CEFR Journal

Research and Practice

JALT
THE JAPAN ASSOCIATION FOR LANGUAGE TEACHING
全 国 語 学 教 育 学 会

Japan Association for Language Teaching (JALT)

CEFR & Language Portfolio SIG (CEFR & LP SIG)

CEFR JOURNAL—RESEARCH AND PRACTICE VOLUME 3

Table of Contents

Articles

Reports

News

Mission statement

The CEFR Journal is an online, open-access, peer-to-peer journal for practitioners and researchers. Our editorial advisory board comprises stakeholders on a wide range of levels and from around the world. One aim of our journal is to create an open space for exchanging ideas on classroom practice and implementation related to the CEFR and/or other language frameworks, as well as sharing research findings and results on learning, teaching, and assessment-related topics. We are committed to a strong bottom-up approach and the free exchange of ideas. A journal by the people on the ground for the people on the ground with a strong commitment to extensive research and academic rigor. Learning and teaching languages in the 21st century, accommodating the 21st century learner and teacher. All contributions have undergone multiple double-blind peer reviews.

We encourage you to submit your texts and volunteer yourself for reviewing. Thanks a lot.

Aims, goals, and purposes

Our aim is to take a fresh look at the CEFR and other language frameworks from both a practitioner's **and** a researcher's perspective. We want the journal to be a platform for all to share best practice examples and ideas, as well as research. It should be globally accessible to the wider interested public, which is why we opted for an open online journal format.

The impact of the CEFR and now the CEFR Companion Volume (CEFR/CV) has been growing to previously wholly unforeseeable levels. Especially in Asia, there are several large-scale cases of adoption and adaptation of the CEFR to the needs and requirements on the ground. Such contexts often focus majorly on English language learning and teaching. However, there are other language frameworks, such as the ACTFL and the Canadian benchmarks, and the Chinese Standard of English (CSE). On the one hand there is a growing need for best practice examples in the form of case studies, and on the other hand practitioners are increasingly wanting to exchange their experiences and know-how. Our goal is to close the gap between research and practice in foreign language education related to the CEFR, CEFR/CV, and other language frameworks. Together, we hope to help address the challenges of 21st century foreign language learning and teaching on a global stage. In Europe, many take the CEFR and its implementation for granted, and not everyone reflects on its potential uses and benefits. Others are asking for case studies showing the effectiveness of the CEFR and the reality of its usage in everyday classroom teaching. In particular, large-scale implementation studies simply do not exist. Even in Europe, there is a center and a periphery of readiness for CEFR implementation. It is difficult to bring together the huge number of ongoing projects from the Council of Europe (CoE), the European Centre for Modern Languages (ECML), and the EU aiming to aid the implementation of the CEFR. This results in a perceived absence in the substance of research and direction. Outside Europe, the CEFR has been met with very different reactions and speeds of adaptation and implementation. Over the last few years, especially in Asia, the demand by teachers for reliable (case) studies has been growing.

For more than a decade, the people behind this journal – the Japan Association for Language Teaching (JALT) CEFR & Language Portfolio special interest group (CEFR & LP SIG) – have been working on a number of collaborative research projects, yielding several books and textbooks, as well as numerous newsletters. This is a not-for-profit initiative; there are no institutional ties or restraints in place. The journal aims to cooperate internationally with other individuals and/or peer groups of practitioners/researchers with similar interests. We intend to create an encouraging environment for professional, standard-oriented practice and state-of-the-art foreign language teaching and research, adapted to a variety of contexts.

Editorial advisory board

- Grant Black (Chuo University, Japan)
- Jack Bower (Tezukayama University, Japan)
- David Bowskill (Humboldt University, Berlin, Germany)
- Jamie Dunlea (British Council, UK)
- Neus Figueras (University of Barcelona, Spain)
- Vincent Folny (CIEP, France)
- Dafydd Gibbon (Bielefeld University, Germany)
- Timothy Goodier (co-author CEFR Companion Volume)[†]
- Marita Härmälä (Finnish Education Evaluation Centre, Finland)
- Bettina Hermoso-Gomez (University of Leeds, UK)
- Bärbel Kühn (TU Darmstadt, Germany)
- Noriko Nagai (Ibaraki University, Japan)
- Naoyuki Naganuma (Tokai University, Japan)
- Pham Thi Hong Nhung (University of Foreign Languages, Hue University, Vietnam)
- Brian North (co-author CEFR and the CEFR Companion Volume)
- Fergus O'Dwyer (Marino Institute of Education, Ireland)
- Barry O'Sullivan (British Council, UK)
- Irina Pavlovskaya (St. Petersburg State University, Russia)
- Cristina Rodriguez (EOI Santiago de Compostela, Spain)
- Judith Runnels (University of Bedfordshire, UK)
- Nick Saville (Cambridge Assessment English, UK)
- Yukio Tono (Tokyo University of Foreign Studies, Japan)
- Matthew Vetrini (Clark International High School, Osaka, Japan)
- Carolyn Westbrook (British Council, UK)
- Aaron Woodcock (University of Reading, UK)

Journal editorial team

- Morten Hunke (Anna-Lindh Primary School, Berlin, Germany)
- Maria Gabriela Schmidt (Nihon University, Tokyo, Japan)
- Alexander Imig (Chukyo University, Nagoya, Japan)
- Fergus O'Dwyer (Marino Institute of Education, Ireland)

Editing and proofreading team

- Mary Aruga (Suwa University of Science, Japan)
- Gregory C. Birch (Seisen Jogakuin University, Japan)
- Wendy Gough (Bunkyo Gakuin University, Tokyo, Japan)
- Mathew Porter (Fukuoka Jogakuin Nursing University, Japan)
- Ryan Richardson (Konan University, Japan)
- Colin Skeates (Keio University, Japan)
- Michael Stout (Hakuoh University, Japan)

Layout

- Malcolm Swanson (Seinan Jogakuin University, Japan)

If you would like to contact the authors, please write to the journal at: journal@cefrjapan.net

Editorial

Alexander Imig

The CEFR and the international CEFR-movement represent an achievement of the emerging world society. Built upon a foundation of long-range language policy by the Council of Europe (CoE), the CEFR was a European project first. But since then, the project has turned out to be a success story beyond Europe as well; as the volume of Byram and Parmenter (2012) impressively demonstrated. The perspective, however, of their book is in two aspects too narrow: 1) the case studies of countries offer only a brief outline for these particular countries; 2) the role of networks of teachers and researchers is only touched upon. A systematic analysis of networks could not be carried out within the limited framework of the book. Inquiring into both aspects is in fact the 'raison d'être' of the CEFR Journal. The first and second volume of this journal illustrate amply that practitioners in the field of language learning, teaching, and assessment also successfully act as researchers and have to offer valuable insights into approaches utilizing the CEFR in different organizations.

This third volume of the CEFR Journal continues this policy and starts with an important companion to the CEFR: the portfolio for languages; the European Language Portfolio (ELP). Maria José Luelmo del Castillo and Maria Luisa Pérez Cavana ask in the first text in the Article section: "Is a self-regulatory eELP the way forward?" and they offer a reflection on "two decades of achievements and failures of the ELP". The next three articles are contributions from Japan. In the second article Rebecca Schmidt and Ellen Head analyze "Initial stages of individual teacher CEFR-related classroom curriculum projects at Miyazaki International College", and in the third article Gary Cook and Yukari Rutson-Griffiths (Hiroshima Bunkyo University) introduce "Learner perspectives: familiarization, knowledge, and perceptions of the CEFR". In the fourth article Takeshi Matsuzaki and Kevin Mark are "Investigating the difficulties for university learners of English in Japan of CEFR B1-level phrases", which provides an in-depth study about the usage of the English Vocabulary Profile (EVP) and its application in Japan.

These longer articles are followed by shorter papers in the Report section, which cover a wide geographical range: First, Cuba. The five authors from Cuba and Germany present an "Interpretation of the CEFR Companion Volume for developing rating scales in Cuban higher education", contributing valuable insights into a flexible and cooperative language policy project in relation to CEFR. The next report, by Déirdre Kirwan (Ireland), offers insights into an example of European linguistic diversity: "Utilising pupils' plurilingual skills: a whole-school approach to language learning in a linguistically diverse Irish primary school" and shows how the CEFR is used to support schooling in the national languages while using the multilingual potential of modern Irish Society. Then, in a report from Yukie Saito (Japan), "Initial stages of individual teacher CEFR-related classroom curriculum projects at a liberal arts university in Japan", the author exemplifies how an individual teacher can align different skills to be acquired to the CEFR. The last article from Russia, by Olga Lankina and Yulia Petc, showcases "Classroom-based assessment of group discussion". It features the "challenges and opportunities" of successful group discussions, including the complex question of how effective group discussions can be evaluated.

Finally, in the News section three pieces of news are reported on: (1) the CEFR & CLIL Symposium/ Conference on 23-25 October 2020 (online), (2) a new research project by the SIG – featuring a Call for Collaboration, and (3) the success and outcome of a research project by members of the SIG, recently published with Springer.

Thank you to all contributors, reviewers, proofreaders and the entire editorial team. I wish for insightful reading.

— Nagoya (Japan), September 2020

Reference

Byram, Michael & Lynne Parmenter. 2012. *The Common European Framework of Reference: The globalisation of language policy*. Bristol: Multilingual Matters.

Is a self-regulatory eELP the way forward? A reflection on two decades of achievements and failures of the ELP

Maria José Luelmo del Castillo, Rey Juan Carlos University

Maria Luisa Pérez Cavana, Open University

The European Language Portfolio (ELP) launched in 2001 was created and promoted by the Council of Europe (CoE) as a tool to foster learner autonomy, plurilingualism, and life-long learning. In spite of the progressive educational principles and promising perspectives to develop and support language learning, the ELP has not become established as a widely implemented tool within the European educational context, not even in its electronic version. This paper starts by briefly introducing the original elements and principles of the ELP in order to evaluate some of its main achievements and failures. After examining different models of electronic ELPs, the paper focuses on ePortfolios as pedagogical tools and, in particular, on the suitability of ePortfolios to develop self-regulation. The authors then present some examples of self-regulatory ePortfolios they have created and implemented in different educational contexts. Finally, they present a new self-regulatory ePortfolio prototype. Although still in an exploratory phase, this prototype seems to offer a flexible, adaptable and powerful tool for a variety of learning contexts and learner needs, including the learning of languages and specifically for a state-of-the-art variant of the ELP. This paper concludes by mapping out the self-regulatory ePortfolio as a possible way forward for the ELP.

Keywords: European Language Portfolio (ELP), ePortfolio, self-regulation

1 Introduction

The creation and launch of the European Language Portfolio (ELP) almost twenty years ago, together with the CEFR, can be considered together as milestones in terms of language policy and language pedagogy. The ELP was conceived as a transnational tool to develop learner autonomy, plurilingualism and lifelong learning. It also represented a substantial educational shift from teacher-centred to learner-centred pedagogies. Although the influence of the ELP pedagogy and its implementation across Europe is undeniable, the ELP has not been able to establish itself in formal educational settings.

This paper starts by looking at the background and principles of the ELP and it looks into some of the factors that might explain its lack of success. It then examines the role of electronic portfolios as one possible version of the language portfolio, before considering, more fundamentally, the pedagogical potential of ePortfolios. In particular, we focus on ePortfolios to develop and foster self-regulation. We present different examples of learning ePortfolios; by these we refer to a type of process-based portfolio, whose main function is to enable learners to take control of their learning, to become more aware of their learning process and to foster meta-cognitive skills. In that sense ePortfolios are more than a tool.

