

Co-constructing the assessment criteria for EFL writing by instructors and students: A participative approach to constructively aligning the CEFR, curricula, teaching and learning

Language Teaching Research

1–29

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1362168820948458

journals.sagepub.com/home/ltr

Huahui Zhao 
University of Leeds, UK

Beibei Zhao
Zhejiang Shuren University, China

Abstract

The current assessment in language classrooms prevalingly utilizes the criteria provided by instructors, regarding learners as passive recipients of assessment. The current study drew upon sustainable assessment and the community of practice to highlight the importance of involving learners in co-constructing the assessment criteria and argued that using the criteria provided by instructors could lead to discrepancy between assessment, teaching, and learning. It adopted a participatory approach and investigated how to involve learners in co-constructing the assessment criteria with instructors in tertiary English writing instruction in China, based on the European Language Profile (ELP), an evolved version of the Common European Framework of Reference for Languages (CEFR). Two writing instructors and 146 tertiary students played different, yet interactive roles in adapting the assessment criteria in the local context. Instructors drafted the criteria in line with curricula, teaching, learning and learners. Learners utilized the draft criteria in a training session and suggested possible modifications to the criteria in a survey. Suggestions were used to revise the descriptors alongside teachers' reflections via reflective logs. A follow-up survey explored students' perceptions of the feasibility and usefulness of the modified descriptors to investigate the effectiveness of co-constructing the assessment criteria for learning and reveal further improvement if necessary. Vigilant decision-making processes were thickly described

Corresponding author:

Huahui Zhao, School of Education, Hillary Place, University of Leeds, Leeds, LS2 9JT, UK

Email: h.zhao1@leeds.ac.uk

regarding how assessment descriptors were selected, arranged, and modified to constructively align them with curricula, teaching, and learning. Statistical and thematic analyses were conducted to examine the accessibility, feasibility, and usefulness of the assessment descriptors prior to and after the modifications. Results substantiated the effectiveness and thus the importance of co-constructing assessment criteria for enhancing the quality of assessment criteria and developing learners' cognitive and metacognitive knowledge of writing and assessment. Implications for language tutors regarding co-constructing assessment criteria in local contexts were deliberated on at the end of the article.

Keywords

assessment criteria, Chinese EFL contexts, community of practice, constructive alignment, sustainable self-assessment, the CEFR

I Introduction

The current study explored how co-constructing assessment criteria could be employed to adapt the Common European Framework of Reference for Languages (CEFR) in tertiary EFL (English as a foreign language) writing instruction in China. It attempted to employ its evolved self-assessment instrument, namely, the European Language Profile (ELP; Figueras, 2012; Lenz & Schneider, 2004), to facilitate self-assessment in the target setting. The students were expected to select specific ELP-based 'I can' statements that best approximated their current level of writing proficiency. As such, the language learning process was intended to be transparent to learners and helped develop their learning autonomy via self-assessment.

The first key issue emerging from introducing the ELP descriptors into the target classrooms was how to adapt the ELP descriptors in the local instructional context. It is important to consider how the CEFR/ELP, which was designed as a common assessment framework across contexts, can be translated into context-relevant forms (Byrnes, 2007) because their application requires a shift in pedagogic routines to bring curricula, pedagogy, learning and assessment into productive interaction with one another (Little, 2007). Nevertheless, despite the wide application of the CEFR in different sectors of language education (Carson, 2016), its function as an assessment instrument in local contexts has remained under-explored (Jin et al., 2017; Runnels, 2014).

As far as the current setting is concerned, the existing assessment practice, similar to most instructional contexts globally, is predominantly driven by the criteria provided by the instructors. It emphasizes students learning how to be compliant with instructors' expectations with the assessment criteria produced by them (Tan, 2007). However, the drawbacks of utilizing the assessment criteria solely created by instructors have been articulated from theoretical and empirical perspectives. From the theoretical perspective, sustainable assessment (Boud, 2000; Boud & Soler, 2016) and the community of practice (Lave & Wenger, 1991) has highlighted the importance of instructors and learners co-constructing assessment criteria within its local contexts to make assessment sustainable and interactive. From the empirical perspective, using the assessment criteria offered by instructors leads to discrepant understanding of the criteria by instructors and students (Andrade & Du, 2007) and subsequent misalignment between teaching and learning.

These pitfalls suggest the necessity of searching for ways of facilitating learners participating in the process of co-constructing assessment criteria alongside their instructors in local settings. The participative approach aimed to adapt the ELP from students' and instructors' perspectives and reshape the prevailing teacher-driven assessment to develop learner autonomy and their assessment literacy. The results in this exploratory study are expected to provide implications for language educators regarding how to co-construct assessment criteria with students within local contexts so that the CEFR/ELP or similar common assessment frames could be effectively implemented locally.

II Theoretical and empirical support for co-constructing assessment criteria

The review of the relevant existing literature discloses the theoretical and empirical support for instructors and learners co-constructing assessment criteria. Below, we first stipulated the potential problems caused by using the assessment criteria solely constructed by the instructors. We then discussed how constructive alignment could guide the process of co-constructing assessment criteria, followed by implications for the current study.

I Theoretical support for co-constructing assessment criteria

Teachers providing the assessment criteria without involving learners in the construction process could either lead to learners' confusion about the meaning of the criteria or their different understanding of the criteria from their instructors'. This would subsequently generate different self- and teacher assessment results, different evaluation of learning achievement between learners and instructors, and consequently, disconnectedness between teaching and learning. If the primary aim of creating the assessment criteria is to 'provide ourselves, our fellow markers and our students, a shared sense of what we think is good work . . . we will need to open ourselves to the opportunities of our students to teach us about the standards, not just the other way around' (Bearman & Ajjawi, 2018, p. 7). Undoubtedly, it is important to provide students with the opportunities of asserting their understanding of good work; however, it is also essential for the students to beware of the standards of good work from instructors' perspective. The bilateral understanding of the standard is vital for promoting the partnership between learners and instructors in assessment. Co-constructing the assessment criteria is conducive to solve the different standards of good work held by learners and teachers (Andrade & Du, 2007) and address the potential problems of students constructing marking criteria based on what they are happy with and their existing yet developing knowledge of the current tasks (Orsmond, Merry & Reiling, 2000).

Using the assessment criteria provided by the instructors mistakenly considers learners as passive recipients rather than conscientious consumers of assessment (Higgins, Hartley & Skelton, 2002). It makes the assessment practice designed and dominated by the instructors which could discourage learners' active participation in assessment, generate misconceptions of their role in assessment, hinder their competence in carrying out shared assessment practice with instructors, and subsequently, stop them from

developing their skills as participants of assessment and hamper their full participation in assessment. In other words, instructors providing assessment criteria for self-assessment regards students as those 'receiving a body of factual knowledge' (Lave & Wenger, 1991, p. 33) and impedes the construction of lively assessment communities, consisting of learners and instructors in the instructional context.

The community of practice highlights active participation as an essential part of learning within a community to help newcomers to gradually make the culture of practice theirs and enable them to fully participate in the community (Lave & Wenger, 1991). To facilitate the process, a community of practice should be an active system wherein participants share understanding of what they are doing and what that means for their practice and community (Lave & Wenger, 1991). The community practice should also be a live curriculum which encourages community members to actively participate in its practice and develop their competence of conducting shared practice against conventions and standards within the community (Wenger, 2011).

In a language classroom, students are apprenticed to instructors to develop their assessment literacy alongside language proficiency. To construct a lively assessment community, students need to foster their shared understanding with their instructors about the purposes, the criteria, and the process of the assessment activity so that they could fully participate in it. They also need to make shared efforts to promote assessment for learning through regular interaction with instructors: discussing the assessment purposes, co-addressing concerns over self-assessment, co-developing assessment criteria, co-designing the process of assessment, and co-reflecting on assessment results.

