session for any questions or comments the lead mappers had during the process.</p>
<p>Consolidation workshop</p>	<p>Lead mappers attended a final live online session with the Cambridge trainers to:</p> <ul style="list-style-type: none"> • reflect on what work had already been completed • reflect on what work still needed to be done • discuss how to complete the mapping work by the deadline

4.4 The process of mapping to the CEFR

During the live sessions, the Cambridge trainers explained and discussed the process of mapping to the CEFR, which was then used by the lead mappers with their language teams to complete the mapping. This process remained more or less the same for each different section of the test. It

assumes the familiarisation and specification stages have already been completed (i.e. the participants are familiar with both the SMEEA tests and the relevant parts of the CEFR). For Reading, Listening, Linguistic Knowledge and Translation sections, mapping was done at the item level. For Writing and Speaking sections, mapping was done at the level of the assessment criteria: for each grade within each assessment criteria scale. To give the example of the Guided Writing section, there were three assessment scales: 'Content', 'Language' and 'Organisational Structure'. The grades available for each scale ranged from A to F (six possible options). Therefore, there were 18 judgements to be made (three scales multiplied by six possible options). When mapping, while the content of the writing or speaking task had to be investigated to find relevant CEFR scales, it was each grade of each assessment criteria scale that was mapped. The mapping process was advised to be carried out as follows:

Individual mapping. Each member of the language team decides individually:

- what each item² in the section is testing
- which CEFR scale(s) are relevant to look at for each item
- roughly which level each item is at
- reading the CEFR descriptors in the relevant scales, at, above and below the estimated level, for each item
- choose the CEFR level which is the best fit for each item. There may not be a perfect descriptor, but the aim is to find the nearest.
- For items focusing on vocabulary, the use of an additional tool (EVP, EGP or CEFRLex) may be helpful to find the CEFR level but should be seen as an aid rather than the answer to which level an item is at.

Pair discussion. In pairs, a discussion takes place:

- Each person explains which level each item is at, and why. They should justify their decision using CEFR descriptors.
- Each person listens to the other's opinion. The aim is to come to a compromise, i.e. one person might be convinced by another's justification and therefore change their mind.

Small group discussion. The process for the pair discussion is replicated at a small group level, e.g. four or five people. The aim is to find agreement between all members of the small group.

Whole group discussion. The process is replicated at the whole group level. The aim is to find agreement between all members in the language team.

Record judgement on spreadsheet. The decision as to which CEFR level each item is aiming at, is recorded on the Excel workbook spreadsheet. The group could give up to three CEFR descriptors (from the same or from different scales) for each judgement, as long as all descriptors were at the same CEFR level.³

Productive performance remarking. For Speaking and Writing mapping, another additional step was carried out:

- Candidate performances at different grades (especially the 'passing' grade) were reviewed.
- Each performance was marked again in two ways:
 - using the usual assessment criteria for the SMEEA test;
 - using CEFR levels and descriptors only
- The CEFR level given to the performance was compared with the CEFR level given to the grade, to

2 'item' here also refers to each grade in each assessment scale, for Writing and Speaking sections.

3 'Plus levels' (A2+, B1+ and B2+) were disregarded, so e.g. B1+ was regarded as the same as B1.

see if they were the same.

- If they were not the same, participants had to explore the reason for this, and if necessary adjust either the grade given to the performance or the CEFR level given to the grade.

5 Results

The results of the mapping are not shared here in terms of how the different versions compared for confidential reasons, but we are able to provide an overview of the methods used to collect and collate the data, and to give a general overview of the results.

5.1. Receptive skills

This section summarises what kind of information was collected and collated in order to provide results. The data collected comprised of tallies of the different CEFR scales used and the different CEFR levels chosen in the mapping of each section of the different language versions of the test, in terms of receptive skills. The number of times a CEFR scale or level was mapped were reported as a raw number of judgements and as a percentage of the number of judgements per language version. Results were collated in tabular form to allow for easier comparison across language versions. It was therefore straightforward to make an initial comparison between the language versions of the levels chosen in each section, and also overall, using simple descriptive statistics.

The average CEFR level (mean, median and mode) per language was also calculated to give a quick indication of comparison between the languages. The mean was worked out by assigning 1 point to Pre-A1, 2 to A1, 3 to A2, 4 to B1 and so on, and after calculating the average, converting it back to a CEFR level by rounding the mean to the nearest whole number. For example, a mean of 4.15 rounded down to B1. Any significant difference in distribution of each CEFR level across the different language versions was calculated at the $p < .05$ level using the chi-square test.