First, however, there is a need to take stock of the achievements of the ELP but also to reflect on its failures and to explore new ways of working with the ELP. We suggest flexible approaches with a strong

focus on 'learning to learn' and argue that the ELP could become a self-regulatory ePortfolio without losing its original spirit.

2 The ELP context and developments

The European Language Portfolio (ELP) is a language learning tool promoted by the Council of Europe in order to adapt to the new intercultural and multilingual reality in Europe at the end of the 20th century. In general terms, we could define the ELP as an educational tool that is the property of the learner. It records their skills in foreign languages and encourages autonomy and reflection on the learning process.

The history of the ELP is closely related to the Common European Framework of Reference for Languages (CEFR) (CoE 2001). The origin of both (Trim 2007) can be traced right back to an intergovernmental symposium held in Rüschlikon, near Zürich, in 1971, where it was decided that a system of transparent objectives for language learning by adults should be created. The subsequent proposal was then, unfortunately, rejected at another symposium in Ludwigshafen in 1977. However, a second Rüschlikon symposium took place in 1991, which recommended the development of a Common European Framework of Reference for Languages (CEFR) and a European Language Portfolio to report personal achievement (CoE 1992). Then in 2001, after ten years of meetings, projects and piloting, both the ELP and the CEFR were launched on the occasion of the Council of Europe Conference in Brussels for the European Year of Languages.

From 2001 to 2012, activity around the ELP was constant: experimentation, launching of different projects to create and implement portfolios, teacher training, data collection, European, national and regional seminars, etc. However, it is difficult to find a clear comprehensive picture of ELP use at the European level after 2012, since projects were no longer being officially tracked.

2.1 The ELP pedagogical rationale

The ELP was created as a practical tool to put the guidelines specified in the CEFR into practice. Thus, both documents share common objectives: to protect and develop the cultural heritage and diversity of Europe as a source of mutual enrichment, to facilitate the mobility and exchange of ideas of European citizens, to develop an approach to language teaching based on common principles and to encourage plurilingualism. Apart from these four common goals, the ELP, being a pedagogical tool, further elaborates on two aims: to promote both autonomous and lifelong learning.

Promoting autonomous learning is closely related to lifelong learning, since the autonomous learner seeks and finds opportunities for learning beyond the classroom. A learner is considered to be autonomous when he/she is able to take responsibility for his/her learning and exercise this responsibility in a continuous effort to understand what, why, and how to learn (Holec 1981; Boud 1981; Little 1991). The ELP helps students to take responsibility for their learning. It not only collects all the learning experiences that the language learner has had both inside and outside the classroom in any circumstance of his/her personal situation, it also fosters reflection and understanding of their own learning. Thus, trying to meet both objectives.

These two objectives of the ELP, to promote autonomous learning and lifelong language learning, show its character as a tool, as an instrument that puts into practice the aims of the CEFR and also presents its own. The ELP is aimed at developing learning awareness, it is aimed directly at the learner, while the CEFR is a document that proposes general guidelines. Both are instruments of learning but they operate at different levels: the ELP with the learner and the CEFR with educational institutions.

Regarding its characteristics, they are specified in the document *ELP Principles and guidelines* (CoE 2000) and could be classified and summarised as follows:

Table 1. *ELP main characteristics*

General characteristics	Plurilingualism and Pluriculturalism	Property of the learner
• Based on the CEFR • Incorporates core elements that make it recognisable and easy to understand • A tool to promote learner autonomy • It has both a pedagogic as well as a reporting function	• It is a plurilingualism and pluriculturalism promoting tool • Values linguistic and intercultural competence in and outside formal education	• Owned by the student • Fosters self-assessment • One of a series of documents that the student will own throughout his life

The ELP consists of three parts: Passport, Biography and Dossier. However, the sections might vary depending on the country or type of learner to whom it is addressed.

The *Language Passport* contains a self-assessment grid through which the holder can reflect on their language competence according to skills (listening, reading, spoken interaction, spoken production and writing). The learner uses language descriptors from the CEFR to assess their level of language proficiency in each of these skills. Certificates and accreditations that the holder has accumulated throughout his/her experience as a language learner are also recorded in the Passport.

The second section of the ELP, the *Language Biography*, is the part of the ELP in which its pedagogical function is realised. It is there the learners describe their learning process. It contains forms where the student self-evaluates, describes the learning activities that help them learn, reflects on the use of learning strategies, sets new objectives, and where he details his linguistic and cultural experiences outside formal education (Lenz and Schneider 2000).

The third section of the ELP, the *Dossier*, is the section that most reminds us of that portfolio of the artist who inspired the idea of the ELP (Little and Perclová 2002). It contains samples of the student's work. It is the holder who must decide what projects, recordings, etc. they will include in their ELP, since these works are the ones that they will present as a sample of what they can do in foreign languages and, therefore, they must be significant tasks for the holder.

These three sections of the ELP fulfil two different distinct functions: the reporting and the pedagogic. The Passport shows the owner's linguistic competence in different languages and has thus mainly a reporting function. The Biography has mainly a pedagogic function, it supports the learner's learning process and the Dossier combines both functions.

The ELP can also be used to illustrate the student's linguistic competence in foreign languages - the reporting function. The ELP documents the linguistic capacities of its holder in a comprehensive and transparent manner. It helps the holder to record their level of competence achieved in one or more languages and in each of the skills; accounts for the formal and non-formal learning experiences that the holder has lived; it shows the self-evaluation that the student has made of his/her capabilities.

The pedagogic function of the ELP involves making the learning process more visible and helping individuals to develop their ability for organisation, reflection and self-assessment. That is, it improves their meta-cognitive skills, and therefore it will foster their learner autonomy. This way, learners will be able, little by little, to assume more and more responsibility for their learning (Little and Perclová 2002). The objective of this function coincides with the emphasis on learning to learn and the development of critical thinking skills increasingly present in regional and national curricula.

To sum up, the ELP was a tool launched by the Council of Europe in 2001 in order to implement CEFR principles and promote student's autonomy and lifelong learning. The ELP was structured around

three recognisable sections (Passport, Biography and Dossier) and had both a pedagogic and reporting function. Its use was implemented throughout Europe (albeit unevenly) from 1998 to 2012, after which the different projects stopped being officially monitored. 32 countries participated in the project during those years and they provided evidence of their successes, failures and challenges. What follows is a summary and a reflection on its adoption.

3 A major change within languages education? Critical evaluation

Viljo Kohonen, one of the most relevant researchers in the field of the European Language Portfolio, writes about the ELP-oriented pedagogy as a major change within language education (Kohonen 2012). In their European Language Portfolio Impact study, Stochieva et al. (2009) concluded that the ELP seemed to have had positive effects on the classroom, textbooks, tests and exams, other educational projects, teacher education and training and on language policy in general. Their personal conclusion about the study emphasises the link between the ELP and CEFR:

> The ELP with its emphasis on learner autonomy, self-assessment and lifelong learning has reinforced some of the basic implications of the CEFR approach – those elements which constitute the underlying concerns behind its conception. By engaging in the ELP development process practitioners, teachers, educators and a wide range of FLL stakeholders have, we believe, achieved a better understanding of these underlying principles of the CEFR. (Stochieva et al. 2009: 20).

However, despite reports of its positive impact at so many levels and the fact that learning portfolios were becoming increasingly popular in different disciplines, the ELP, both in the original paper format and the more recent versions in electronic design (e-ELP), was not adopted as widely as had been expected. We do agree that the ELP had the potential for a major change in languages education, however, this change did not materialise.

David Little, a leading figure in the ELP project, admits that in spite of the large number of portfolios validated and registered, 118 portfolios, the ELP has not been successfully implemented on a large scale in any educational system: "the ELP has never been used on a large scale in most national education systems and seems to be largely forgotten in some of those that were among the first to develop ELPs and submit them for validation" (2016: 162). However Little adds, the ELP has had a major impact on transforming curriculum, textbooks, and teaching practice.

It is paradoxical that, once the pedagogical value of the ELP was demonstrated after its pilot phase and the years in which the implementation projects were active and after the use of learning portfolios appeared in various educational areas, ELPs have not been disseminated and implemented in a systematic way. Little (2016: 166-167) points to four main reasons for this failure:

- The ELP did not live up to the expectation of being the magic bullet that would provide the universal remedy for language learning and teaching ailments: in many countries the ELP needed much more support than the authorities could provide.
- The ELP's pedagogical approach was not well aligned to most educational systems; it was strange (Little calls it *alien*).
- The ELP encountered integration problems in three areas: Most ELP models were not developed as part of a broader curricular reform, making self-assessment descriptors difficult to relate to curriculum objectives. In addition, in most educational centres a textbook was used, which meant that the ELP was an extra burden. Furthermore, the culture of self-evaluation and reflective learning that underlies the ELP was unthinkable in many educational systems.
- The ELP itself presents some problems, such as the dichotomy between the use of the target language and plurilingualism.

However, Little wonders whether the time is now ripe for a revival of the ELP (2016). He considers that "the educational ideals on which the CEFR and the ELP are founded have lost nothing of their relevance and urgency" (2016: 168). Therefore, he proposes different ideas to guarantee a successful return of the ELP and suggests a bottom-up implementation model starting in the primary classroom and spreading to higher education. He furthermore recalls the importance of establishing clear links between official syllabi and the ELP. This involves the reformulation of the objectives of the official curricula following the CEFR scales. Lastly, Little proposes that it is important to redesign the tool according to each particular context, now that there are no longer any validation or registration processes and, therefore, the creation of new ELPs can be more flexible, affirming that "sticking to the three core sections is now a *moot* question" (2020: 15, emphasis in the original). He suggests the following guidelines for the development of new ELPs:

> Some form of language biography is clearly essential to provide a reflective accompaniment to learning and support the recurrent cycle of goal setting and self-assessment; and a dossier is an obvious way of storing both work in progress and work that can be used to support self-assessment claims. But a version of the language passport might be used for a reporting function only. In a school system, for example, students might need different curriculum frameworks and portfolios for first, second and subsequent foreign languages but could use the same language passport to summarize their learning achievement at the end of schooling. (Little 2020: 15).

Finally, we note that Stoicheva et al. (2009: 27, original emphasis) have reported a "*loss of momentum*" in use of the ELP. We believe that this momentum can be re-established through investment in a structured reconceptualisation of the ELP as a Personal Learning Environment (PLE) (Haines and van Engen 2012: 143). These authors consider that the original structure of the ELP produced "artificial separation of the language learning experience" and propose a "more organic representation of the language learning process from the learner perspective" (2012: 139).

Regarding the implementation of the ELP, Forster Vosicki (2012) argues that different key factors have to be taken into account. She draws from her experience implementing the ELP at the University of Lausanne Language Centre for more than a decade and offers a solid analyse of the challenges that working with an ELP at higher education present. These key factors offer a clear vision of the issues that need to be addressed to ensure successful ELP adoption; at the management level, the awareness that implementation is a long process that needs to be facilitated and monitored; at the teacher level it needs the integration of teacher training; and in term of learning infrastructures, the design of tasks and learning support systems in line with the objectives of the ELP (Forster Vosicki 2012).

The above discussion and evaluation clearly demonstrate the complexity and multiplicity of factors that are involved in the successful implementation of the ELP. In this paper we would like now to focus on two aspects of the ELP: the pedagogical concept and the use of an ePortfolio. In the following sections these two fundamental aspects are explained and documented.