Teachers as the sole agent to create the assessment criteria could lead to unsustainable assessment practice as it would induce learners to keep asking instructors for information regarding what and how to assess their performance. 'Learning cannot be sustainable in any sense if it requires continuing information from teachers on students' work' (Boud & Soler, 2016, p. 403). Sustainable assessment aims to 'meet the needs of the present and [also] prepares students to meet their own future learning needs' (Boud, 2000, p. 151). It requires the development of learners' skills in making informed judgement on their current and expected performance and taking those skills forward to their future professional practice (Boud & Soler, 2016). This resonates the community of practice regarding making education serve lifelong learning beyond schooling (Wenger, 2011). Developing learners' ability to construct the assessment criteria alongside instructors is an essential skill for sustainable assessment. It makes students understand how to negotiate with instructors about what they are expected to achieve (i.e. learning objectives), how it could be achieved (e.g. via making action-oriented assessment criteria) and how to evaluate where they are (e.g. using the action-oriented criteria to reflect on their current performance). As Boud and Associates (2010) suggested, assessment has the most effect when students and teachers become responsible partners with students progressively taking responsibility for the assessment process and developing and demonstrating their abilities to make sound judgment of their work.

2 Empirical support for co-constructing assessment criteria

The benefits of learners and instructors co-constructing the assessment criteria have been substantiated in other instructional settings. Orsmond et al. (2000) noted that students

constructing criteria with support from instructors developed their sense of ownership of the meaning and effective use of the criteria in marking (e.g. appropriate weight of each category). The ownership could shape learners' beliefs in assessment and willingness to participate in assessment activities (Nelson & Schunn, 2009). We argue that the ownership could also cultivate their motivation for participating in assessment intrinsically (as students understand why and how they are doing assessment to support their learning) and extrinsically (e.g. students know how to go about their future study based on assessment criteria and following results). It can also develop learners' belongingness to the assessment community (i.e. integrated motivation) and foster their shared assessment literacy (e.g. how and why assessment should be designed and conducted) with instructors within the local instructional context. This echoes the notion of situated learning emphasized by the community of practice: Learning is a social practice which is shaped by where and when learning takes place with whom (Lave & Wenger, 1991). This is particularly important when a unified framework such as the CEFR is applied in a local context, considering learning and teaching varying from contexts. The social nature of assessment has been substantiated by the misalignment of the CEFR with the local syllabi (Zou & Zhang, 2017) and instructors' difficulties in understanding them within their local teaching contexts (Zheng, Zhang & Yan, 2016).

The theoretical and empirical evidence has stressed the importance of avoiding driving assessment with the criteria offered by instructors. This study adopted a participatory approach to involving students in producing co-constructed assessment criteria, aiming to foster their belongingness to the assessment community and develop their assessment literacy via negotiating and understanding the purpose, design, conduct and use of assessment for their current and future study and professional practice (i.e. sustainable assessment). The specific steps of designing the co-constructed criteria were guided by constructive alignment.

III Co-constructing assessment criteria through constructive alignment

Constructive alignment emphasizes the congruence between assessment and characteristics of the learning environment in order to make assessment focus on what learners should be learning and serve for learning (Biggs, 2003). Dochy, Segers, Gijbels and Struyven (2007) argued that constructive alignment made instruction and assessment more integrated and learners more active in sharing responsibilities of developing the criteria and the standards for evaluating their performance. Zou and Zhang (2017) reported that the CEFR descriptors related to writing activities required by the syllabus were perceived by teachers to be easier and more applicable than those not required or less frequently used and practised in writing classes. Gielen, Dochy, & Onghena (2011) stated that while co-developing the assessment criteria, teachers considered the criteria from the pedagogical perspective (e.g. syllabi) and learners cogitated them from the learning perspective (e.g. learning motivation and objectives).

The current study encouraged students and teachers to take into consideration syllabi (e.g. teaching objectives and learning outcomes), teachers (e.g. workload, perceptions of assessment and teaching styles), and learners (e.g. language proficiency and assessment

experience) when co-constructing the assessment criteria to achieve productive interaction between assessment, curricula, teaching and learning within the local assessment community. Such a process offered the instructors the opportunities of reflecting on their assessment design from learners' perspectives and modifying it to better meet learners' needs. As such, assessment based on teaching could constructively align with learning and learners. The process guided by constructive alignment made students associate their learning outcomes with teaching objectives and identify learning support that is needed based on shared assessment criteria. It facilitated students in (1) understanding the criteria (i.e. accessibility), (2) using the criteria to prepare and assess their assessment tasks (i.e. feasibility), and (3) applying the criterion-based assessment results to plan future study to fill in the knowledge gap between what has achieved and what needs to be achieved (i.e. usefulness). These three facets guided the design and evaluation of the co-constructed assessment criteria, intending to generate comprehensible, actable and helpful information for students about their current and future learning tasks.

IV Design of the pre-modified assessment criteria

The design of the pre-modified criteria aimed to align assessment with curricula, teaching, learning, teachers, and learners (Appendix 1). They were shaped by the local teaching and learning culture, teaching/learning activities (e.g. the syllabus, teaching objectives and learning outcomes), and participants (teachers and learners).

1 Choosing the relevant ELP descriptors

The two writing instructors and the researchers selected the ELP descriptors to construct the draft assessment criteria for summaries and argumentative essays. Students were not involved in this process for three reasons.

One, the host institution required the writing instructors to create the teaching materials before the semester started so that they could cope with other workloads in the forthcoming term including pastoral support for students. Two, the writing instructors suggested that the tight timeframe to complete the syllabus could not afford the class time to construct the criteria from scratch with students. Furthermore, students' limited knowledge of assessment and ELP descriptors and their unfamiliarity with the teaching objectives would make the process of co-constructing the criteria challenging and time-consuming. Three, students' perceptions of the assessment criteria would be unreliable without grounding their opinions on their use of the criteria. Therefore, it would be more practical, efficient, and reliable by inviting students to comment on the draft assessment descriptors based on their experience of using them and modified them accordingly afterwards.

2 Constructing the pre-modified criteria: macro- and micro-aspects

Summaries and argumentative essays were the two genres included in the syllabus; therefore, the pre-modified criteria consisted of macro- (i.e. how to structure summaries and argumentative essays) and micro-aspects (i.e. language use in summaries and

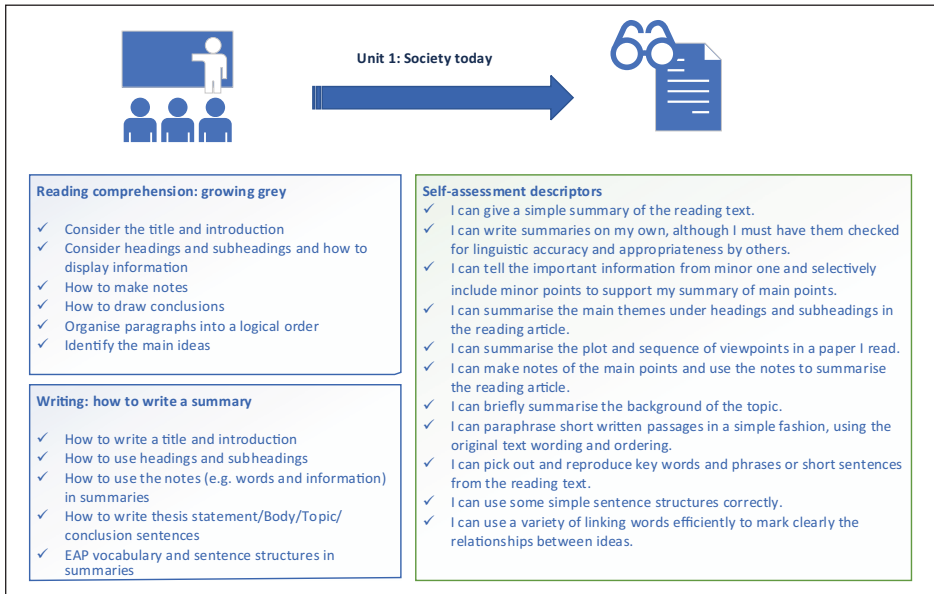


Figure 1. Constructive alignment among pre-modified assessment descriptors, teaching and learning.

argumentative essays). Macro-descriptors were mainly originated from the ELP descriptors on essays and reports, considering the relative similarity of genres. Descriptors of the language use were selected from ELP descriptors of written interaction and language competence/linguistics in accord with language instruction in sessions. In addition, the writing instructors created three new ‘I can’ descriptors and adapted two descriptors from the ones on argumentative essays for summaries to match their writing instruction (Appendix 1). Figure 1 exemplifies how the pre-modified assessment criteria were created to align with curricula, teaching, and learning, using Unit 1 as an example.