For Listening overall, the three most frequently mapped scales were *Overall oral comprehension*, *Understanding conversation between other people* and *Understanding audio media and recordings*. For Grammar and Vocabulary overall, the three most frequently mapped scales were *Grammatical accuracy*, *Identifying cues and inferring* and *Vocabulary range*. In the Cloze tests (Reading Section A), the most commonly mapped scale was *Identifying cues and inferring*. In the Reading (Sections B and C) test, the most commonly mapped scale was *Overall reading comprehension*, followed by *Identifying cues and inferring* and *Reading for information and argument*. In the Translation test overall, the two most frequently mapped scales were *Translating a written text in writing* and *Grammatical accuracy*.

Writing in the Shanghai Gaokao is marked according to assessment criteria. CEFR level judgements were made to map a CEFR level to each grade for each assessment criterion. Guided Writing and Summary Writing scores were calculated by asking the language teams to allocate a CEFR level to each of the points in their grading scale and converting these CEFR bands to a numerical value. It was often felt that it was not possible to map the lowest grade to the CEFR, because this grade indicates no, or almost no, performance displayed. In these cases, a zero (0) was allocated in place of a CEFR level. In the collation across language versions, a dash (-) was used to indicate the grades are not available within the mark scheme for that language version. Data was collated by criterion and across language versions. In a second collation, the number of times each CEFR level judgement (including 'zero' (0)) was chosen was made across language versions. As assessment criteria score across a range of possible marks, a simple tabulation indicated to what extent each language version had been mapped to a range of possible CEFR levels, and this was also displayed as a stacked bar chart, for a quick visual comparison. The bar charts showed the percentage of possible grades for two different types of writing task. A chi-square test was used to determine any significant difference in distribution of each CEFR level or for zero across the different language versions overall, at the $p < .05$ level. The mean average CEFR level per grade and per language, across all the

assessment criteria for Writing was also calculated. To do this, the CEFR level judgements are converted into a number, as described above, i.e. Pre-A1 = 1, A1 = 2, A2 = 3 and so on, and the mean average found. The numerical mean was converted back into a CEFR level as described above. In order to find out if the differences in the distribution of CEFR levels against possible scores for the grading scale for each language version was significant, we carried out a simple analysis of variance (ANOVA). This analysis takes a distribution of scores for different groups (in our case English, Japanese etc) and calculates whether the difference *between* the groups is greater than the difference *within* the groups. Analysis was also carried out for each of the criteria individually, to identify any significant differences between criteria.

In the Guided Writing test overall, a large variety of scales are mapped across all the language test versions. These include generic scales such as *Overall written production*; scales looking at specific aspects of writing such as *Thematic development* and *Coherence and cohesion*; scales for particular genres such as *Creative writing*; and linguistic scales such as *Grammatical accuracy*. This indicates that the Guided Writing test has a broad construct in all language versions of the test.

Only four of the language version contain a Summary Writing section. As with the guided writing test, the summary writing maps to a wide range of scales across all language versions, including scales which address technical skill in this area such as *Adapting language*, *Processing text in writing*, *Propositional precision* and *Streamlining a text*. In addition, this section maps to linguistic scales such as *General linguistic range* and *Grammatical accuracy*.

For Speaking, each question in each part is awarded a different number of marks, which makes it more difficult to make meaningful comparisons between language versions. To address this, an initial tally was made of how often each CEFR level was chosen, and then expressed as a percentage. The difference in distribution of CEFR levels across the different language versions was analysed for significance at the $p < .05$ level. In order to start to try to compare meaningfully one language against the other, when the tasks within the Speaking paper have different ranges of marks possible, the mean average CEFR level for the *highest* mark for each of the six tasks within the Speaking paper and the *half way* mark for each of the six tasks were collated and compared. Speaking and Listening-Speaking scores were calculated by asking the language teams to allocate a CEFR level to each of the points in their grading scale and converting these CEFR bands to a numerical value. This was done in order to compare the mean maximum scores and the mean mid-point scores across language versions. In order to find out if the differences in the distribution of CEFR levels against possible scores for the grading scale for each language version was significant, we carried out a simple analysis of variance (ANOVA). This analysis takes a distribution of scores for different groups (in our case Spanish, Russian etc) and calculates whether the difference *between* the groups is greater than the difference *within* the groups. The CEFR levels were converted into a numerical figure using the same procedure as described above, and then the mean average of all the highest marks was calculated, and the mean average of all the half-way marks was calculated. Simple analysis of variance (or ANOVA) was used to determine whether there was any significant difference between the language versions at the 0.05 level.