4 The need for an electronic e-ELP

Originally, the European Language Portfolio was a paper document, but soon it became apparent that a more flexible and accessible format would be the future of the tool. Furthermore, the use of electronic portfolios in various academic fields was spreading more and more (Haines and van Engen 2012).

The first accredited electronic ELP was developed in 2001 by EAQUALS/ALTE. From then on, the development of electronic portfolios was promoted by the Council of Europe. In recent years, several versions have appeared, despite the fact that neither its use nor its creation have had the expected impact.

The creation of an electronic portfolio brought with it many advantages. Haines and van Engen pointed out, for example, that the use of various languages in different settings was becoming commonplace

and that it could be made visible through the different components of the portfolio in digital format: "The use by learners of several languages in a variety of settings and with a variety of goals is becoming the norm, and the production of digital language biographies, dossiers and language passports can make this explicit" (2012: 131).

It is important to note, however, that adapting the ELP to electronic format was not simply a matter of updating its format. This is how García explains it: "Electronic portfolios cannot remain a digital version of their pencil and paper peers, as many authors suggest. They must go beyond a mere compilation function and assume functions of learning management" (2005: 115). Despite this recommendation: "The tendency has been mostly to replicate the paper design and format", without taking into account the technological advantages that can improve the learning process through eportfolios (Álvarez 2012: 131).

The first European Language Portfolio in electronic format was developed by EAQUALS and ALTE and accredited in 2000. After this one, many others followed, such as the Lolipop project (2003-2007), the European Language Portfolio for Professional Purposes (Prof-ELP) for vocational training students, developed by the Employment Service of the Community of Navarra, the e-ELP of the University of Montesquieu (2010) or the electronic ELP of the Regional Centre for Education and the Official School of Languages in Pilsen, in the Czech Republic (2014).

More recently, other e-ELP models have been developed and used by various international institutions. As an example, we could mention *Peppels*, which is a Dutch e-ELP model created in 2007. Since then, its use has spread in schools in the Netherlands in a commercialised version that is currently active. *LinguaFolio* is not a European project, however, both its design and the underlying principles are the same as those of the European Language Portfolio. It was created for students between 12 and 18 years old and, mainly, to be used in secondary schools. Currently, it is also used in some universities, such as the University of Oregon, and in intensive language learning programs in other institutions where it has proved to promote self-regulation and motivation (Ziegler and Moeller 2012).

The last example we would like to mention is EPOS, which was launched in 2012 to take advantage of technological innovations that could facilitate learning through portfolios, creating a tool that offered more modern functionalities, while adapting to learner-centered methodologies that began to be used more and more. Already in the first version launched in 2012, EPOS allowed for the creation of groups to use it cooperatively, carry out projects together, create collections of works and allow other students to see and comment on them (Fehse et al. 2011: 5).

The principles on which EPOS is developed make it clear that this portfolio is much more than a change of format from a paper to electronic medium. EPOS already introduced more functions that were implicit in the ELP, such as Learning objectives, *Lernziele,* Projects, Learning journal etc. Friedrich and Kühn highlight how EPOS goes beyond the ELP, in terms of flexible self-assessment, the possibility to work with different descriptors, e.g., CercleS, CARAP, Intercultural communication etc, and the possibility of collaborative learning and group work (Friedrich 2019; Kühn 2016). It was implemented at the Universities of Bremen Language Centre for language exchange tandems and in many other German and European universities. This e-ELP model is especially relevant in the context of our work, since the underlying principles and functionalities of EPOS have been used as a basis for the ePortfolio prototypes that will be described below.

5 Developing a learning ePortfolio for languages
5.1 The pedagogical role of ePortfolios

As the previous section has shown the move to an electronic Portfolio was both a 'natural' and necessary step within the development of the ELP. Over the last decade ePortfolios have become an increasingly common component of Higher Education (HE) programmes, serving as constructivist learning spaces where students can reflect on their learning journeys, where they can be assessed, collect their work and

demonstrate their achievements to potential employers (Pegrum and Oakley 2017). The recent saliency of ePorfolios has been stressed (Chaudhuri and Cabau 2017) as they are demonstrating in different contexts and across disciplines how they might fit with institutional objectives as well as allowing for a greater personalisation of learning. As Pegrum and Oakley state:

> It is suggested that ePortfolios may have a role to play in supporting a shift away from today's administratively oriented, pedagogically limited learning management systems (LMSs), and towards personal learning environments (PLEs) where students can engage in more individualised, autonomous learning practices. (Pegrum and Oakley 2017: 21)

In line with this position held by Pegrum and Oakley that ePortfolios foster the shift to a more personal and autonomous learning, we have brought this claim further with regard to two aspects:

- by stressing the fundamental role that an ePortfolio can play as a learning tool.
- applying and integrating the pedagogical structure of self-regulation to different ePortfolios.

5.2 The way to a self-regulatory ePortfolio

Drawing from our own experience designing and working with different electronic ELPs, it appeared clear that the pedagogical potential of ePortfolios had not been fully explored and acknowledged. Elsewhere we argued (Pérez Cavana 2012) that there was a need for a *soft* portfolio'. With the word *soft* we were not only referring to an electronic portfolio, but also to the **pedagogical component** of the ELP as opposed to the *hard pages* or **reporting function** of the ELP. In that paper we showed how the ELP can foster strategic self-regulated learning and metacognitive knowledge.

Following this line of inquiry over the last years, we have focused on the development of an ePortfolio that specifically fosters and develops self-regulation. As explained in the sections above, the ELP is based on the principle of learner autonomy. In this paper we argue that an ePortfolio can drive this principle further to promote self-regulation. Thus, we endeavoured to make the most out of the affordances of an ePortfolio to develop a learning, self-regulatory ePortfolio, that can be used in a variety of contexts, and definitely as a language portfolio.

5.3 Integrating Self-regulated Learning (SRL) functions in the ePortfolio structure

Self-regulation is widely recognised as an important factor in active control of the learning process (Goulão and Cerezo Menedez 2015) and consequently, in students' academic performance. According to Zimmerman (2000) self-regulated learning in education is based on the premise that students use metacognitive, motivational, and behavioural processes in their learning.

As described above, autonomous learning requires self-management; that means being proactive and developing self-knowledge and control of the learning process. Bjork et al. (2013) have demonstrated in their research that for a learner to become effective in the learning process, they should not only be able to assess accurately the states of their own learning, but also be able to manage their own learning and activities in response to such monitoring.

Gavaldón (2019), drawing from her research on studies on ePortfolios for student teachers, stressed that for the ePortfolio method to be effective, teachers need to direct students progressively toward self-regulated learning. In order to functionalise the principles of self-regulation and adapt them to fit within an ePortfolio structure, we used as the starting point the five stages of learning recommended by the Open University for students to work on their Personal Development Planning as seen in Figure 1 (Open University 2020).
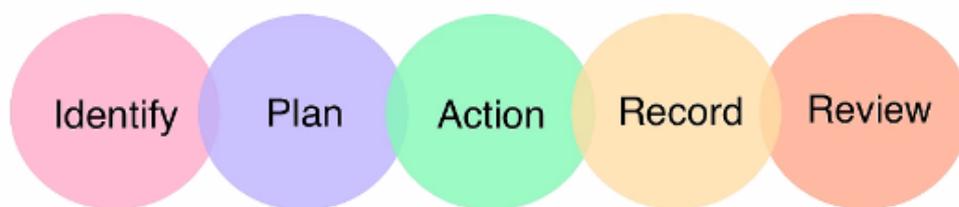
Figure 1. Five-step learning cycle.

As you can see this series of actions is understood as a learning cycle which describes the types of actions characteristic of self-regulation. The cycle comprises of the following stages: identifying (learning goals, weaknesses etc.); planning (how to work with these weaknesses, how to achieve the learning goals); action (performing the planned actions); recording (evidence of the actions performed, successes etc.); and reviewing the whole process (has it worked as planned, have I achieved what I wanted? If not, what can I do differently?).

Although there are different models of SRL, according to the cyclical model proposed by Zimmerman (2000) there are three phases in SRL: forethought, performance and self-reflection, as can be seen in Figure 2. These phases closely match the pedagogical cycle shown and described above:

- Forethought – (Identifying, Planning).
- Performance – (Action, Recording).
- Self-reflection – (Reviewing, Evaluating).
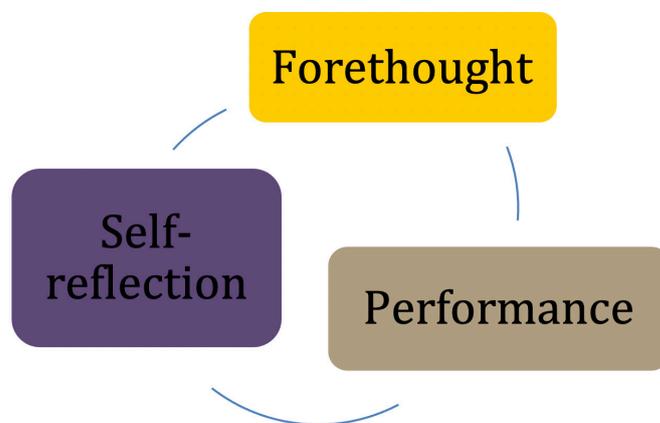
Figure 2. SRL pedagogical cycle.

Drawing from the pedagogical principle of self-regulation and the functions described above we designed and refined our learning ePortfolio prototype, which we could call a self-regulatory ePortfolio.

6 Applying the self-regulatory ePortfolio to different contexts
6.1 ePortfolios for Personal Development Planning (PDP)

From 2016-2018, we carried out a pilot to implement a new approach to Personal Development Planning (PDP) using the Three-layered model (Pérez Cavana and Lowe 2018) in an ePortfolio. Through successive pilots we designed and refined our learning ePortfolio prototype. For the first pilot we used EPOS, the ePortfolio developed by the Language Centre for the Universities of Bremen. As explained above, it was

originally designed as a language ePortfolio following the structure of the European Language Portfolio. We adapted EPOS specifically for PDP at the Open University (OU) by integrating the pedagogical functions (identify, plan, record, review) in the main tabs (see Figure 3) in order to facilitate the visualisation of the pedagogical process.

Figure 3. The ePortfolio EPOS for PDP.

The rationale behind inserting these pedagogical functions in the ePortfolio was twofold: cognitive, to help to understand what PDP is about, and practical, to facilitate the factual work with learning by doing.

The findings of those studies have been published elsewhere (Pérez Cavana and Lowe 2018) but the main finding was that the visualisation of the pedagogical cycle as part of the structure of the ePortfolio helped students to become aware of their learning, to manage, to plan it and to take control of their learning.

At the time of the second pilot study, the OU was starting to provide students with Microsoft Office 365, a cloud-based suite of tools including OneNote. This aligned to Kim et al.'s (2010) proposal of a cloud-based approach for ePortfolios. We therefore piloted the use of OneNote as a means of supporting students in their PDP.

As advocated by Howes et al. (2011), we provided a structure within the ePortfolio through the creation of a template in OneNote which we then made available to the students. Once set up with Office 365, students installed the OneNote template on their own devices. They had a choice of desktop version and cloud-based, potentially syncing the two and using both depending on their location and device at hand.