The left two boxes summarize the foci of the reading and writing sessions of Unit 1. The right box lists sample descriptors in the pre-modified assessment criteria that encouraged students to assess learning outcomes related to the left boxes. As such, self-assessment was served as a tool for learners to reflect on what was taught, whether they had attained the learning outcomes, and what they should go about their study next. Self-assessment and teacher assessment using the same criteria could also be served as the tool for the instructors to reflect on how well the students had learned and what they could do to support the students to achieve learning outcomes.

3 Selecting the language proficiency levels of the descriptors

Descriptors at four levels of English language proficiency (A2 to C1) were selected to make the criteria inclusive and facilitative for learners. The two instructors suggested that some of their students were at A2 whereas a few students were at C1, although the

majority of students were at B1–B2 in general. However, the students' proficient level also varied from specific assessment aspects. For instance, most students were believed to possess a higher level of spelling and grammar (e.g. subject–verb agreement) than structuring essays. Therefore, it would make the draft criteria more inclusive by selecting descriptors across the four proficient levels. Additionally, including A2 descriptors was also perceived to motivate students to keep up their good work with what they had achieved. The inclusion of C1 descriptors could help students to envisage their next learning objectives.

After selecting the descriptors, they were collated and presented in a mixed order to avoid students classifying the levels of descriptors based on their orders and giving the same rating to the ones belonging to the same level without careful consideration of each descriptor and their writing performance. This would be conducive to develop evaluative skills, an essential skill for future study and beyond.

4 Using emoticons to suggest achievement levels

Emoticons rather than numbers were used for students to reflect on their writing proficiency: 😊 standing for *achieved*, 😐 standing for *nearly there* and 😞 standing for *not there yet*, due to mainly three reasons. One, the emoticons have been widely used in students' daily conversations in social media to express their reflections of daily life. Therefore, it was reasoned that they should be more familiar and friendlier to use emoticons than the numbers to assess how well they had done in their writing. Two, this study intended to make students aware that self-assessment should be part of their learning processes like their momentary reflection of daily life. This message could be more easily delivered via using emoticons than numbers as students used the former in their daily life reflections. Finally, numbers have been dominantly used in summative assessment (i.e. to evaluate what has been learned). Using numbers in the grid could mislead students to consider self-assessment as rating the quality of their own writing. This conflicted with self-assessment in this study as formative assessment (i.e. providing information for future learning) to facilitate the process of learning.

All in all, using emoticons connected self-assessment to learners' daily reflections, made the criteria aligned with learners, and would hopefully make self-assessment more authentic, learner-driven, and sustainable as lifelong learners.

V Training in self-assessment

All students had limited experience in self-assessment due to the traditional teacher-driven and examination-oriented learning culture in China (Zhao, 2018). Training is also expected by sustainable assessment to gradually develop learners' knowledge of and skills in self-assessment (Boud & Soler, 2016). A one-hour training session was provided to encourage full participation in the assessment community, develop learners' assessment literacy, and address learners' insufficient knowledge of the ELP descriptors and self-assessment (Little, 2009).

Phase One of the training focused on developing shared understanding and practice of assessment between instructors and learners. The instructors encouraged students to

confer their understanding of self-assessment, its advantages and disadvantages, and possible ways of undertaking it. The instructors referred to the existing literature about benefits and limitations of, and training in, self-assessment to develop the class discussion. The instructors and the students reached a consensus about the three-step self-assessment in the end: (1) reading and comprehending assessment descriptors, seeking the instructors' support if necessary, (2) reading their writing, and (3) selecting the descriptors that best approximated their reflection of their writing. The phase continued for about 25 minutes.

In Phase Two, the instructors briefly introduced the CEFR and ELP and their use and popularity in education globally. They also highlighted its similarities to the forthcoming China's Standards of English Language Ability (Jin et al., 2017) to make students understand why the ELP descriptors were chosen as the assessment criteria. The discussion was expected to motivate students to engage in the ELP-based assessment. Students were invited to ask questions about the CEFR and ELP. The phase lasted about 7–10 minutes.

In Phase Three, the instructors thought aloud to demonstrate how they used the ELP descriptors to assess a student's summary to develop shared assessment practice with students. They articulated their understanding of each descriptor and their decision-making processes of selecting the emoticon related to each descriptor. The instructors reiterated the importance of understanding the descriptors and encouraged students to seek help with comprehension of the descriptors if needed. The students then used the pre-modified descriptors to assess their summaries (about 250 words written in Week 1) for approximately 15 minutes. They were offered support to understand descriptors when required. The whole stage lasted for about 30 minutes.

Although extensive training was only provided once, the instructors provided ongoing support for students. They used projects to explain the descriptors as the whole class, instructed in the use of the descriptors for self-assessment, and helped individual students understand the assessment descriptors in each self-assessment session.

VI Questions to guide the co-construction process

The pre-modified assessment criteria were utilized in the training session to elicit learners' and instructors' perceptions of their accessibility, serving as the basis for the co-constructed criteria. Three questions were asked to guide the co-construction process, seeking to improve the accessibility, feasibility, and usefulness of the ELP-based assessment descriptors and develop shared interpretation of the descriptors, and hopefully reached shared understanding of learning achievements between learners and instructors.

1. How accessible were the pre-modified ELP-based assessment descriptors perceived by the students and the instructors based on their experience of using them?
2. What modifications were suggested by learners and instructors to improve the accessibility of the pre-modified assessment descriptors?
3. How feasible and useful were the modified descriptors perceived by the students and the instructors?

Table 1. Overview of the syllabus.

Week	Focuses of each session
Week 1	<ul style="list-style-type: none"> • Overview the content over the academic term • Introduce key skills in EAP [English for academic purposes] reading and writing
Week 2	<ul style="list-style-type: none"> • Conduct a training session on self-assessment • Mock self-assessment of summaries with pre-modified ELP descriptors • Elicit students' perceptions of pre-modified descriptors
Week 3–5	Unit 1 Society today: two reading sessions & one summary writing session and self-assessment
Week 6–8	Unit 2 Food security: two reading sessions & one summary writing session and self-assessment
Week 9	Unit 3 Sustainable energy: reading session 1
Week 10–11	mid-term break: revision and exam weeks
Week 12–13	Unit 3 Sustainable energy: one reading session & one argumentative essay writing session and self-assessment
Week 14–16	Unit 4 Sustainable fashion: two reading sessions & one argumentative essay writing session and self-assessment
Week 17	<ul style="list-style-type: none"> • Review the content covered in the whole semester • Reflect on teaching and learning, including self-assessment

These three questions were answered when the syllabus (Table 1) was delivered. Each unit consisted of three sessions across three weeks, including one writing session. Each writing session lasted for 90 minutes with a 10-minute break. Self-assessment took place in the second half of the writing session, lasting for approximately 15 minutes. The first question was answered after students used pre-modified ELP descriptors in the training session in Week 2. This provided the basis for modifications to the draft descriptors asked in RQ2. Students' and instructors' views of the feasibility and usefulness of the co-constructed criteria were elicited after their use for four writing tasks, aiming to make continuous improvement of the co-constructed assessment criteria.

VII Eliciting students' and instructors' views to facilitate co-constructing the assessment criteria

To integrate students' and instructors' perceptions of the co-constructed assessment criteria, two surveys were developed to elicit learners' perceptions of (1) the pre-modified descriptors to improve their accessibility, and (2) the modified descriptors for continuous improvement of feasibility and usefulness. Reflective logs were employed to elicit the two instructors' views based on their classroom observation and experience of using the pre-modified and modified descriptors.

1 Students' and instructors' background

Two writing instructors and 146 second year university students in China participated in the study for one semester on a voluntary basis. They were Chinese and spoke English as

Table 2. Student participants' background.

Class ID	Number	Gender	Major	Final writing marks (average)
1	35 (taught by Tutor 1)	Male: 16 Female: 19	Network media	69.69 (SD = 7.66)
2	35 (taught by Tutor 1)	Male: 13 Female: 22	Network media	67.79 (SD = 7.17)
3	29 (taught by Tutor 2)	Male: 8 Female: 21	Public Management	71.31 (SD = 7.79)
4	47 (taught by Tutor 2)	Male: 3 Female: 44	Chinese Language and Literature	75.98 (SD = 6.22)

a foreign language. One tutor held a master's degree in English Literature whilst the other held a doctoral degree in language education. Both had been working at the university for more than 10 years and taught the academic reading and writing integrated module since its introduction in 2016 as an optional module.