In cases where there was a significant difference, an additional post-hoc test was run to compare each language version against each other.

The Speaking test requires candidates to read sentences aloud, or describe a picture, and therefore it is unsurprising that this section maps to scales connected to production, to features such as pronunciation, intonation and stress, and to the ability to sustain a monologue, for all language versions. *Overall oral production* and *Overall phonological control* are the most commonly mapped scales, followed by *Sound articulation*, *Prosodic features* and two sustained monologue scales. Given the limited nature of the task design, it maps to a considerable number of scales, indicating a relatively wide construct for this section for all versions. The four most commonly mapped scales for the Listening-Speaking section of the test are the two generic scales of *Overall oral interaction* and *Overall oral production*, and two scales which reflect the task format: *Understanding an interlocutor* and *Conversation*.

6 Evaluation

This section evaluates the mapping process by reporting on the closing survey. It also acknowledges limitations in the study.

6.1 Closing survey findings

After having completed the judgements for all sections of the test, a closing survey was administered to the lead mappers to gauge their feelings on the mapping process as a whole. Nearly all the lead mappers took part in the closing survey, with the three lead mappers from the English team and the two lead mappers from all other teams except Spanish, where one lead mapper participated. In total therefore, 12 responses were collected, see Figure 1.

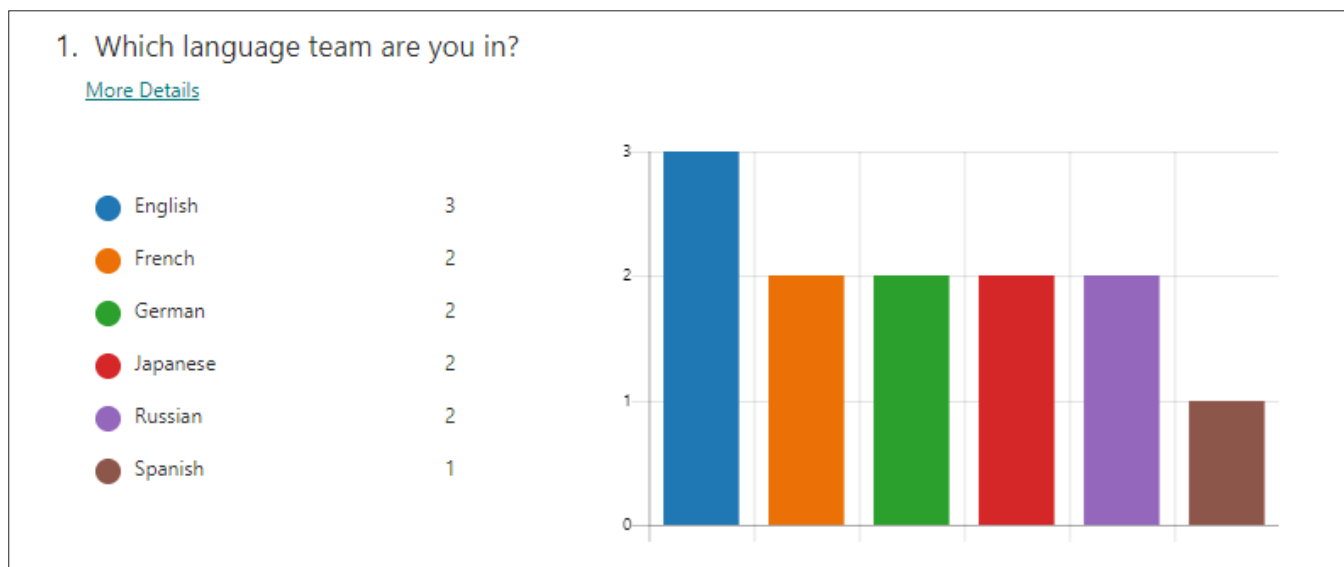


Figure 1. Which language team are you in?