The template in OneNote (Figure 4) was a simplified version of the EPOS ePortfolio used in our first pilot (Figure 3) but maintained the idea of the tabs to provide guidance through the stages of PDP (identify; plan; record; review). Under each of the tabs was a space for students to use, in some cases with minimal scaffolding in the form of framework or prompt questions, and in other cases, space for students to use as they wish.
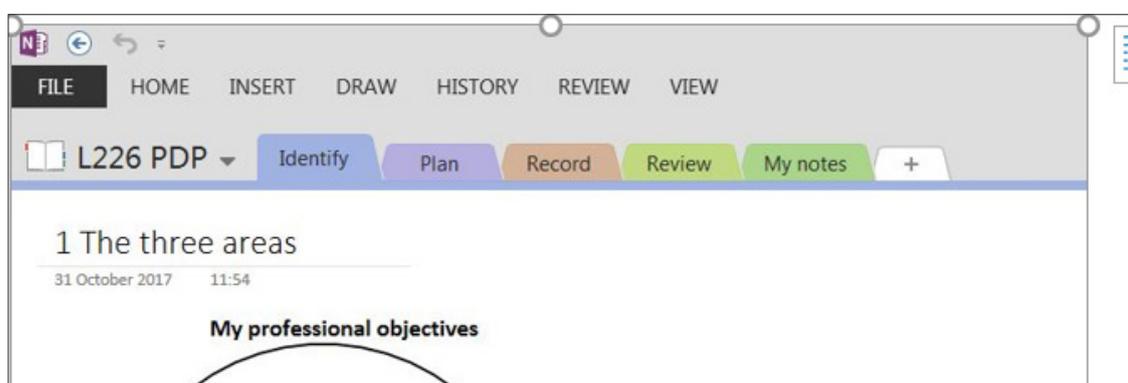
Figure 4. Tabs created within OneNote.

One of the advantages of using OneNote is that it is very easily customised by the users. Under the tabs and in their personal space, students were able to enter their thoughts as text, in paragraphs, in tables, in lists. They can upload photos of work done or inspirations, upload audio/video recordings of themselves reflecting, and upload their assignments containing tutor feedback. They can make use of checkboxes to help prioritise and to keep track of progress.

6.2 ePortfolios for languages

Following this prototype self-regulatory ePortfolio and the promising findings from the first pilot study on PDP, a similar OneNote ePortfolio was designed based on the European Language Portfolio using the same pedagogical functions (identify; plan; record; review). The pedagogical self-regulatory cycle is also the basis of the structure, although the tabs have been kept relevant for the specifics of language learning and follow the traditional structure of the European Language Portfolio.

The OneNote template was introduced in a number of hands-on workshops for language teachers in the context of a European Centre for Modern Languages' project. The language teachers were highly positive in their response to the ePortfolio. They also valued the ability to customise and adapt it to their specific needs. These teachers are now able to use their own adapted versions in their teaching, but have yet to report on their experiences.

Figure 5 shows one example of a language ePortfolio. In it the traditional parts of the ELP (Passport, Biography and Dossier) have been freely adapted to the learners' needs. In this particular example, the self-assessment section – normally included in the Passport – has been presented in three language skills (speaking, listening and writing) and an additional section on *learning objectives* has been added to allow students to plan and manage their learning and to strengthen self-regulation. Another main section is the *Dossier* as in the original ELP concept, but with the facilities an ePortfolio provides, such as the possibility to store and collect all types of files and documents, including video and audio files, pictures etc. Finally, the learning journal part fulfils the function of the biography.
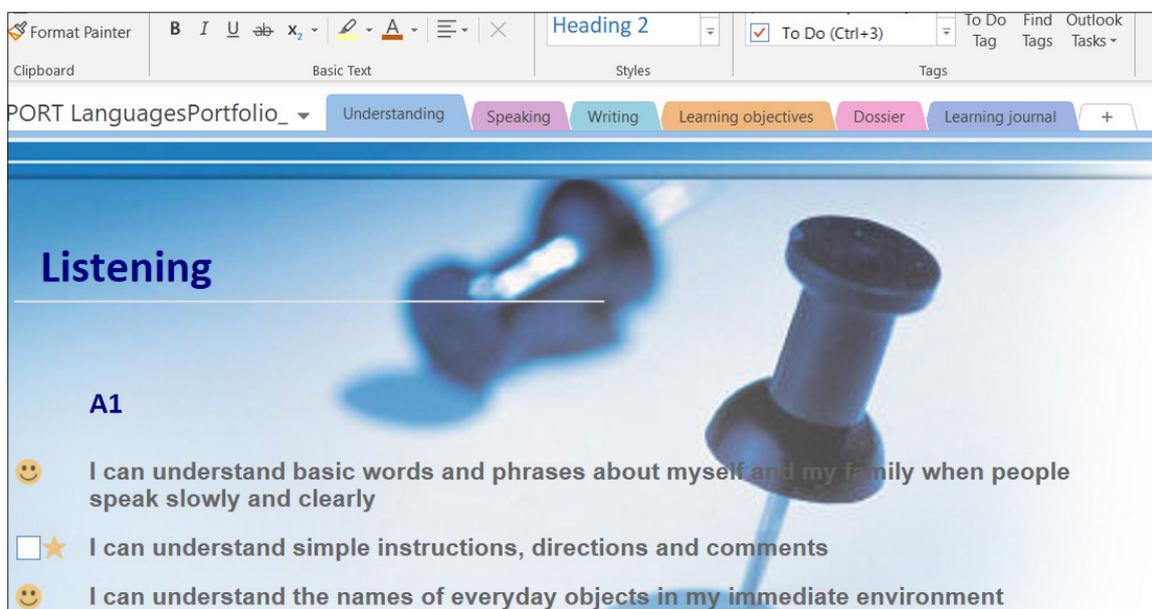
Figure 5. Structure of OneNote Languages ePortfolio.

6.3 ePortfolios for English student teachers

Since the OneNote ePortfolio template could be tailored to suit different target groups, a different prototype was designed on the same basis. Using the same pedagogical functions (identify; plan; action; record; review), another OneNote ePortfolio was created for future pre-primary teachers. In this case, the students were taking the subject English Language and its Pedagogy within their Pre-primary Education Degree and the portfolio was used not only to promote their language learning but also the core content of the subject: English Teaching Methodology. The portfolio was meant as class activity and the students were asked to complete the different portfolio sections after each unit with the work being supervised by the teacher.
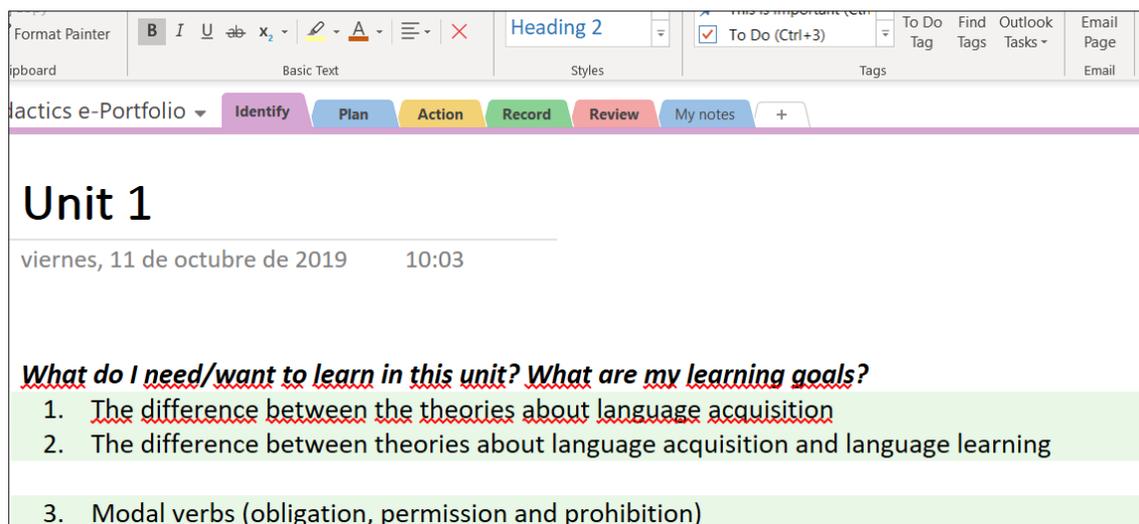
Figure 6. Structure of OneNote Portfolio for student teachers.

Figure 6 shows how OneNote was adapted for these students. Again, the original sections of the ELP (Passport, Biography and Dossier) were adapted to this new context. The Identify tab helps the student set goals and identify what they can already do, resembling the Passport. The Plan and Action tabs are for designing a plan of action and developing specific activities to cover the contents of the subject in a flexible way. Together with the Review tab, where a learning journal can be found, they keep the spirit of the Language Biography. Lastly, evidence can be stored in the Record section (Dossier).

The ePortfolio was used as a voluntary class activity and students were surveyed about the experience when the course was over. Student teachers showed a very positive attitude towards it and most of them felt it helped them with self-assessment and getting more control over their learning process. In general terms, they felt it was a useful reflection tool. As it has been shown, we have opted for a flexible ELP model to suit different contexts and target students.

7 Discussion

Continuous development has resulted in the latest examples of language eportfolios we have described. Is it possible to still recognise the original ELP within these new examples of the Self-regulatory portfolio that we are suggesting?

The answer is 'yes' and 'no'.

Yes: the new learning ePortfolio for languages we are proposing is clearly based on the original spirit of the ELP: to develop learner autonomy and supporting plurilingual lifelong learning. It also includes the reporting and the pedagogical functions. That means it works as a product (Dossier/Record) and as

a process, although the emphasis is more on this latter, on the scaffolding[1] function, aiming to make learning visible and to facilitate the learner taking control of their learning.

Yes: it is property of the learner, and self-assessment is a main element in the ePortfolio. In fact, this aspect of self-reflection is much more developed in the SRL ePortfolio. The learning function in the SRL ePortfolio has taken over the reporting function (although both are present since the ePortfolios we are suggesting can be used to assess learners).

No: the original three-part structure (Passport, Biography and Dossier) might not be recognisable at first sight, it has become much more flexible and fluid. The main elements of these parts are either integrated or expanded upon, however there are some elements such as the record of formal language qualifications, that do not have a main role in the SRL ePortfolio, although they can be easily included.

The SRL ePortfolio we are proposing is much more flexible than the original one and much more learner- and context-centred. In that sense we are in line with Blanch et al. (2011) who believe that portfolios "have to be flexible and must promote self-reflection and autonomy in students' learning: therefore, proposing a standardised learning portfolio model that homogenises the results and allows for a statistical approach would be incoherent with this conception". Thus, the flexible approach we are proposing enables the purposeful development of the three parts. In the example discussed above, the passport has been designed around the different language skills (writing, listening, speaking) and a new section to work with learning objectives has been introduced. Also, in that example, the Biography was adapted to follow the self-regulation cycle (Plan-Action-Record-Review). Besides, when the original ELP is integrated in the SRL ePortfolio as we are proposing, it can support other disciplines or skills such as Personal Development Planning, language teaching or teaching practice as shown in the examples.

Another major difference with the original ELP is that it is not meant to be implemented at a regional-national-European scale like the old projects, but rather to be adapted to one's context and used following the initiatives or demands of individual teachers or educational institutions, as suggested by Little (2016).

We are aware of the limitations of this new ePortfolio prototype. It is a work in progress and we only have incidental evidence of its effectiveness. Therefore, it is essential to carry out studies in different contexts and to collect and analyse significant amounts of data. Regarding the digital platform of the SRL ePortfolio, the pilots we carried out with OneNote showed several technical drawbacks, such as the co-existence of different versions, the difficulties of importing a template on some computers and the limited inter-connectivity of the parts. Therefore, we will be using a different platform, Mahara, for the design and development of our next SRL ePortfolio. We are also in the process of developing measurement tools to collect a robust set of data.