The student participants were from three majors. They had been learning English for more than 10 years, with a low intermediate level of English language proficiency (an approximate B1–B2), judging by their university entrance English exam scores and instructors. Table 2 summarizes their background information.

Class 1 and 2 consisted of a more balanced number of female and male students than Class 3 and 4. Descriptive analysis and an ANOVA test of average final English writing scores¹ showed no significant differences among Class 1–3. Independent t-tests showed that students in Class 4 had a 5 point significantly higher average writing score than the rest of the three classes. The different writing scores across classes were considered when the data were interpreted.

2 Eliciting students' perceptions of pre-modified ELP descriptors

Learners' perceptions of the accessibility of pre-modified descriptors were mainly investigated in a survey. It consisted of two sections and nine questions (Appendix 2). Section one gathered student background information to identify its impact on the perceived accessibility so that adjustment might be made to create more inclusive criteria. Section Two asked students to indicate how easy they could understand descriptors on a four-point scale from extremely easy to extremely difficult. The short questionnaire was written in English as the instructors wished to create additional English learning opportunities and they were confident in their students' capability to understand the questions in English. They also provided explanations of the survey before distribution as requested by the researchers. The survey results were used to modify the draft criteria to increase accessibility.

3 Eliciting students' perceptions of the co-constructed modified descriptors

Students' perceptions of the modified assessment descriptors were investigated with a survey with a mixture of closed and open-ended questions after they used them for two

summaries and two argumentative essays. Three questions sought information about (1) feasibility on a five-point scale, (2) justifications of their selection, and (3) suggestion on improving feasibility. One question was asked about usefulness with four options: extremely useful, very useful, useful, and not useful. It was followed by an open-ended question eliciting justifications of responses related to usefulness. The last open-ended question collected other information the respondents wished to add about the co-constructed assessment criteria. The questionnaire in English was explained by instructors and administered in Week 17.

4 Eliciting instructors' perceptions of assessment criteria via reflective logs

Logs were used for the instructors to record, reflect on and evaluate the quality of the pre-modified and modified descriptors, based on their classroom observations (e.g. student questions) and their use of the criteria for teacher assessment of the same assignments. Logs are a type of diaries/journals which 'contain observations, feelings, attitudes, perceptions, reflections, hypotheses, lengthy analysis, and cryptic comments' (Hook, 1985 cited in McKernan, 1996, p. 84). Reflective logs create flexibility in time and space. This was particularly important for the two instructors who were heavily involved in administration apart from teaching. Compared with interviews, reflective logs could be completed at their convenient time, thus supporting longer and possibly deeper reflection of their observations and experience.

The less interactive nature of logs than interviews was addressed by prompts and follow-up conversations. Prompts were provided for the instructors to (1) minimize the distortion effects of assumption, biases and recollection on the accuracy of log entries, and (2) steer teachers to provide similar information to what was elicited in student surveys. The prompts consisted of:

1. How do you think of the role of self-assessment for writing?
2. Can the students understand the descriptors based on your observations? What questions did they ask regarding the descriptors?
3. Can you understand the descriptors when you used them for teacher assessment? What are your difficulties in understanding them if there is any?
4. How should the descriptors be modified to promote its accessibility and feasibility based on your reflections?
5. Is it useful for your students to use the ELP descriptors to assess their writing? If so, in which ways are they useful?
6. Would you like to add any other comments regarding the ELP descriptors?

The prompts were explained in the introductory meeting with the instructors where the popularity of logs in professional development and methods of keeping logs were discussed. Suggestions were provided to the instructors about how to keep a log, including:

- make good use of prompts
- enter logs about students' experience of the descriptors at the earliest convenient time after each self-assessment session

- enter logs about their own experience at the earliest convenient time after they used the assessment criteria for teacher assessment
- take onsite notes of what occurred related to the criteria and using them as examples to support opinions in logs
- conduct continuous observations and reflections of experience to assist log entries.

The same prompts were used for the pre-modified and modified descriptors. Reminders were sent to the instructors after each self-assessment session. Instructors could use both English and Chinese to enter their reflective logs. Instructor 1 used Chinese whilst Instructor 2 used English only. To remedy the limitations of unidirectional log data, clarifications of meanings of entries (e.g. what made you think students' low commitment affected the usefulness of self-assessment?) and confirmation of the researcher's understanding of their log entries (e.g. do you mean the use of descriptors develop their understanding of good writing?) were sought after receiving each reflective log.

VIII Integrating students' and instructors' perceptions into the co-constructed assessment criteria

Based on the students' and instructors' perceptions of the pre-modified and modified descriptors, the assessment descriptors were evolved to enhance their accessibility, feasibility, and usefulness.

1 Students' perceived accessibility of pre-modified ELP descriptors

A relatively low level of accessibility was suggested about the pre-modified ELP descriptors, suggesting the necessity of modifications. Based on the 143 respondents, about 20% more students reported that the pre-modified descriptors were difficult to understand (i.e. 57.3%), compared with 37% of respondents who claimed that the draft descriptors were easy to understand. Only a few students evaluated the pre-modified descriptors as being either extremely difficult or extremely easy to understand (see Figure 2).

A Kruskal–Wallis test suggested significant differences in the perceived accessibility across classes ($H = 10.78$, $df = 2$, $p = .013$). However, further ad-hoc pairwise comparisons showed that the significant differences only existed between Class 1 and 2 from the same subject. Considering the value for each scale (extremely easy = 1, easy = 2, difficult = 3 and extremely difficult = 4), the mean rank of accessibility for the four classes (54.65 for Class 1, 82.09 for Class 2, 75.79 for Class 3 and 74.41 for Class 4) suggested that Class 1 reported the highest level of accessibility whilst Class 2 reported the lowest level of accessibility. The results aligned with the descriptive data in Figure 2. Considering the lowest writing proficiency of Class 2 among the four classes (Table 2), the result might suggest the importance of taking into consideration the language proficiency when the assessment criteria were designed and introduced.

To improve the accessibility of the pre-modified descriptors, the students were asked to identify difficult items and explain why they were difficult to understand and how to improve them. Table 3 shows that Items 11 and 23 were the most difficult items for the students, followed by Items 10 and 21.

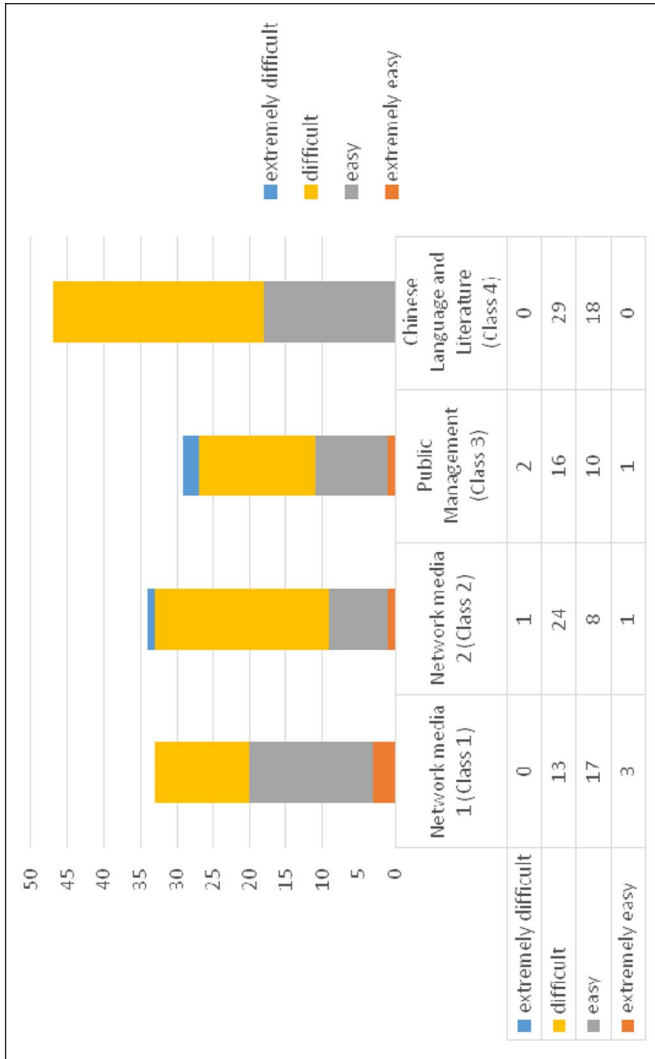


Figure 2. Perceived accessibility across four classes.