These findings can be used to give an indication as to the success of the mapping project, particularly in terms of the aims to train SMEEA staff and consultants in using the CEFR, encouraging them to reflect on how the different language versions of the SMEEA Gaokao tests relate to the CEFR, and training them on mapping procedures to be able to use these techniques again in the future. The results of the closing survey responses will now be evaluated in the order that they appeared to the lead mappers, see Figure 2.



Figure 2. Feedback about the mapping process and timing of activities

Lead mappers were overwhelmingly positive about the mapping process and timing of activities within the mapping process agreeing to all of the following statements, with the vast majority **strongly** agreeing with them:

- The training provided helped me to understand the mapping process
- I understood how to complete the mapping spreadsheet
- The resources provided were helpful
- Graham and Jane [the Cambridge trainers] were helpful and knowledgeable
- The Q&A sessions with Graham and Jane were useful
- We had enough time between the mapping workshops to carry out the mapping in our language teams

- We had enough time between the mapping workshops to do the pre-workshop activities for the next workshop
- The workshops with Graham and Jane were the right length of time
- The Q&A sessions with Graham and Jane were the right length of time

This shows that the general format of the project, with pre-workshop activities, training input sessions, and cascaded sessions to the language teams, worked well and took place in a demanding but achievable timeframe, see Figure 3.



Figure 3. Feedback about the workshop experience

Additionally, almost all lead mappers agreed (or even strongly agreed) with the following statements about their experience in the project:

- The online environment (Microsoft Teams) worked well for the purpose
- I felt comfortable doing this activity in an online environment
- I felt comfortable taking the information from Jane and Graham and then planning the workshops with my language team
- I felt comfortable leading the workshops with my language team
- Overall, the experience was successful

Cambridge trainers originally intended to conduct the activities in a face-to-face environment as it was anticipated that it is easier to check participation and understanding in such a manner. Given the necessity of conducting the project in a virtual format, the responses to this question suggest that online activities could also work to fulfil the aims of the project, see Figure 4.



Figure 4. Feedback on familiarity with the test and with the CEFR

The fifth question explored whether the lead mappers' familiarity and understanding of the tests in question developed through the course of the project. While it is expected that the SMEEA lead mappers would have a certain amount of familiarity with their tests before the workshops started, a third of respondents said that they were only somewhat familiar, with one respondent of the opinion that they were not very familiar with the test before the workshops started. However, after the end of the project, all respondents reported being very familiar with the test. This shows that lead mappers became more familiar even with their own tests over the course of the project.

One of the key aims for the project was to increase familiarity and understanding of the CEFR. While only one of the lead mappers claimed to be very familiar with the CEFR before the start of the project, half of them said they were not very familiar, and one respondent being not at all familiar with the CEFR. Over the course of the project, familiarity with the CEFR clearly increased, so that by the end of the project, three quarters of the lead mappers stated that they were now very familiar with the CEFR, with the remaining quarter now responding that they were somewhat familiar. This is a pleasing result showing that this aim of the project has been achieved. The next two questions asked about the lead mappers' experience with the different parts of the SMEEA Gaokao test, cf Figure 5.

The perceived difficulty of being able to map the different parts of the test to the CEFR can give an indication as to the confidence level of the lead mappers, and shows where further training may be needed. The results showed a broadly similar picture across all sections, with both types of Listening sections being observed as the only sections which were not found difficult to map by anybody. The Speaking section, however, appeared to be the most difficult to map to the CEFR, although all sections had at least five out of the twelve respondents who found it easy, or fairly easy to map to the CEFR, see Figure 6.