8 References

Álvarez, Inma. 2012. From paper to the web. In Bärbel Kühn & María Luisa Pérez Cavana (eds.), *Perspectives from the European Language Portfolio. Learner autonomy and self-assessment*, 125-142. London and New York: Routledge.

Bjork, Robert, John Dunlosky & Nate Kornell. 2013. Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology* 64(1). 417-444. doi:10.1146/annurev-psych-113011-143823 (accessed 14 May 2020).

Blanch, Sílvia, Alejandra Bosco Paniagua, Xavier Gimeno Soria, Neus González Monfort, Marta Fuentes Agustí, Mercé Jariot García, Sònia Oliver del Olmo, Montserrat Rifà Valls, Noemí Santiveri Papiol, Dídac Segura Aliaga, Laura Trafí i Prats. 2011. *Portfolios in higher education: an opportunity to rethink teaching*. Bellaterra: Universitat Autònoma de Barcelona.

---

1. "Scaffolding" is a concept introduced by J.Bruner to refer to the support given to students when they learn something new. Bruner, J. S. (1978). The role of dialogue in language acquisition. In A. Sinclair, R., J. Jarvelle, and W. J.M. Levelt (eds.) The Child's Concept of Language. New York: Springer-Verlag.

Boud, David. 1981. *Developing student autonomy in learning*. London: Kogan Page.

Chaudhuri, Tushar & Béatrice Cabau. 2017. *E-Portfolios in Higher Education*. Singapore: Springer.

Council of Europe. 2000 (rev January 2011). European Language Portfolio (ELP): Principles and Guidelines. Strasbourg: Council of Europe. https://rm.coe.int/CoERMPublicCommonSearchServices/ DisplayDCTMContent?documentId=09000016804586ba (Accessed 7 September 2020).

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. http://www.coe.int/en/web/common- european-framework-reference-languages (accessed 03 June 2020).

Council of Europe. 1992. *Transparency and coherence in language learning in Europe: Objectives, assessment and certification*. Proceedings from the intergovernmental symposium at Rüschlikon, 199. Strasbourg: Council of Europe. https://rm.coe.int/0900001680707cde (accessed 03 June 2020).

Forster Vosicki, Brigitte. 2012. Overcoming the challenges of implementing the European Language Portfolio (ELP) in higher education: the experience of the University of Lausanne Language Centre. In Bärbel Kühn & María Luisa Pérez Cavana (eds.) *Perspectives from the European Language Portfolio. Learner Autonomy and self-assessment*, 100-121. London and New York: Routledge.

Fehse, Rüdiger, Jürgen Friedrich & Bärbel Kühn. 2011. EPOS- An ePortfolio for Language Learning. Paper presented at the International Conference ICT for Global Learning International Conference, Florence, 20-21 October.

Friedrich, Jürgen. 2019. EPOS- Das elektronische Sprachenportfolio. Entwicklung, Didaktik, Perspektiven, In Bärbel Kühn & Sandra Ballweg (eds.), *Portfolioarbeit im Kontext von Sprachenunterricht*, 171-188. Göttingen: Universitätsverlag Göttingen.

García Doval, Fátima. 2005.El papel de los portafolios electrónicos en la enseñanza-aprendizaje de las lenguas. *Glosas Didácticas.* 14. 112-119*.* http://files.geohistoria5.webnode.es/200000009- dcd1eddcbf/10.pdf (Accessed 07 September 2020).

Gavaldón, Guillermina. 2019. ePortfolio as a Means of Scaffolding Reflection. From Written Reflections to Graphic Narratives. Portfolioarbeit im Kontext von Sprachenunterricht. Neue internationale Entwicklungen, In Bärbel Kühn & Sandra Ballweg (eds.), *Portfolioarbeit im Kontext von Sprachenunterricht*, 171-188. Göttingen: Universitätsverlag Göttingen.

Goulão, María de Fatima & Rebeca Cerezo Menedez. 2015. Learning autonomy and self-regulation. Paper presented at International Conference on New Horizons in Education, INTE 2014, Paris, France, 25-27 June. Procedia Social and Behavioural Sciences 174. 1900-1971.

https://www.sciencedirect.com/science/article/pii/S1877042815009052 (accessed 14 May,2020).

Haines, Kevin & Jeroen van Engen. 2013. Re-conceptualizing the ELP as a Web 2.0 Personal Language Learning Environment. *Language Learning in Higher Education* 2(1). 129-144. doi:10.1515/ cercles-2012-0008.

Holec, Henri. 1981. *Autonomy and foreign language learning*. Oxford: Pergamon for the Council of Europe.

Howes, Anne-Marie, Niall Dew, Pat Harkin & Tamsin Treasure-Jones. 2011. *ePortfolio Project Report*. University of Huddersfield. https://core.ac.uk/ display/30730045?source=2&algorithmId=10&similarToDoc=%27CO (accessed 14 May 2020).

Kim, Paul, Chen Kee Ng & Gloria Lim. 2010. When cloud computing meets with Semantic Web: A new design for e-portfolio systems in the social media era. *British Journal of Educational Technology* 41(6). 1018-1028. doi:10.1111/j.1467-8535.2010.01055.x, https://doi.org/10.1111/j.1467-8535.2010.01055 (accessed 1 May 2020).

Klenoski, Val. 2014. *Desarrollo de portafolios para el aprendizaje y la evaluación*. Madrid: Narcea, S.A. de Ediciones.

Kohonen, Viljo. 2012. Developing autonomy through ELP-oriented pedagogy. Exploring the interplay of shallow and deep structures in a major change within language education. In Bärbel Kühn & María

Luisa Pérez Cavana (ed.), *Perspectives from the European Language Portfolio. Learner Autonomy and self-assessment*, 22-42. London and New York: Routledge.

Kühn, Bärbel. 2016. EPOS – the European e-portfolio of languages. *Language Learning in Higher Education*, 6(2). 335-354.

Language Policy Division. 2000. *European Language Portfolio: Principles and guidelines*. Strasbourg: Council of Europe.

Lenz, Peter & Günther Schneider. (2000). European Language Portfolio: Guide for developers. Strasbourg: Council of Europe. http://www.coe.int/t/DG4/Portfolio/documents_intro/Eguide.pdf. (accessed 14 May 2020)

Little, David. 1991. *Learner Autonomy: Definitions, Issues and Problems*. Dublin: Authentik Language Learning Resources.

Little, David. 2016. The European Language portfolio: Time for a fresh start? *International Online Journal of Education and Teaching*, 3(3). 162-172.

Little, David. 2020. The European Language Portfolio: Past Success, Present Reality, Future Prospects. Portfolioarbeit im Kontext von Sprachenunterricht. Neue internationale Entwicklungen, In Bärbel Kühn & Sandra Ballweg (eds.) *Portfolioarbeit im Kontext von Sprachenunterricht*, 17-37. Göttingen: Universitätsverlag Göttingen.

Little, David & Radka Perclová. 2002. *The European Language Portfolio: a guide to teachers and teacher trainers.* Strasbourg: Council of Europe. https://goo.gl/z21dV4. (accessed 14 May, 2020).

Open University Help Centre, The. 2020, Develop your career: Personal development plan (PDP) | *help.open.ac.uk*. https://help.open.ac.uk/pdp (accessed 14 May 2020).

Pegrum, Mark & Grace Oakley. 2017. The Changing Landscape of E-Portfolios: Reflections on 5 Years of Implementing E-Portfolios in Pre-Service Teacher Education. In: Tushar Chaudhuri & Bèatrice Cabau (eds.), *E-Portfolios in Higher Education. A Multidisciplinary Approach*, 21-34. Singapore: Springer.

Pérez Cavana, Maria Luisa. 2012. Fostering strategic, self-regulated learning: The case for a 'soft' ELP, In Bärbel Kühn & María Luisa Pérez Cavana (eds.), *Perspectives from the European Language Portfolio. Learner Autonomy and self-assessment*, 143-160. London and New York: Routledge.

Pérez Cavana, María Luisa & Sue Lowe. 2018. More than a tool: integrating the 'teacher voice' into an ePortfolio using the Three-layered PDP model. *The International Journal for Recording, Achievement, Planning and Portfolios (RAPPORT)* 3(1). 24-40

https://view.joomag.com/rapport-vol-3-rapport-vol-3-issue-1/0412019001542135381?short (accessed 14 May 2020).

Pérez Cavana, Maria Luisa & Sue Lowe. 2019. Personal Development Planning (PDP) and ePortfolios: Arranged marriage or true love? In Bärbel Kühn & Sandra Ballweg (eds.), *Portfolioarbeit im Kontext von Sprachenunterricht*, 171-188. Göttingen: Universitätsverlag Göttingen.

Stoicheva, Maria, Gareth Hughes & Heike Speitz. 2009. *The European Language Portfolio: An Impact Study*. Strasbourg: Council of Europe.

Trim, John. 2007. *Modern languages in the Council of Europe 1954-1997*. Strasbourg: Council of Europe Language Policy Divisio.

Ziegler, Nicholas A. & Aleidine J. Moeller. 2012. Increasing self-regulated learning through the LinguaFolio. *Foreign Language Annals*, 45(3). 330-348.

Zimmerman, Barry. J. (2000) Attaining self-regulation: a social constructive perspective. In: Monique Boekaerts, Paul. R. Pintrich & Moshe Zeidner (eds.), Handbook of Self-Regulation, San Diego Ca: Academic Press. 13-39.

8 Biographies

Dr. Maria Luisa Pérez Cavana (SFHEA) studied philosophy, history and education in Spain and Germany. She is a lecturer in languages at the Open University. Her main research interests have been learner autonomy, learner centred pedagogy and learning ePortfolios in relation to language learning. She is part of the expert team for the European Language Portfolio (ELP) at the European Centre for Modern Languages (ECML) in Graz, Austria. She has also developed a model of Personal Development Planning (PDP) adapted to distance education. She is currently using phenomenology as a research method to study the lived experience of learning a language.

María José Luelmo holds a PhD from the Rey Juan Carlos University and a Bachelor of English philology from the Autonoma University of Madrid (UAM). Since 2003, she has worked permanently at the Rey Juan Carlos University where she teaches various undergraduate subjects, mainly related to the teaching of the English language at different levels: Degrees in early childhood education, primary education and Master in teacher training. Her research work has focused on educational innovation, specifically on the implementation of Active Learning Methodologies with special attention and monitoring of the implementation of the European Language Portfolio (ELP).

Initial stages of individual teacher CEFR-related classroom curriculum projects at Miyazaki International College

Rebecca Schmidt, Miyazaki International College

Ellen Head, Miyazaki International College

This paper explores steps individual instructors have undertaken to incorporate the CEFR into oral communication, reading, writing, and team-taught CLIL (Content and Language Integrated Learning) courses absent of administrative mandates, within a large-scale curricular framework of the International Liberal Arts Department at Miyazaki International College (MIC). Although the curricular framework specifies general learner and course goals, a high level of teacher autonomy allows instructors to embark on their own projects integrating CEFR-related goals. For example, instructors have found ways to input larger conceptual goals of the CEFR through Can Do statements, and other reflective activities aimed to raise learner awareness, as well as noting where curriculum already aligns with established descriptors of the CEFR scales. In addition, aspects of the newer Illustrative Descriptors Scales of the CEFR Companion Volume (CEFR/CV) including reception, production, interaction and mediation provide benchmarks for instructors to reflect on the way language is taught and negotiated in the CLIL program. While the paper focuses on projects within such an autonomous environment, it offers a case study of steps individual instructors can attempt in order to incorporate the CEFR into curricula absent of administrative mandates and provides guidance for those receiving various mandates to reform curriculum.