Table 3. Difficult items which were mentioned more than 10 times.

Difficult items	Frequencies
Item 11: I can write with <i>reasonable</i> grammatical accuracy and can correct mistakes if they are identified by others.	25
Item 23: I can convey simple information of <i>immediate relevance</i> , getting across which point I feel is the most important.	25
Item 10: I can have good control of <i>elementary vocabulary</i> , but major errors still occur when expressing more complex thoughts or <i>handling</i> unfamiliar topics and situations.	23
Item 21: I can produce <i>continuous</i> writing, which is generally <i>intelligible</i> throughout.	19
Item 5: I can write summaries on my own, although I must have them checked for <i>linguistic accuracy</i> and <i>appropriateness</i> .	17
Item 18: I can link a series of shorter, <i>discrete</i> , simple elements into a connected, <i>linear sequence</i> of points.	13
Item 9: I can <i>briefly</i> summarize the background of the topic.	11
Item 4: I can tell the important information from <i>minor</i> one and selectively include minor points to support my summary of main points.	10

Unspecified or vague wording was presented as the main threat to their understanding of the draft descriptors. Respondents identified the words in italics in Table 3 as ‘being vague’, ‘complicated’, ‘rarely heard’ and therefore difficult to understand their meanings. Statements like ‘I can’t understand the meaning of these difficult words’ were reiterated. Knowledge of writing also affected their understanding of the descriptors. Fifteen respondents stated that they struggled to distinguish different aspects of writing in the criteria because they were never explained about them, e.g. the difference between linguistic accuracy and appropriateness. Seven respondents thought their limited understanding of how to write summaries leading to their difficulties of understanding the descriptors. Language proficiency was presented as another main reason: Fifteen respondents indicated that their low English language proficiency caused their difficulties, including their ‘poor’ grammatical knowledge and limited vocabulary.

To improve the accessibility of the pre-modified descriptors, 69 among the 115 respondents conveyed developing their English language proficiency as the best way to improve their understanding of the descriptors (e.g. increasing vocabulary sizes). Thirty-seven recommended replacing difficult and vague words with simple or common words. Students also advised to use bilingual descriptors, reduce the number of descriptors, and give examples of each descriptor.

2 Instructors’ perceived accessibility of pre-modified ELP descriptors

The two writing instructors echoed most of the students’ views in their reflection logs. Both identified unfamiliar wording as the main reason impeding the accessibility of descriptors as all student questions asked in class were about the vocabulary used in the pre-modified descriptors. Lack of exposure to authentic English expressions of writing

was perceived as another major reason for students' difficulty in understanding the descriptors. Both instructors considered their students' low English proficiency affecting the accessibility of descriptors. Instructor 2 believed that difficulties only resided in a minority of learners with low English proficiency who seemed to spend a longer time selecting the emoticons for certain descriptors than their peers. Instructor 1 thought that the descriptors were difficult for most students considering their low English language proficiency as a whole. Their different concerns could possibly be due to the slightly higher English writing proficiency of Class 3–4 taught by Instructor 2 than of Class 1–2 instructed by Instructor 1 based on the final writing scores reported above.

Both instructors suggested reducing the number of descriptors as they observed that a small number of students could not finish assessing their summaries within the allocated time; therefore, a shorter grid should be covered in each session. Instructor 2 commented that similar items placed closely to each other could increase the difficulty level of understanding because students sought clarifications of their differences in class, such as items: (1) I can summarize the short text by using words from the original reading text, and (2) I could find key words, phrases and short sentences in the original reading materials and use them to summarize the short text. She also mentioned that requiring students to figure out the subtle differences between similar descriptors should be minimized within the limited available class time for self-assessment. By contrast, Instructor 1 believed that placing items with similar focuses could help students to distinguish their differences and provided more accurate evaluations. The conflict viewpoints suggested retaining the order of items but promoting the accessibility of them with other approaches.

3 Applying students' and instructors' suggestions to modify ELP descriptors

After synthesizing students' and instructors' views on the pre-modified descriptors, two main changes were made to the draft descriptors: created bilingual versions of descriptors and separated descriptors across sessions in terms of instruction focuses.

The researchers and instructors agreed that creating bilingual versions of descriptors was a fair compromise to address both students' and tutors' concerns. For one thing, the instructors believed that reading the descriptors in English helped the students to develop their English proficiency (e.g. learning new and authentic vocabulary related to writing and motivating them to increase vocabulary). This would also address students' limited exposure to authentic English language use, one reason for comprehension difficulties of the descriptors. Therefore, the English version should be retained. For another thing, the two instructors deliberated on the difference in Chinese and English which would make it hard to find equivalent alternatives to the problematic wording (e.g. the terminologies related to writing). Different vocabulary knowledge among individual learners would mount the challenge of replacing words. Therefore, a Chinese version could be created to serve a secondary role in assisting students' comprehension of the descriptors in English. Furthermore, they believed that students' difficulties caused by their limited language knowledge could not be addressed within a short period; therefore, Chinese translation could be served as temporary solutions to solving the difficulties caused by unfamiliar words/expressions. Although a bilingual version might distract learners from

the English descriptors, the accessibility of the descriptors should be prioritized as the main aim of self-assessment was to encourage learners to reflect on their learning.

To reduce the number of descriptors in each session but align with the syllabi, the modified descriptors were separated into two assessment grids on macro- and micro-aspects, respectively. As such, students would have fewer descriptors to handle and gain more reflection time on their writing and future study plan with reference to the descriptors. The four self-assessment sessions focused on the following aspects (Appendix 3–6), respectively:

- Constructing summaries (9 modified descriptors) (Week 5)
- Language use in summaries (12 modified descriptors) (Week 8)
- Constructing argumentative essays (7 modified descriptors) (Week 13)
- Language use in argumentative essays (14 modified descriptors) (Week 16).

Instruction focuses for each writing session were adjusted accordingly: Week 5 focused on how to construct summaries; Week 8 discussed the language use in summaries; Week 13 dealt with how to write argumentative essays; and Week 16 analysed the language use in argumentative essays.

4 Perceived feasibility of modified co-constructed assessment descriptors

Students and instructors' perceptions of the feasibility of the modified descriptors were explored in order to identify the effectiveness of the modifications and seek further suggestions to improve the quality of the co-constructed assessment criteria.

The students reported a relatively high level of feasibility of the modified descriptors. Only 15.4% of students chose 4 (13.3%) or 5 (2.1%) to indicate the modified criteria difficult to use for assessing their writing. 23.8% of students reported the descriptors easy to use by choosing 1 (2.1%) or 2 (21.7%). 60.8% of students selected 3 (neutral), suggesting a moderate of feasibility. A Kruskal–Wallis test analysis suggested significant differences in perceived feasibility across the four class ($H = 14.15$, $df = 3$, $p < .05$). However, follow-up pairwise comparisons suggested a significant difference only existed between Class 3 and 4. Class 3 reported the lowest level of feasibility whilst Class 4 with the highest level of writing proficiency reported the highest level of feasibility among the four classes.

The relatively high level of feasibility of the modified descriptors was confirmed by both instructors. They reported no difficulty in using the modified descriptors for teacher assessment of the same assignments. In the first self-assessment session where the modified descriptors were applied, their students seemed to be more confident and quicker in selecting emoticons for each descriptor than they were in the training session with the pre-modified descriptors.

The students were asked to recommend how to further improve the feasibility of the assessment criteria. Continuing using bilingual versions were reiterated in the survey. Over half of the respondents presented keeping improving accessibility could further increase feasibility as difficulty in understanding the descriptors (e.g. accessibility) was the most likely reason for their difficulties of using the descriptors for self-assessment.