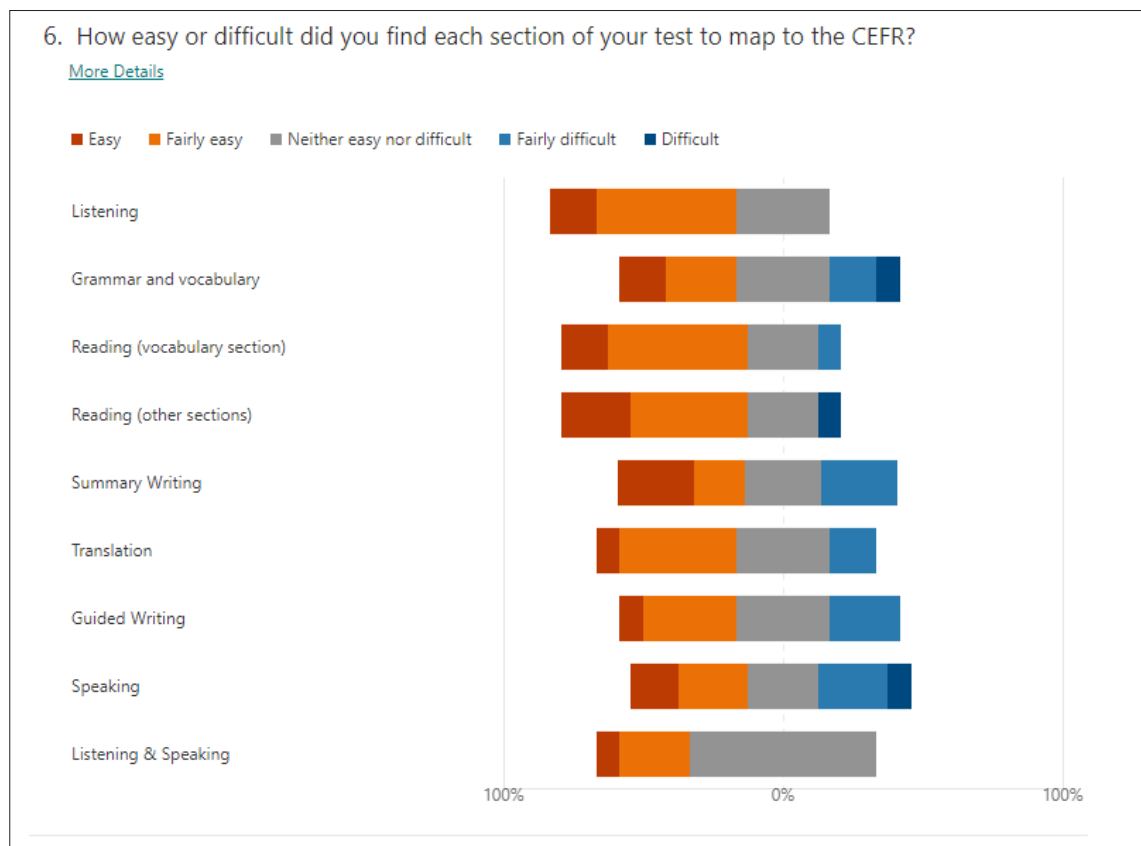


Figure 5. Feedback on perceived difficulty of carrying out the mapping

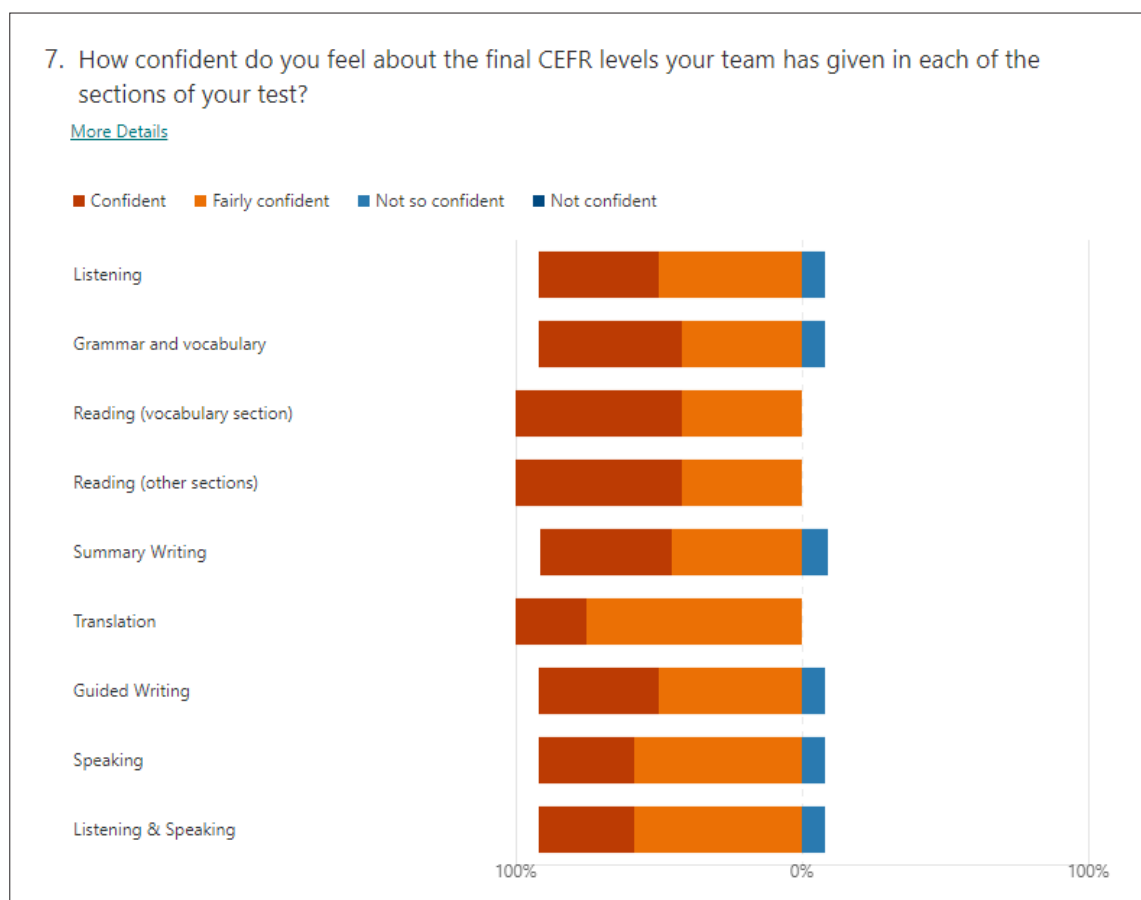


Figure 6. Feedback on perceived confidence in mapping ability

The following question showed how confident the lead mappers felt about the final judgements. For the most part, all respondents were confident or fairly confident about all the parts, with the Reading sections' judgements being the parts that had most confidence. In some parts, one respondent was not so confident in their judgements. Although opportunities were given for lead mappers to check understanding in order to aid confidence, further time could be devoted if necessary to avoid any uncertainty in judgements. Of course, it must be said that all judgements are by their nature subjective to a point, and it could be that this explains any potential lack of confidence. The final three questions required open-ended responses. The questions are therefore set out here, with a summary of the responses given. Respondents were allowed to answer the question in either English or Chinese.

Question 8. What have you learnt about your test? Are there any strengths or positive points you or your team have discovered by carrying out the mapping activity?

Many respondents commented that the test measures proficiency well, with a balanced structure assessing different aspects of language proficiency. The difficulty of the test was appropriate with useful, detailed scoring criteria, which lead mappers found out more about during the project. Respondents also took the opportunity in their responses to this question to praise the training and benchmarking activities, and one, for example, noted that it would affect their teaching techniques too by putting the CEFR's underlying theory in practice. Respondents also took the opportunity to note possible mismatches between the can-do focus of the CEFR and the selective focus of the SMEEA tests, which resulted in some difficulty in mapping, especially in Writing. One interesting perspective showed this difference:

'I feel that the CEFR focuses on affirming the content that students have mastered, while the test is mainly to find out students' deficiencies.' [translated from Chinese]

There was also an awareness of how the quality of the item can have an effect on the mapping:

'The language ability represented by the question itself does not match the ability required by the examinee to answer the question correctly. If the option is too simple, the answer will be correct regardless of the question.' [translated from Chinese]

Question 9. Is there anything you would like to change about the test in the future?

Some respondents to this question touched on considering changes in the writing and speaking assessment criteria to more fully reflect the CEFR approach. Some others thought that the Listening section was too easy. Overall, there was reflection on whether the fundamental basis for the test could be more in line with the CEFR's outlook: *'not based on errors but based on whether students can achieve a certain goal.'* [translated from Chinese].

Question 10. Do you have any other comments about the workshops or the process?

Some points for improvement were seen in comments about a better online platform to be used, the advantages of a face-to-face process, and increasing the time available for mapping. On the other hand, praise was given to the successful process as well as the Cambridge trainers. Respondents also commented on having a better understanding of the CEFR in order to be adequately familiarised to engage in the mapping process. Furthermore, there was again the desire to include CEFR principles within their own teaching. The actual process of coming to CEFR level judgements was seen as useful:

'Errors in judgement were inevitable, but we enjoyed the adequate and heated negotiation about these items.'

One respondent's comment sums up many of the others:

'After this benchmarking, I have a better understanding of the European standard framework, and I have a lot of reflections on my own teaching. I hope I can have more exchanges with experts face to face in the future.' [translated from Chinese].