Keywords: CEFR-informed curriculum design and mapping; teacher autonomy; learner awareness; learner reflection; CLIL; EMI institution; Can Do statements; Illustrative Descriptor Scales, CEFR/CV; action research approach

1 Introduction

Since the official launch in 2001, the Common European Framework of Reference for Languages (CEFR) has significantly affected language learning, curriculum planning, teaching practices, and assessment worldwide (Byram and Parmenter 2012; Piccardo, North and Goodier 2019; Szirmai 2014). Its action-oriented approach considers language learners as "users and social agents", promotes learner awareness, autonomous learning and "builds on and goes beyond the communicative approach" of functional-notional syllabi and curriculum planning (Council of Europe 2001 and 2018: 25-26). As Japanese universities and the Ministry of Education (MEXT) aim to improve English language education, the CEFR has emerged as a framework for curriculum development in Japan both at the secondary and tertiary levels (Morrow 2004; Nagai and O'Dwyer 2011; Cook and Rutson-Griffiths 2018; Schmidt, Runnels and Nagai 2017). Although this emergence was influenced by well-documented CEFR informed curricula within the European Context, many challenges remain with implementing the CEFR in European public education curriculum development as correlation between language "learning outcomes to CEFR levels lack in general empirical evidence," (EU 2013: 13). Nonetheless as the EU (2013) study concludes, "the more the CEFR is implemented

and used in policy documents, the more the CEFR is used in examination, schoolbooks and teacher training," (13). Indeed, MEXT's core curriculum document for junior and senior high school English language mentions the CEFR B2 as a target goal for language teachers to obtain (MEXT 2019: 7). Several students in MIC's School of International Liberal Arts (ILA) are also enrolled in the teacher certification course. Given this environment, there exists a potential for future mandates from MEXT or universities to show where curricula or class syllabi align with the CEFR A1-C2 reference levels. During the period from 2001 to 2010, over one hundred CEFR-based "language portfolio models" were validated by the Council of Europe (Little, Goullier and Hughes 2011). Creating local, context-sensitive models and having projects validated by a central authority was an important dynamic in the way that applications of the CEFR unfolded. In Asia, on the other hand, the dynamic interaction between the center and the periphery has generally been missing and the understanding of the CEFR was not so strongly rooted in a culture of locally initiated curricular development. CEFR is seen more as a comprehensive set of targets to be achieved. Awareness of possibilities of future mandates from MEXT, along with a commitment to learner autonomy and positive prior experience with the CEFR and portfolios, inspired the authors to investigate how the current curricular goals of the School of International Liberal Arts could be re-framed and clarified by drawing on the CEFR and CEFR Companion Volume (Council of Europe [CoE] 2001, 2018).

After explaining the background of the university as an English as a Medium of Instruction (EMI) institution, we proceed to detail the classroom initiatives taken to incorporate the CEFR concepts into courses being taught by the authors using an action research approach. First, the language program's oral communication course is discussed in terms of how its functional objectives aligned with CEFR descriptors and other reflective activities implemented. Secondly, we examine the use of Can Do statements for setting goals in reading and writing skills courses in the light of the CEFR. Finally, we discuss ways the descriptors of the CEFR Companion Volume (CV) can be used to analyze practices within one CLIL classroom.

It is worth saying at the outset that the two authors represent different perspectives on the CEFR. R. Schmidt had experienced the process of aligning to the CEFR, seeing how it could bring a whole staff together to improve materials, teaching and assessment at Hiroshima Bunkyo University. Bringing accuracy to a large amount of curricular data to sequence course levels and facilitate students' path through the curriculum therefore emerged as her major focus of interest. On the other hand, E. Head had worked in various communities of practice related to assessment for learning, starting with involvement in Cambridge suite exams in the 1990s, going on to work with the FLP and Language Portfolio SIG on classroom applications of the CEFR, and most recently working in China, where she was engaged in helping teachers to work with CEFR-based criteria as assessment for learning for a CEFR-based test called Aptis. Head's vision of the CEFR was as a reference point for cyclic processes of assessment in a learning community. Establishing processes of standardization of language assessment emerged as her major preoccupation during the time working on the project with R. Schmidt. Both authors share a strong commitment to transparency of assessment and strong linkage of assessment with classroom practice, which the CEFR promotes.

2 The context

2. 1 The School of International Liberal Arts

The School of International Arts (SILA) at MIC offers a liberal arts degree, a combination of an English Language Curriculum and CLIL courses. Generally, students study in Japan for the first three semesters, followed by a semester abroad. In years 3-4 they write a senior thesis with 1:1 supervision. Content includes English literature, sociology, political science, economics, history, psychology, anthropology, IT and science. In 2019-20, around 18% of freshmen were non-Japanese students. The trend towards recruiting overseas is creating new challenges in terms of identifying the level of content in classes and verifying acceptable standards for incoming students and for graduation (Brown 2017: 8).

2.2 EMI and CLIL courses at MIC

MIC College was the first university in Japan to allow students to graduate by taking all their courses in English (Mulvey 2018: 42). Before going on to consider the current English program in the light of the CEFR, we will look briefly at the domains of target language use (TLU) for students across their four years at MIC. Although much is made of the fact that students spend a semester abroad in second year, they generally join English-language programs appropriate to their level abroad, and so their on-campus classes aim to provide as much of a challenge as anything they encounter abroad (Bennett 2018; Bishop 2018). There has been a shift in practice at MIC from EMI (the term used by Mulvey (2018), referring to practice in 1994) towards CLIL, particularly in the content classes for first and second-year students. Originally classes were team-taught by content-experts with a language teacher to provide support but without structured language development. Mulvey explains how the MEXT accreditation process in 2007-8 led to improvements to the program. "MIC used feedback from its assessment review to initiate clearer delineations of level and learner appropriate outcomes for its EFL classes, not to mention better coordination between these classes and the EMI curriculum," (Mulvey 2018: 41). Looking at reports of first year content courses, we can see that teachers make an effort to tailor the material and goals to fit the needs of students. Hamiuc and Parker (2016) describe how they worked to make the targets of a religious studies course achievable despite conceptually challenging content. Only 81% of the vocabulary in their initial texts belonged to the first 2000 words of English of the K1 and K2 lists. Content objectives were redefined to suit the abilities of the students. These procedures are similar to those carried out on an ongoing basis by teachers of first year courses. Both authors have experienced this process as co-teachers of Psychology, Sociology, Anthropology, and History. We note that although these courses have been adapted by language teachers, the amount of subject-specific vocabulary is still overwhelming for students. Schmidt, Jiang and Grabe (2011) believe that readers can only read independently if they have 98% coverage of the vocabulary level of the text. A study by Bennett (2018) of vocabulary needed by students for aural comprehension of the Japanese Popular Culture course showed that the vocabulary used in lectures and videos was 95% covered by the NGSL list of 2,800 most frequent words. Assuming students had mastered the NGSL by the time they took the Popular Culture course they would still have 5% not covered. Bennett estimated that one in 15 words heard would be unknown to students (Bennett 2018: 17-18).

The issue of starting to do English-medium courses with a very low level of English has been the subject of global debate in recent years. Breeze (2014: 146) raises important questions: "How far does knowledge of the target language impact [students'] chance of doing well?" and "How can content teachers provide help without sacrificing quality?" She reports a study of correlations between listening levels and success in content classes such as law and medicine at a Spanish university, concluding that above a certain threshold, study skills, commitment, memory and information-gathering can help students succeed even with a relatively low listening level. In our context, introducing a minimum entry requirement has been discussed but currently most first and second year courses do not have such a minimum. However, students' feedback questionnaires show that CLIL classes are motivating for lower level students, particularly when they have the opportunity to work in a group with peers who are of a higher level. The more proficient students can help by explaining concepts and modelling the desired products and behavior to their classmates, and students employ different levels of mediation to gain understanding. We perceive this as an area in which CEFR/CV would be an invaluable tool for developing our understanding of student needs in relation to CLIL in our specific context. Hitherto, approaches have focused on quantitative analysis of material encountered by students. In section 3.5.1 below, we explain how text analysis tools were used to simplify the vocabulary required for the Introduction to History course.

2.3 The English language curriculum: Core Program

The program, outlined in the language program handbook (Bennett 2017), began in 2011 and has been in its current form since 2015. Students are streamed into levels for the 3 core courses: oral communication, reading, and writing. Content classes are mixed-ability, but a language instructor team-teaches with a content instructor. Besides the language teachers being familiar with content of the language program, there is little sequencing or scaffolding for language needed to do the content classes beyond overarching goals to encourage discussion and critical thinking. The English program, "planned, developed, and revised by faculty" aims to "foster students' ability for self-expression, questioning, evaluation analysis, and creative thought through the medium of English" (Bennett 2018: 1). Rather than CEFR, vocabulary was the guiding principle in the construction of the curriculum, as Bennett (2018: 1) explains: "The underlying ethos of the program loosely follows Nation's (2007) *Four Strands* approach to curriculum design, in which learners are exposed to an approximately equal balance of activities designed to provide meaning-focused input, meaning-focused output, language-focused learning, and fluency development." Courses were developed considering these strands, where material for reading (input) and speaking and writing (output) aimed to be familiar to the learners (Nation 2007). All of these components were incorporated into a spiraled curriculum of "topical linking of classes" for the three courses in the Core Program (Bennett 2017: 1).

Although the CEFR was not an inspiration for the course objectives, there appears to be overlap in its goals. Moreover, conceptual goals to "develop sufficient language proficiency and learner autonomy to become successful participants in the MIC setting and in the larger English-medium global community" (Bennett 2017: 1) coincide with conceptual goals of the CEFR. It is from this starting point that the authors began to evaluate specific ways the curriculum already aligned with the CEFR and how it could be improved by using CEFR in future.

2.4 Teacher autonomy and larger curricular goals

The learning objectives for each course remain general enough to allow for teacher autonomy in the ways teachers create or supplement program material in each of their individual classes. Such high teacher autonomy and flexibility provides an ideal environment for teachers to embark on their own classroom level projects. Cook (2019) discusses the disadvantages of teachers being too independent, "working in silos" as they often tend to do in university settings to the point that it can "inhibit the exchange of knowledge amongst staff" (Swap and Wayland (2013) in Cook 2019: 2). However, he outlines ways that these individual classroom practices, initiatives, and activities can be brought together to enhance a more collaborative program. It is possible our projects could influence the Core SILA Language Program. As CEFR-informed curriculum benchmarks, assessments and conceptual CEFR classroom practices grow in importance in Japan, it will be beneficial for the language program at MIC to have explicitly defined ways in which its curriculum aligns with the CEFR.