Sixteen students suggested shortening the descriptors or dividing long descriptors into a few short sub-descriptors to accelerate their processing of the descriptors; by contrast, 14 students hoped to make the descriptors longer, adding more details to each descriptor. None of the instructors supported long descriptors. Instructor 2 suggested that adding more details to each descriptor might help students understand it, but over-lengthy descriptors would make students spend too much time on the descriptors and hence insufficient time for self-assessment. Instructor 1 recommended spending more time explaining the descriptors in class would be more effective than extending the descriptors. Students also proposed to use a wider scale as sometimes they felt difficult to choose between the three emoticons. They suggested adding to the reflection scales to differentiate varied feelings. In addition, lack of self-assessment experience resulted in their uncertainty about how to use descriptors to assess their writing. More training and support in self-assessment was expected by students. These suggestions should be used for continuous improvement of the descriptors.

Students reported that their limited language and writing knowledge could refrain them from fully understanding how they could use the descriptors to assess their writing. They stated that their low writing proficiency caused their ‘fluctuating’ feelings about their own writing proficiency on different occasions and made them question themselves whether they had selected the appropriate descriptors last time. Therefore, they suggested improving their own English proficiency was the key to enhancing feasibility. However, the impact of writing proficiency on feasibility was not fully supported by statistical analysis of the whole dataset: Class 3 possessed the second highest writing score but the lowest feasibility among the four classes, although Class 4 possessed the highest writing scores and the highest level of feasibility. Nevertheless, it would be important for the instructors to raise learners’ self-esteem of their language proficiency and encourage them to downplay the role of their language proficiency in understanding the assessment criteria and seek support for comprehending the criteria.

5 Perceived usefulness of modified ELP descriptors for self-assessment

The usefulness of the modified descriptors for self-assessment was investigated to decide the future use of the co-constructed assessment criteria for the writing instruction. Nearly all the 135 survey respondents (i.e. 97.9%) confirmed the usefulness of the modified ELP-based assessment descriptors for self-assessment and writing development. Seventy-three (i.e. 54%) students reported that the descriptors helped them identify their weaknesses of writing. Among them, 22 further explained that the descriptors made them attend to underachieved items in their next assignments (e.g. intelligibility of their writing and the use of link words). Students also reported that the descriptors increased their learning motivation as they understood their knowledge gaps and related descriptors made them aware of how to fill in the gap. Students also asserted that the self-assessment activities developed their skills in evaluating their own learning and understood how to do it for other subjects in future. The perceived usefulness confirmed the facilitative role of co-constructed criteria in conducting formative and sustainable assessment.

The two writing instructors echoed the students’ statements in their reflective logs. Both believed that the descriptors had helped their students identify their writing weakness and motivated them to work on those aspects. They assisted learners to set up

learning targets for their next assignments. They also made learners realize the existence of various aspects of writing apart from grammar and vocabulary and developed their understanding of the standards of good work. They further commented that the self-assessment results driven by the co-constructed criteria allowed them to better understand 'students' internal dialogue about their own writing' and helped them to integrate student voice in writing instruction and writing support. They praised the development of meta-cognitive knowledge of knowing how to assess themselves and how to improve their writing, resulting from the co-construction activity. These benefits would motivate them to carry on co-construction activities in their future classrooms.

Furthermore, the usefulness could be boosted via providing students with follow-up activities. Twelve students highlighted in the last survey question that 'It can make me sometimes find my own problems, but I don't know how to fix them' and 'Even though it helped me to realize my weak areas, it was not that useful to improve my writing proficiency.' They suggested teachers to design follow-up activities to assist them in addressing those weak areas or provide advice on how to achieve learning objectives related to the items rated as *nearly there* and *not there yet*. The two instructors resonated with students about the importance of follow-up activities yet expressed the difficulties in providing individualized support with regard to the heavy teaching workload, the large number of students, different learners' needs (e.g. different weaknesses in writing), and learning motivation.

IX Discussions and implications

The current study explored how to involve instructors and learners in co-constructing the assessment criteria to adapt the ELP for local use through constructively aligning the assessment criteria with curricula, teaching, learning, teachers, and learners within the local instructional context. It has shaped the entrenched teacher dominated assessment practice in the local context which was conducted with the assessment criteria provided by the instructors and encouraging instructors to invite students in designing the assessment criteria and integrate their voices into the criteria for self- and teacher assessment. It has shifted learners' roles from passive recipients of assessment design to active participants in designing assessment criteria, undertaking assessment, and interpreting assessment results for future learning. The study has demonstrated that the participative approach to involving students in constructing the assessment criteria could convey information about whether and how to adapt the ELP for local use from teaching and learning perspectives. This study has provided important implications for how to co-construct the assessment criteria while adapting the CEFR/ELP and other similar frameworks which were designed as a common framework across contexts.

1 Importance and considerations for involving learners in co-constructing assessment criteria

This study revealed that language instructors need to strive to involve learners in co-constructing the assessment criteria to improve their accessibility, feasibility, and usefulness of the ELP descriptors for local teaching and learning. It echoes North (2007)'s argument that the insufficient involvement of learners in the process of developing the ELP descriptors contributed to their continuous accessibility problem.

In the current study, learners and instructors were actively engaged in the whole process: from developing their shared understanding of the benefits and limitations of self-assessment to co-designing its procedure to address students' and instructors' concerns within reality constraints (e.g. limited class time and crowded curricula), from eliciting students' voices of the draft assessment descriptors to using them to modify the criteria, and from using the criteria to reflect on learning to seeking instructors' suggestions on facilitating students to address underachieved assessment aspects.

A few considerations need to be taken when the process of co-constructing the assessment criteria is designed. This study has demonstrated that the instructors need to provide students with the experience of using the pre-modified criteria before eliciting their suggestions on improving the criteria. The experiential suggestions would be more insightful and practical than perceptual ones, evidenced by the detailed information about the specific difficult items and associated reasons in the student survey data on pre-modified descriptors which led to a high level of feasibility and usefulness of the modified descriptors. The study has also revealed that the instructors need to synthesize students' recommendations and their judgements with their pedagogical knowledge (e.g. the nature about language development) and reality constraints (e.g. varied language proficiency levels and learners' need). This led to the use of bilingual versions of the descriptors, considering the accumulative nature of language development, the difference between learners' first and target language, and the limited class time.

2 Ongoing yet varied formats of interaction in the co-construction process

Language instructors should regard the co-construction process as a live community of practice with ongoing interaction. To remedy the constraints posed by the limited face-to-face contact time between instructors and students and the crowded curriculum, it is important to be aware that interaction could take place in the varied forms.

In this study, the two instructors reiterated about the infeasibility of engaging learners in the co-constructed process due to the overcrowded syllabus and heavy workload at the very beginning. Subsequent conversations about alternative interaction channels made the co-construction possible. Apart from traditional classroom-based interaction, opportunities of students interacting with the draft criteria designed by the instructors enabled students to express their difficulties of understanding individual descriptors and provide recommendations to improve their accessibility. Later, instructors' interaction with students' views on the pre-modified descriptors expressed in the survey assisted their understanding of students' difficulties and integration of their feedback into the modified criteria, achieving shared understanding. The continuous and varied forms of interaction had led to a high level of feasibility of modified co-constructed assessment criteria and overwhelmingly positive views about their usefulness for writing, regardless of the heavy teaching and administration load.

3 Power distribution between learners and teachers in the co-construction process

This study suggests an increasingly balanced power distribution between learners and instructors in the co-construction process, aiming to change teachers' dominant roles in

designing assessment and related criteria and promote learner engagement and agency in assessment. To make the co-construction process sustainable within the community of practice, learners and teachers in the local setting must play interactive yet different important roles at every stage. Practicality and flexibility should be considered locally.

In this study, learners' prolonged reliance on instructors to construct the assessment criteria suggested teachers' leading roles in drafting the assessment criteria as an appropriate approach in that context. Learner agency indicated their leading role in modifying the assessment criteria to cater for their learning. Teachers' expertise knowledge in pedagogical design decided their positions to finalize the assessment criteria to achieve constructive alignment between assessment, the syllabus, learners' requests and the characteristics of the local instructional context (e.g. learners' language proficiency, teaching materials and class parameters). Learners' perceptions of the criteria should be elicited continuously to facilitate ongoing refinement of the criteria and related writing instruction. The responsibility distribution across stages reflects the complex set of social relations in the community of practice (Wenger, 2011).