6.2 Overall evaluation and limitations

The mapping process was completed successfully and satisfactorily in that it achieved its aims. The aims written in Section 2 are re-stated here, together with a justification as to why they have been achieved.

- to train participants about the CEFR and its uses. This aim was achieved by the workshops provided and the positive responses by lead mappers in understanding the CEFR and being able to cascade the information in order to then make mapping judgements.
- to encourage participants to reflect on how the specific parts of the tests relate to the CEFR and therefore map the six Shanghai Gaokao language tests to the CEFR. The project succeeded by not giving SMEEA the answer to mapping the language tests, but letting SMEEA staff and consultants take part in the mapping themselves. Not only were they able to make judgements to map the test to the CEFR both in terms of scales and levels, but also were able to provide reflections as to how the test did, or did not, align to the CEFR's "AoA".
- to observe the similarities and differences of the different language versions of the tests by comparing the six language test versions to each other, especially in terms of CEFR language proficiency levels. The results from the mapping enabled this aim to be achieved in comparing the language test versions through the analysis.
- to train up key people within SMEEA's staff and consultants in the use and application of the CEFR according to their context, so that they can carry out the mapping activity by cascading information to others; and also be able to plan and consider how to use the mapping tools and activities in the future. This aim is achieved by the nature of the workshops allowing the cascade of information from the Cambridge trainers to the lead mappers and then to the language teams. The results of the closing survey also show the success in training of the CEFR and its uses.

It is acknowledged again that the online virtual nature of the project, mandated by the coronavirus pandemic, was a limitation and lead mappers may have benefited from face-to-face training. In this way, the Cambridge trainers may also have been better placed to respond immediately to issues arising. In particular familiarisation with the CEFR and standardisation among participants with how to interpret the CEFR in a particular context is a prerequisite for CEFR mapping, and any deficiency in familiarisation may result in a less than accurate mapping. There is the potential therefore that differences in opinion of, for example, CEFR levels between languages are a result of different understandings of how to apply the CEFR rather than actual differences in test difficulty. It is hoped that the familiarisation activities employed in the project have limited this, but a face-to-face check may have been able to improve a standardised understanding of the CEFR more. It was the first time that SMEEA had carried out such a large-scale project, which involved more than 70 people and lasted over two months, blending online and offline delivery. The online communication platform facilitated meetings, sharing of materials and communication online. However, it also involved holding sessions which were early in the day for UK based presenters and late in the evening for Shanghai based staff, and there were sporadic connection issues. Connection failure does discourage participants who lack confidence working in a computer mediated environment. Organisations have to decide whether to opt for highly intensive training, with little time for reflection or cascading in between sessions, or whether to space out sessions, and weigh up the risks of participants going off track. These considerations apply both to offline and online modes of delivery. Organisations need to weigh up which styles and modes of delivery will be efficient, and what will be effective, and accept that these may differ.

The limitation that this is still a relatively small-scale CEFR mapping has already been acknowledged. A full alignment involving standard setting procedures would involve a much deeper investigation at each level for each language, and therefore would take a longer time and would likely be more costly. Nevertheless, it is believed that this mapping procedure has been useful in achieving its aims and has provided a satisfactory outcome of purpose for SMEEA.

Finally, a note on the languages used. The multilingual nature of the project is to be praised and also fits in well with the CEFR's multilingual and plurilingual approaches. If all participants would be able to communicate at a high level of proficiency in all seven languages used, it would enhance standardisation and a common understanding. As this is an unrealistic scenario, the language of the workshops with the Cambridge trainers was English and the common language of the cascaded workshops was Chinese, although language teams would also make use of the language of the test too. This limitation was mitigated with the translation of the materials from English into Chinese and the encouragement of all languages to be used, rather than 'forcing' English, for example, on to all participants. We also noted that for participants who are new to the CEFR, simply providing a version of the CEFR in a language they are more familiar with, such as Chinese, does not mean that their understanding of the terminology, particularly as used in the descriptors, can be taken for granted. SMEEA staff reported that the translated version sometimes complicated their understanding. A lack of examples in the descriptors made it hard for those who have no CEFR-related experience to understand or distinguish key terms.