3 The classroom projects

3.1 An action research approach

Action Research is a common approach in Educational academic research which follows a cyclical method. Researchers identify a problem, attempt to implement a solution, collect data, reflect on the process and results, and evaluate those results to create a better solution (Mills 2006). However, as Burns (1999: 35) notes, "models such as these" can be seen as "too prescriptive" as researchers "will need to make their own interpretations of what are appropriate processes for the circumstances of the research." Therefore instead of a cycle, researchers may prefer to see the process "as a series of interrelated experiences involving the following phases: 1-exploring, 2-identifying, 3-planning, 4-collecting data, 5-

analyzing, reflecting 6-hypothosising, speculating, 7-intervening, 8-observing, 9-reporting, 10-writing, 11-presenting," (Burns 1999: 35). The authors have chosen to adopt this model of action research to explain classroom projects being undertaken by the authors. This paper discusses the initial stages of these projects to attempt to investigate ways the CEFR could more explicitly be incorporated into their individual classroom projects and ultimately investigate ways the CEFR could be used in the overarching curriculum. As discussed in the introduction, MEXT explicitly states goals in its core curriculum for junior and senior high teachers to have a CEFR B1 level (MEXT 2019: 7). Thus, it is possible, MEXT could ask universities to demonstrate ways the CEFR informs certain course curricula. As both authors had experience with the CEFR before joining MIC, they noticed the lack of any official mention of the CEFR in the curricular goals both conceptually and in ways its learning objectives overlap, despite existence implicitly. Therefore, in terms of action research, this lack of explicit mention serves as the 'identified problem'. Yet, much of what has been done is still in the exploratory and planning stages as they have attempted to collect preliminary data to note existing ways the CEFR fits into the current curriculum. This paper will explore the processes of 'exploring', 'identifying' and 'planning' their classroom projects.

In addition to noting explicit lack of mention of the CEFR as the identified problem, the authors seek to identify other ways the CEFR might possibly improve aspects of the curriculum, such as leveling of classes, material development, assessment and allowing students to identify their level of English beyond the TOEIC. Nonetheless, in the case of these initial classroom projects, the goal is showing ways the CEFR could be used in the classroom as evidence for how it could be implemented at a wider curricular level. To put it another way, the CEFR contains tools and processes for sorting out levels in language programs, and these tools would add clarity and coherence to our program in terms of organizing materials and assessment. Thus, the projects discussed serve as attempts to 'identify' the issues, clarify 'planning' and ways to go about 'collecting data' in current and potential future projects with the ultimate goals of making the CEFR more explicit in individual teacher practices and ultimately program curricular goals. We have chosen to define our discussion of projects in this paper as *initial stages* to stress that we are still in the 'planning' steps of collaborative projects involving CEFR use at MIC.

3.2 The oral communication course

The oral communication course is one of the three main courses in the Core Program of the language program. According to the SILA English Program Handbook (Bennett 2017: 8), the primary aims of oral communication are to fluently perform "communication tasks for an academic setting, using controlled vocabulary and grammar." Benchmark goals are set in terms of six functional objectives and correlating grammar points for each level. In addition, vocabulary targets are set, based on the most frequently-occurring words according to the NGSL. Vocabulary development is assessed within the Reading course and will be discussed further below. Students are assessed through speaking production and interaction tasks, grammar quizzes and a final oral and grammar exam. Although the goals of the program are centered on an academic setting, many of the functional objectives are everyday communicative tasks, such as describing daily routines and past experiences. The functional-notional approach for these learning objectives, and overall curricular goal of learner autonomy, coincide with the emphasis on social agency and action-oriented approach of the CEFR (CoE 2001, 2018). Thus, this course provided an excellent starting place to explore ways the language program aligns with CEFR goals and descriptors.

3.2.1 Using the Hiroshima Bunkyo University model

In 2012, the administration of the English Communication Center at Hiroshima Bunkyo University decided to revise its General English (GE) curriculum to make it more CEFR-informed. As one author was a part of this project from 2012 through March 2018, it was the inspiration and model to begin evaluating ways the CEFR could be implemented in the oral communication course at MIC. Although course

objectives and specified functional objectives were clearly defined in the SILA language handbook, high teacher autonomy was granted to individual instructors to develop material and decide approaches to teaching the functional objectives. This provided an ideal environment to use the experiences at Hiroshima Bunkyo University to map functional objectives and conceptual course objectives in the Oral Communication course at MIC aligned with the CEFR.

Although the Bunkyo project began with administrative mandates and all faculty were given roles in the project, it still provides a model for how one teacher could initiate a similar approach within a single classroom without administrative mandates. For the purpose of briefly explaining the CEFR project at Bunkyo, the author has defined it in terms of three stages. The first stage involved mapping the existing Communicative Language Teaching curriculum to the CEFR-J, the Japanese version of the CEFR. Mapping in this project and in the ones discussed in this paper refers to the process of investigating current curriculum content and goals to see how they align to Can Do descriptors. The CEFR-J was chosen for its greater micro-levels, which were deemed more applicable within the context of low-level of English of the students at a Japanese university (Negishi, Takeda and Tono 2013). First, teachers wrote Can Do statements for each lesson in the General English Program. Then these were mapped (examined to see how they aligned) to the descriptor levels of the CEFR-J. One Can Do statement was chosen for each lesson and placed at the beginning and end of it. Also, overall curriculum level Can Do statements were created. More details on this first stage of the Bunkyo curriculum reform can be found in Bower et al. 2017 (also Kodate 2017; Schmidt 2018). The mapping revealed many inconsistencies and gaps in the program so it was decided to rewrite the entire curriculum. Due to lack of external resources like grammar and vocabulary lists, a decision was made to use the CEFR instead of the CEFR-J. Thus, stage two consisted of redesigning the entire curriculum based on the CEFR A1-A2 and A2-B1 levels, aligning tasks in the Self Access Learning Center (SALC) and developing CEFR-informed assessments. Stage three involved creating and mapping Can Do statements for Global Communication Department courses (Cook 2019). Experiences learned in stages 1 and 3 on writing Can Do statements and noting their alignment with the CEFR provide the background model for mapping the Oral Communication functional objectives at MIC.[1]

3.2.2 Mapping the functional objectives

The Language Program Handbook (Bennett 2017) for SILA at MIC clearly lists learning objectives, including six functional speaking objectives for each level of Oral Communication. Thus, the first step was to evaluate where these goals are similar to the CEFR descriptors. Using the experience from experiences at Bunkyo as discussed above, the author chose to investigate how they aligned with the Eaquals (European Association for Quality Language Services) Core Inventory for General English, an interpretation of the CEFR into English language learning goals. In addition, other sources such as the digital grammar analysis provided by English Profile website were used to help map functions and grammar targets. (English Profile 2015). Analysis of the first three functional objectives can be seen in Table 1, below.

Although the curriculum was not based on CEFR-informed goals and benchmarks, our functional objectives and target grammar correlate well with descriptors for understanding, spoken interaction and production in the inventory. Students enter the university with an average TOEIC score of around 350 and start their second year around 460. With scores 225-545 correlating to an A2 level (Tannenbaum

---

1. It should be noted that the mapping of the Oral Communication discussed below is the work of one instructor, the first author of this contribution, who had the experience of mapping course objectives at Hiroshima Bunkyo University, but is by no means an expert. While we recognize that the process of mapping to the CEFR has been carried out numerous times by curriculum and test-developers, it was decided to draw on the models known for this initial step. The primary purpose of this project is to give a case study of what individual instructors could do to initially see where the material they are teaching lies on the CEFR scales. Identifying correspondences between our curricular goals and CEFR descriptors was the first step taken.

and Wylie 2015), we can see the program is fairly well leveled for the first semester but with an increase in B1/B2 content in the second semester, it becomes increasingly challenging for lower level students (see Appendix 1).

Table 1. *(see Appendix 1 for all functional speaking objectives for the entire Oral Communication course)*

Functional Speaking Objective (as listed in the language Program Handbook)	Functions	Grammar
Oral Communication 1: Spring Semester (1st year)		
1. **Managing conversations** (*wh-questions / yes / no questions*)	**A1**- Giving Personal Information (family life, hobbies, leisure activities)	**A1:** Questions / To be / Verb + -ing
2. **Classroom English**	**B1**- Expressing Opinions / Expressing agreement and disagreement / Taking the Initiative in Interaction / Checking Understanding / Managing Interaction	**A2:** Modals - should
3. **Describing routines & habits** (*present simple tense*)	**A1/ A2** - Describing habits and routines	**A1 / A2:** Adverbs of Frequency **A2:** Adverbial phrases of time **A1:** Present Simple Tense

A brief overview of the mapping also reveals a general progression in difficulty from A1 to B2, although it is not completely linear. However, it should be noted that in other classes students are tackling CLIL content in the B2-C1 range, despite efforts to simplify. Giving opinions, expressing an argument, and joining in group discussions, which fall into B1 and B2 descriptors, are of utmost importance in CLIL courses. Thus, these functions are placed at the beginning of each semester in Oral English. This discrepancy of levels also highlights weaknesses in the program where A2 level students are mixed with B2 level students in the CLIL courses, yet under these circumstances it is appropriate to have some oral communication course goals at the B2 level. The Oral Communication classes are streamed into four levels and teachers have freedom to modify content to best meet the needs of their students. This need for flexibility thus means the functional objectives remain as overarching goals and make a true mapping of content to CEFR difficult at this stage. This initial investigation thus reveals that actual course content each teacher uses at each level should be examined more closely. Such projects could be embarked on as a way to collect more data to determine real alignment of the oral communication course.

3.2.3 Can Do statements and reflection tasks

In addition to the mapping of the functional objectives, the first author aimed to add more conceptual goals of the CEFR to enhance learner awareness of autonomous learning, in accordance with the overarching goals of the SILA. One of the main purposes of incorporating the CEFR into the Oral Communication course is to increase awareness of the learning objectives and provide means for reflection. Again, experiences learned from the Bunkyo project were applied. First, one single Can Do statement was created for each functional objective. This Can Do statement, with a four-point Likert scale was placed at the beginning of a unit created for each functional objective. This Can Do had two boxes, one to check before the lesson and one after.

Table 2

How well can you do the following objective?		I Can Do it easily	I Can Do it	I Can Do it, but it is difficult	I can't do it
I can ask and answer questions and discuss my life, daily routine, and habits with classmates.	Before				
	After				

Although there were various handouts for each functional objective created by present and past instructors, the sequencing was disjointed. It was unclear whether students were even aware of what the learning objectives were unless they read the syllabus. Although there was some sharing of material, this high teacher autonomy had resulted in numerous single page handouts without any overarching coherence. Thus, creating one single handout has also helped increase collaboration with material design. Finally, reflective activities were created for students to note weaknesses and goals to improve for each speaking production and speaking interaction tasks. Functional Objective performances were recorded using *Flipgrid,* a website designed for students to upload videos with space for teacher evaluation and feedback. Other tasks, such as presentations were recorded. The new reflection worksheet helped students to reflect on their pronunciation, fluency, conversation strategies (interaction tasks) and actual use of the target functional objective for each speaking task. It is hoped these could be used by other teachers on a voluntary basis in future.

3.3 The Reading Course

The MIC Reading course was designed to build the skills and vocabulary students need in content classes. The goals cover reading strategies, grammar, vocabulary, dictionary use and reading fluency (Bennett 2017: 39). In contrast to the Oral Communication course, the materials and sequence are specified in detail. In-house readings 500-1000 words long provide the backbone of the course. The second feature of the Reading course is mandatory use of applications to achieve targets for vocabulary study and extensive reading. This reduces the freedom teachers have to create extra projects as students must devote homework time to these applications, leaving no time for tasks such as book reports or reading circles. The targets are demanding in terms of breadth and volume, leading many students to struggle with managing their time outside class. Table 3 gives an overview of the program.

Table 3

CEFR/CV	MIC Reading program	
Reading for orientation	Reading for speed and fluency	2 passages/week
Reading for information/ argument	In-house intensive reading passages	8x 500 word passages per term
Reading for information/ argument	Read Theory (2020) (reading comprehension training app for native speaker children)	12-15 sessions per month
Reading for leisure	Extensive reading using Xreading VL (2020) Virtual Library app	15,000 words/month
	Vocabulary study using Praxis Ed (2008/2020) app	12-15 sessions/month

In addition to time management issues, the variety of ability levels adds to the problem. Even though the classes are streamed, reading speeds and vocabulary levels vary widely. This became the focus of the second author's action research project for the reading course. In the next sections we attempt to relate the goals of Reading 1 and 2 to the CEFR, before examining how Can Do lists were used to help students to make individualized goals.