4 Evolving pedagogical knowledge and routine to facilitate co-constructing

Co-constructing the assessment criteria within the community of teaching and learning involves learning on the part of learners as well as teachers, as Wenger (2011) highlights. Co-constructing assessment criteria with learners would require instructors to restructure their pedagogic routines (e.g. providing the criteria for students). Listening to students' voices needs them to critically reflect on their teaching objectives and learner needs and act on them. Constructive alignment demands instructors to commit time to search for ways of aligning assessment (via assessment criteria) with teaching and learning: adapt rather than adopt existing assessment criteria. Co-constructing the criteria obliges instructors to share the floor of designing instruction and assessment with learners and raise learners' roles in the community of teaching and learning.

This could be challenging for instructors in an entrenched teacher-driven assessment culture as in China (Zhao, 2018) as the two instructors in this study constantly mentioned the reality constraints of the tight framework of completing the curriculum, heavy workload, learners' incompetence of co-designing their criteria and learning motivation, and the prolonged teaching and learning culture. In this sense, teacher training in empowering learners in assessment is crucial to embark them on involving learners in the assessment design with innovative methods (e.g. enabling different forms of interaction).

X Conclusions

This study involved learners in co-designing the assessment criteria alongside instructors through various interaction channels in an attempt to adapt the ELP descriptors for local use and achieve constructive alignment with assessment, curricula, teaching, and learning. It approached the co-construction process as a form of sustainable assessment within a community of practice, highlighting and substantiating the importance of continuous bilateral effort between learners and instructors within local contexts. The collaborative process improved the feasibility and usefulness of the ELP descriptors for this group of students. It also enabled the learners to develop their cognitive (e.g. learning

achievement and gap and assessment skills) and metacognitive (e.g. how to plan future study based on the assessment results) knowledge. It developed their skills in setting up assessment criteria with other community members (i.e. instructors) and making informed judgement on their performance against the criteria.

In this way, assessment is not only oriented to develop subject knowledge (e.g. writing quality in this case) but also foster assessment skills for the current and future learning. This would be extremely important for the use of the CEFR or similar common assessment framework in local instructional settings because these frameworks/standards play a continuous role in learners' life beyond their current programmes. In other words, learners need to develop their competence of setting up assessment criteria and conducting future-drive self-assessment against prescribed standards in a certain community independently of an academic (Tan, 2007, p. 120).

Acknowledgements

We would like to thank the writing tutors and the students who participated in the study for their time and commitment.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This document is an output from the ELT Research Award scheme funded by the British Council to promote innovation in English language teaching research. The views expressed are not necessarily those of the British Council.

ORCID iD

Huahui ZHAO  <https://orcid.org/0000-0002-9502-4763>

Note

1. The final writing mark consisted of 10% of the weekly writing task marks, 30% of the mid-term writing exam mark and 60% of the end-term exam mark.

References

- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 159–181.
- Bearman, M., & Ajjawi, R. (2018). From 'Seeing through' to 'Seeing with': Assessment criteria and the myths of transparency. *Frontiers in Education*, 3, 96.
- Biggs, J. (2003). *Teaching for quality learning at university*. 2nd edition. Buckingham: Society for Research into Higher Education and Open University Press.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22, 151–167.
- Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41, 400–413.
- Byrnes, H. (2007). Perspectives. *The Modern Language Journal*, 91, 641–644.

- Carson, L. (2016). Fostering engagement with the Common European Framework of Reference for Languages and the European Language Portfolio: Learning from good practice in university language centres. *Language Learning in Higher Education*, 6, 275–282.
- Dochy, F., Segers, M., Gijbels, D., & Struyven, K. (2007). Assessment engineering: Breaking down barriers between teaching and learning, and assessment. In Boud, D., & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term*. Abingdon: Routledge.
- Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66, 477–485.
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, 36, 137–155.
- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27, 53–64.
- Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the China Standards of English: Challenges at macropolitical and micropolitical levels. *Language Testing in Asia*, 7, 1.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lenz, P., & Schneider, G. (2004). *A bank of descriptors for self-assessment in European Language Portfolios*. Strasbourg: Council of Europe: Language Policy Division.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91, 645–655.
- Little, D. (2009). *The European Language Portfolio: Where pedagogy and assessment meet*. Strasbourg: Council of Europe.
- McKernan, J. (1996). *Curriculum action research: A handbook of methods and resources for the reflective practitioner*. 2nd edition. London: Kogan Page.
- Nelson, M.M., & Schunn, C.D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37, 375–401.
- North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91, 656–659.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25, 23–38.
- Runnels, J. (2014). Japanese English learners' self-assessments on the CEFR-J's A-level can-do statements using four and five-point response scales: Volume 1. *The Asian Journal of Applied Linguistics*, 1, 167–177.
- Tan, K. (2007). Conceptions of self-assessment: What is needed for long-term learning? In Boud, D., & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 114–127). Abingdon: Routledge.
- Wenger, E. (2011). Community of practice: A brief introduction. Paper at STEP Leadership Workshop, University of Oregon, USA. Available at: <http://hdl.handle.net/1794/11736> (accessed July 2020).
- Zhao, H. (2018). Exploring tertiary English as a foreign language writing tutors' perceptions of the appropriateness of peer assessment for writing. *Assessment & Evaluation in Higher Education*, 43, 1133–1145.
- Zheng, Y., Zhang, Y., & Yan, Y. (2016). *Investigating the practice of the Common European Framework of Reference for Languages (CEFR) outside Europe: A case study on the assessment of writing in English in China*. ELT Research Paper. London: British Council.
- Zou, S., & Zhang, W. (2017). Exploring the adaptability of the CEFR in the construction of a writing ability scale for test for English majors. *Language Testing in Asia*, 7, 18.

Appendix I. Pre-modified self-assessment grid: Summary.

☺ = achieved; 😊 = nearly there; ☹ = not there yet



1. I can give a simple summary of the reading text. [B1]
2. I can paraphrase short written passages in a simple fashion, using the original text wording and ordering. [B1]
3. I can pick out and reproduce key words and phrases or short sentences from the reading text. [A2]
4. I can tell the important information from minor one and selectively include minor points to support my summary of main points.
5. I can write summaries on my own although I must have them checked for linguistic accuracy and appropriateness. [B2]
6. I can summarize the plot and sequence of viewpoints in a paper I read. [B2]
7. I can summarize the main themes under headings and subheadings in the reading article. [created by the tutors based on teaching objectives]
8. I can make notes of the main points and use the notes to summarize the reading article. [created by the tutors based on teaching objectives]
9. I can briefly summarize the background of the topic. [created by the tutors based on teaching objectives]

Language use:

10. I can have good control of elementary vocabulary, but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations. [B1]
11. I can write with reasonable grammatical accuracy and can correct mistakes if they are identified by others. [B2]
12. I can write with accurate punctuation. [B2]
13. I can spell accurately, apart from occasional slips of the pen. [C1]
14. I can use some simple sentence structures correctly. [A2]
15. I can use reasonably accurately a repertoire of frequently used 'routines' and patterns for summaries. [adapted from B1 descriptor about argumentative essays]
16. I can use some simple structures correctly but may mix up tenses and forget to mark agreement. [A2]
17. I can use a variety of linking words efficiently to mark clearly the relationships between ideas. [B2]
18. I can link a series of shorter, discrete simple elements into a connected, linear sequence of points. [B1]
19. I can use the most frequently occurring connectors to link simple sentences like 'and', 'but' and 'because'. [A2]
20. I can make it clear what I am trying to express although language errors could occur. [adapted from B1]
21. I can produce continuous writing which is generally intelligible throughout. [B1]

(Continued)

Appendix 1. (Continued)

☺ = achieved; 😊 = nearly there; ☹ = not there yet



-
22. I can qualify opinions and statements precisely in relation to degrees of, for example, certainty/ uncertainty, belief/doubt, likelihood, etc. [C1]
23. I can convey simple information of immediate relevance, getting across which point I feel is the most important. [B1]
24. I can express the main point to make comprehensibly. [adapted from B1 descriptor about argumentative essays]
-

Appendix 2. Your perceptions of the assessment descriptors.

Dear All,

This questionnaire aims to investigate your perceptions of the self-assessment grid. Please read questions, instruction and options carefully. There are no 'right' or 'wrong' answers. As this questionnaire is intended for research purposes only, the information provided is considered anonymous, confidential and will not be disclosed to third parties without your permission.