7 Next steps

This section will suggest possible next steps for further alignment to the CEFR and improvement of test design. As stated, this mapping activity was not a full CEFR standard setting activity. Given that there are six language versions, and several different test sections, a standard setting activity would require considerable resources in terms of time and people. However, this may be a potential next step if a more accurate alignment is required.

As the project has shown some differences in level across the different language versions, work could be conducted to try to ensure a more balanced level of difficulty between languages. Appropriate item writer training may be needed, before the relevant teams rewrite test tasks as needed. A similar mapping activity to that used in this project would then need to be carried out, in order to ensure that the process had been successful in that the language versions were now more similar in terms of CEFR level.

In order to more fully reflect the CEFR's underlying theory of an action-oriented approach, a small-scale report could be commissioned to consider which elements of the test may need some rethinking, and how this might be achieved. This should be done in line with a needs analysis of the target language use domain, to ensure contextual validity. For example, the current test places considerable emphasis on discrete grammar and especially vocabulary parts, and indeed some of the Reading test is primarily concerned with vocabulary application. This may be appropriate, if the target language use context reflects this, but changes could be made to the tests to better reflect the CEFR. Furthermore, while the test involves some acts of interaction and mediation, these areas could be expanded.

Participants commented on the need to reflect on the current wording of the assessment criteria, and whether this is something that can be modified. Work could be commissioned to investigate this further, including how the CEFR could be utilised more in the assessment criteria, as well as in the tasks for the productive skills in the test.

In their later informal reflections, the SMEEA staff emphasised the importance of standardisation through training. They recommend that training should not take place only before the planned alignment starts but whenever any problems with mappers' understanding occur, and that training be closely tailored to the test that will be aligned. This of course needs to be balanced against security concerns and decisions over access to test content, if individuals outside the assessment organisation are involved.

Finally, it is always recommended that periodic mapping or alignment activities are carried out, or if there is a revision of the test. This is because different test versions contain different items, and the content and the difficulty may thus change over time. Where there are different language versions of the same test, the risk is greater that each language version may develop at different rates, and thus periodic mapping or alignment will help keep different versions broadly similar.

8 References

- British Council, UKALTA, EALTA and ALTE (2022) *Aligning Language Education with the CEFR: A Handbook*. Available from <https://www.alte.org/resources/Documents/CEFR%20alignment%20handbook%20layout.pdf> (accessed 20 December 2023).
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Available from <https://rm.coe.int/1680459f97> (accessed 20 December 2023).
- Council of Europe (2009) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. A manual*. Available from: <https://rm.coe.int/COERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d> (accessed 20 December 2023).
- Council of Europe (2020) *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume*. In English, available from: <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4> (accessed 20 December 2023).
- Council of Europe (2020) *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume*. In Chinese, translated by Xiangdong Gu and Zehan Chen (forthcoming).
- Council of Europe (2020) *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume*. In various other languages: <https://www.coe.int/en/web/common-european-framework-reference-languages> (accessed 20 December 2023).

9 Biographies

Jane Lloyd is a Senior Research Manager at Cambridge University Press and Assessment. Her current responsibilities involve working on standard setting and alignment of Cambridge English tests and advising other national test providers. Her interests include alignment to the CEFR and other external frameworks, validation and standardisation of tests and curricula at national level, and assessment literacy training for non-specialists. She is currently studying for a PhD in Language Testing at CRELLA, investigating the decision making of expert judges taking part in standard setting of English language reading tests.

Graham Seed is the European Research Manager in the Research – English team at Cambridge University Press and Assessment. His current professional interests include the operationalisation of the CEFR, and multilingualism in assessment contexts. He holds an MA in Language Testing from Lancaster University, CELTA and DELTA, and previously taught EFL in Bosnia, Russia and the UK before joining Cambridge Assessment as Senior Assessment Manager overseeing various test products. He is currently also the Secretariat Manager of ALTE – the Association of Language Testers in Europe.

Xu Wen is a project manager at Shanghai Municipal Educational Examinations Authority. Her current responsibilities involve working on designing and developing Shanghai English Gaokao tests and coordinating workgroups of other language tests of Shanghai Gaokao. Her interests include language test development and validation.

Hu Yue graduated from Fudan University with a Master degree in German Language and Literature. She is a staff member at Shanghai Municipal Educational Examinations Authority, where she is responsible for NCEE foreign language exams (German, French, Spanish, Russian and Japanese). Her research interests include language testing, translation, natural language processing, and any language-related technologies.