I truly appreciate your volunteering to cooperate and spend time completing the questionnaire. This questionnaire consists of 8 questions. You will need about 10 minutes to complete the questionnaire. Thank you.

Section One: Biographic information (your student ID:)

1. What is your English exam score in College Entrance Examination?
2. Have you passed the following tests? (*You may choose more than one option if necessary.*)
 - a. CET-4 what is the result?
 - b. CET-6 what is the result?
 - c. TOEFL what is the result?
 - d. IELTS what is the result?

Section Two: Your viewpoints of the self-assessment grid

3. How do you think about the accessibility of the assessment grid?
 - a. Extremely easy to understand
 - b. Easy to understand
 - c. Difficult to understand
 - d. Extremely difficult to understand
4. Which descriptor or descriptors are difficult for you to understand? Please write down the order number of the descriptor or descriptors (e.g. 1a) and explain why (e.g. difficult wording, not applicable to you, difficult to measure, etc.).

I feel Descriptor () is difficult because
 I feel descriptor () is difficult because
 . . .

5. How do you think to improve the descriptor or descriptors that you mentioned above? Please comment on them one by one if you identify more than one descriptor.

I think Descriptor () could be improved by
 I think Descriptor () could be improved by
 . . .

6. Is the assessment grid useful for you to assess and understand your writing proficiency?
- Extremely helpful
 - Very helpful
 - Helpful
 - Not helpful
7. Please explain your answer to Question 6.

8. Do you have any other comments about the assessment grid?

Thank you for completing the questionnaire.

Appendix 3. Modified self-assessment grid: Constructing summaries.

Student number: _____ Class: _____.

☺ = achieved; ☹ = nearly there; ☹ = not there yet

☺ ☹ ☹

25. I can give a simple summary of the reading text.
 我能写简单的文章概述。
26. I can paraphrase short written passages in a simple fashion, using the original text wording and ordering.
 我能用原文的措词和顺序简单地概述短小的文章。
27. I can pick out and reproduce key words and phrases or short sentences from the reading text.
 我能从原文中找到关键词、短语或者短句并用于文章概述。

(Continued)

Appendix 3. (Continued)

☺ = achieved; ☹ = nearly there; ☹ = not there yet

☺ ☹ ☹

28. I can tell the important information from minor one and selectively include minor points to support my summary of main points.
我能区分重要信息和次要信息，并能有选择性地用次要信息来帮助概述重要信息。
29. I can write summaries on my own, although I must have them checked for linguistic accuracy and appropriateness by others.
我能自如地概述，虽然我需要别人帮忙检查语言的精确性和表达的合适度。
30. I can summarize the plot and the sequence of viewpoints in a paper I read.
我能概述文章的情节与其关联的观点。
31. I can summarize the main themes under headings and subheadings in the reading article.
我能概述文章中标题和小标题下的主题。
32. I can make notes of the main points and use the notes to summarize the reading article.
我能记笔记，并用笔记来概述文章。
33. I can briefly summarize the background of the topic.
我能简要地概述话题的背景知识。

Appendix 4. Modified self-assessment grid: Language use in summary.

Student Number: _____ Class: _____.

☺ = achieved; ☹ = nearly there; ☹ = not there yet

☺ ☹ ☹

1. I can have good control of elementary vocabulary, but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
我能较好地使用基本的词汇，但是当我表达比较复杂的思想或者处理不熟悉的话题和情形的时候，会出现较大的错误。
2. I can write with reasonable grammatical accuracy and can correct mistakes if they are identified by others.
我能相对准确地使用语法，并能纠正别人指出的语法错误。
3. I can write with accurate punctuation.
我能正确地使用标点符号。
4. I can spell accurately, apart from occasional slips of the pen.
我能正确地拼写单词，除了个别笔误的情况。
5. I can use some simple sentence structures correctly.
我能正确地使用简单的句子结构。
6. I can use some simple structures correctly but may mix up tenses and forget to mark agreement.
我能正确地使用一些简单的句子结构，但是会混淆时态和忘记主谓一致。
7. I can use a variety of linking words efficiently to mark clearly the relationships between ideas.
我能有效地使用各种各样的连接词，清楚地标注各种关系。

(Continued)

Appendix 4. (Continued)

☺ = achieved; 😊 = nearly there; ☹ = not there yet



-
8. I can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
我能把一系列短小的、零散的成分串成有关联的点。
 9. I can use the most frequently occurring connectors to link simple sentences like 'and', 'but' and 'because'.
我会用诸如“和”、“但是”、“因为”等使用频率很高的简单的关联词连接句子。
 10. I can make it clear what I am trying to express although language errors could occur.
我能清楚地表达自己的想法，尽管会有语言方面错误发生。
 11. I can qualify opinions and statements precisely in relation to degrees of, for example, certainty/ uncertainty, belief/doubt, likelihood, etc.
我能不同程度地限定观点和陈述，比如确定/不确定，相信/怀疑，类似，等等。
 12. I can convey simple information of immediate relevance, getting across which point I feel is the most important.
我能传达简单并直接关联的信息，表达我认为最重要的信息。
-

Appendix 5. Modified self-assessment grid: Constructing argumentative essays.

Student Number: _____ Class: _____.

☺ = achieved; 😊 = nearly there; ☹ = not there yet



-
1. I can summarize information from the reading texts and use them to support my arguments.
我能概述阅读内容并用于支持我的观点。
 2. I can develop different main points to form an argument line.
我能基于不同的要点形成一条论证主线。
 3. I can support the main points with relevant supporting detail and examples.
我能使用支撑细节和具体例子来支持论文的主要论点。
 4. I can develop a clear argument line, expanding and supporting the main points with relevant supporting detail and examples.
我能形成一条明显的议论文线索，并能利用相关的支撑细节和具体例子拓展和支持主要论点。
 5. I can elaborate a detailed argument, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.
我能综合不同分主题，详尽地阐述一个具体的论点，得出一个合理的结论。
 6. I can comment on and discuss contrasting points of view and the main themes of various written resources.
我能评论和讨论不同的观点和来自不同阅读渠道的主题。
 7. I can reconstruct arguments in different sources in such a way that the overall result is a coherent presentation.
我能重构来自不同渠道的论点并能连贯地陈述总体的结果。
-

Appendix 6. Modified self-assessment grid: Language use in argumentative essays.

Student number: _____ Class: _____.

☺ = achieved; ☹ = nearly there; ☹ = not there yet



1. I can have good control of elementary vocabulary, but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.
我能较好地使用基本的词汇，但是当我表达比较复杂的思想或者处理不熟悉的话题和情形的时候，会出现较大的错误。
2. I can write with reasonable grammatical accuracy and can correct mistakes if they are identified by others.
我能相对准确地使用语法，并能纠正别人指出的语法错误。
3. I can write with accurate punctuation.
我能正确地使用标点符号。
4. I can spell accurately, apart from occasional slips of the pen.
我能正确地拼写单词，除了个别笔误的情况。
5. I can use some simple sentence structures correctly.
我能正确地使用简单的句子结构。
6. I can use reasonably accurately a repertoire of frequently used 'routines' and patterns for argumentative essays.
我能比较准确的应用议论文经常使用的语言。
7. I can use some simple structures correctly but may mix up tenses and forget to mark agreement.
我能正确地使用一些简单的句子结构，但是会混淆时态和忘记主谓一致。
8. I can use a variety of linking words efficiently to mark clearly the relationships between ideas.
我能有效地使用各种各样的连接词，清楚地标注各种关系。
9. I can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
我能把一系列短小的、零散的成分串成有关联的点。
10. I can use the most frequently occurring connectors to link simple sentences like 'and', 'but' and 'because'.
我会用诸如“和”、“但是”、“因为”等使用频率很高的简单的关联词连接句子。
11. I can make it clear what I am trying to express although language errors could occur.
我能清楚地表达自己的想法，尽管会有语言方面错误发生。
12. I can qualify opinions and statements precisely in relation to degrees of, for example, certainty/ uncertainty, belief/doubt, likelihood, etc.
我能不同程度地限定观点和陈述，比如确定/不确定，相信/怀疑，类似，等等。
13. I can convey simple information of immediate relevance, getting across which point I feel is the most important.
我能传达简单并直接关联的信息，表达我认为最重要的信息。
14. I can express the main point to make comprehensibly.
我能把主要观点表达清楚。