3.3.1 Mapping curricular content to the CEFR/CV

CEFR/CV conceptualizes reading progression in terms of ability to understand information on a cline in terms of "length and complexity/simplicity of message", "the extent to which the subject is an everyday one, related to interests, or specialized" (CoE 2018: 61), and "type of language, from simple to stylistically complex" (CoE 2018: 65). Our intensive reading material follows a progression from the familiar to unfamiliar. The initial texts describe familiar subject matter: the classroom, successes and failures of English study and school life. However, students encounter complex syntax from the start. This is done deliberately to challenge students (Bennett 2018: 17). Questions focusing on the line of argument appear early in the course. This equates to a B1 level goal: "Can recognize the line of argument in the treatment of the issue presented, though not necessarily in detail" (CoE 2018: 63). The readings include summarizing and paraphrasing tasks. This coincides with B1 level of processing text in speech in the CEFR/CV: "Can summarize the main points made in clear, well-structured spoken and written texts" (CoE 2018: 111).

The intensive readings require sophisticated syntactic and vocabulary knowledge. The following example, from Reading 1, may serve to illustrate the kind of complexity students encounter. "As a human being, you have a natural talent for learning the advanced form of communication that we call language." The accompanying questions include "What do humans have a talent for?" Most students respond "the advanced form of communication that we call language", while the desired answer is "language". This creates an opportunity to teach students about reading for the main idea. By working on grammar and rhetorical structures in this way, it is hoped students will learn to analyze complex sentences by themselves. But in practice, many students try to translate into Japanese. Beginning readers might be better served by easier material.

CEFR resists specifying numerical parameters for reading speed and vocabulary size, due to the variations in parameter/skill correlations for different languages (Milton and Alexiou 2009: 196). At first sight, MIC Reading course goals appear unrelated to CEFR, in that they refer to mastery of vocabulary according to NGSL levels and increasing reading speed (Bennett 2018: 5-8). The intensive reading texts in the first semester draw on NGSL level 1 to 3 words (roughly the most frequent 1,500 words) and vocabulary gradually expands to include level 4 and 5 (most frequent 2,800 words) by third semester. The same wordlists are assessed in a vocabulary test which makes up part of their reading class score. The test is done at the start of the program and at the end of each semester, to measure not only individual performance but also program effectiveness (Bennett 2017: 17-19).

The applications (Praxis Ed, Xreading VL, Read Theory) give instant feedback, showing not only correct or wrong answers, but also the number of attempts, the time spent on each page or problem, and the stats for their previous sessions, in great detail. Tracking quantitative information such as vocabulary scores and reading speeds allows students to see their progress. However this approach is less strong on the instrumental and integrative aspects of reading, which the CEFR highlights. The emphasis on quantitative targets makes it difficult to focus on the development of integrated skills and overarching competencies.

3.3.2 Using Can Do lists for goal-setting in Reading 1 and 2

A Can Do list was constructed based on the program handbook goals for Reading 1 in Spring 2019 and 2020 (see Appendix 2). Using the list at the beginning, middle and end of the course was intended to

help students to prioritize and make choices between various activities, particularly for out-of-class study. At the time the Can Do list was first administered, students were told that they should try to work on their own goals and make some progress from their own initial level rather than competing with each other. Different students responded to the Can Do list in various ways. For lower level students, improving their reading speed, vocabulary and reading amount were the most popular goals. The list worked well within the program because feedback was available quickly on quantitative goals such as reading speed. Higher level students picked more sophisticated goals such as inferencing, identifying tone, trustworthiness and humor.

Since both authors had been using Can Do lists in their classes individually, in Spring 2019 they decided to try working together to relate their work more closely with CEFR and CEFR/CV. This project stimulated the second author to investigate how her program-based Can Do statements could be calibrated with the CEFR/CV (see Appendix 2). This process highlighted the challenging nature of Reading 1 for students who enter the program below B1 level. Of the 11 items on the Reading 1 Can Do list, six relate to B1, two to B2, two to A2 and two are not in the CEFR (using punctuation marks to help with reading aloud, reading 200 words per minute). The possibility of creating differentiated targets to cater for lower levels in the program was discussed during 2018-19 but rejected. It was thought that it might be difficult to change the targets, because it would be perceived as unfair by high level students who had more demanding targets. It might also be unfair to deprive lower level students of the potential gains they would make if they studied hard.

Given the many fixed elements in the reading curriculum, it is difficult to combine the mandatory elements into an assessment which is learning-oriented, achievable for lower levels and motivating for higher level students. Most teachers include the recognition of effort and improvement in the final grade, evidenced by hours spent in on-line study and meticulous performance on the intensive reading worksheets. The adoption of Can Do lists in one individual class was an attempt to help students to communicate with the teacher about what they found difficult, and calibrate their own progress independently. Systematic feedback needs to be gathered in the next stage of the project in order to determine whether this was effective from the students' point of view.

3.4 Academic Writing Within the English Program

In this section we will look at first year Academic Writing at MIC and describe current practice and potential improvements which could be made by implementing CEFR-derived tools and processes such as Can Do lists, portfolio assessment and collective standard-setting. The course raises similar issues to those described in the section on oral communication, in that there is a high degree of teacher autonomy and a wide range of levels being graded within one program. In contrast with the Oral Communication course, there is no program-wide collaboration on grading of the final test, even though the attainment targets and test are the same. Although the situation is not viewed as problematic by faculty, the need for increased accountability in 2020-21 due to grade-linked criteria for government scholarships may create a need for working towards standardization in the future.

Over two semesters (Academic Writing 1 and 2), the course aims to teach students how to write an academic essay, building up from sentence to paragraph to a five paragraphs essay over three semesters. A third semester offers students a chance to work more on five paragraph essays, but we focus on first year classes here. Goals are specified in the program handbook in terms of grammar and functions ("rhetorical styles") such as classification, persuasion or comparison. (Bennett 2017: 23). The functions mirror those introduced in the Oral Communication course. Writing fluency is a further goal, addressed through timed writing practice and journaling. The course is assessed through continuous assessment and a program-wide final exam featuring essay-writing and sentence-writing tasks focused on the material introduced in class. An in-house student writing handbook provides back-up in the form of example essays and paragraphs, detailed explanation of relevant grammar and cohesive devices. As

with the other courses, most of the tasks require B1 or B2 in CEFR terms to be completed adequately. Teachers in the lower level classes spend more time working on basic grammar and paragraph-level. In practice, teachers' norm-reference to individual classes to enable weak but industrious students to pass. Problems sometimes arise when students' progress to third year and start writing their senior thesis, as the three-semester course is not adequate to prepare them to organize a 4,000 word academic essay.

3.4.1 Supporting transparent assessment: rubrics and Can Do lists

During 2015-19 the whole college implemented a drive to use rubrics for assessment and make the rubrics available to students through documents such as syllabi. The student writing material contains a bilingual rubric in English and Japanese, for scoring paragraphs and essays, using the criteria organization, lexis, grammar and content. However the rubric is designed to assist teachers with grading. It is too condensed to be used for scaffolding an on-going awareness of the development of subskills. Yet for students to achieve the goal of writing better paragraphs or more complex types of text, subskills need to be listed and worked on one by one. The author felt that Can Do lists would give students a sense of their own progress and needs, in a more dynamic way than the rubrics. The effectiveness of this methodology should be investigated by survey research in 2020-21.

As part of a classroom project to increase students' understanding of the mechanics of writing, Can Do lists were created based on the goals set out for the writing courses in the program handbook (Bennett 2017) (see Appendix 3). The list was used at the start, middle and end of term aiming to familiarize students with the terms needed on the course (punctuation, paragraph, draft, re-draft) and give students a chance to make short-term goals. Keeping a tally of students' self-rating on Can Do lists helped the instructor decide when to move on from single-paragraph to two-paragraph essays. The Can Do list was also useful to guide reflection at the end of the course. Most of the Can Do statements for our course were detailed, task-related statements which fit into B1 "Overall Written Production." But our goals also drew on overarching production-related competencies: planning, grammatical accuracy and vocabulary range. Goals which related to various sub-tasks of process writing were difficult to account for within the CEFR/CV framework and often overlapped with competencies related to speaking. For example, planning writing using a mind-map or list, seemed to be subsumed in CEFR/CV under planning a speech (CoE 2018: 78). Similarly, developing an argument, accounted for in CEFR/CV under "coherence and cohesion," appears to be presumed transferable from speech to writing (CoE 2018: 141). The process of attempting to map our curriculum-based Can Do list to the CEFR/CV appeared to highlight areas of the CEFR/CV which would benefit from further details being added.

3.4.2 Future steps: Portfolios and a round table on writing

Although the course encourages process-writing and learning journals, portfolios were not being used for assessment by any writing teachers in 2017-2019. The college had recently started using the Mahara system to make e-portfolios for IT, with the result that "portfolio" was understood by students to mean something to do with IT skills rather than reflection. However, the second author was very interested in trying out portfolios after using them in her former workplace. Re-reading "Developing an ELP Model" she was struck by the question "Will the learner be able to trace their development?" as it points towards the need for students to go through a process of looking at their old work and comparing their old and current productions. From 2017 to 2019, a final, reflective task was set where students mentioned increases in writing speed, sentence complexity and the ability to construct essays in their final reflections. However, in summer 2020, an example portfolio was made using written work collected from a student in third year (with permission) and students were asked to select work for their own portfolios, aiming for a balance of spontaneous and planned writing, showing evidence of their progress between April and July. At the time of writing we await the first round of portfolios.

E-portfolios have a great merit in bringing together student work in a form that can be conveniently sampled by various faculty members. In order to make grading fair, it seemed important to discuss samples of student output in relation to grading criteria. Two out of five writing teachers met to compare the grading of selected essays written during the tests in summer and fall 2018. The teachers agreed on the order in which they ranked the essays but their way of justifying their decision varied widely. However, in 2019, such collaboration did not happen. It is hoped that carrying out successful pilot projects using Can Do lists and portfolios will inspire faculty to work together on a standard setting.

3.5 CLIL Courses

3.5.1 Text Analyzers as a means to identify levels

CLIL classes present a particular challenge both for students and teachers. Students select their courses according to interest, leading to a wide range of ability in one class. The average student is around A2 level yet there are numerous students above and below. The authors were concerned about vocabulary difficulties encountered by students in the first semester. As mentioned above, Hamiuc and Parker (2016) and Bennett (2018: 14-16) have both used lexical frequency profiling to explore the vocabulary encountered in CLIL courses, in Religious Studies and Japanese Pop Culture respectively. Hamiuc was concerned with the vocabulary encountered in reading. Bennett analyzed samples of the vocabulary encountered aurally, by transcribing lectures. In 2019, the language instructor of a new Introduction to History course (author 1) analyzed the class texts to evaluate CEFR levels and identify CEFR vocabulary levels and structures requiring simplification (Figure 1). The project is ongoing in that only initial analysis has been done. The next stage is to evaluate vocabulary, rewrite the texts with simpler grammar, and rerun the analysis, aiming to produce material at B1 level. This project is challenging to follow up, since not only vocabulary but also syntactic complexity impact on difficulty of understanding, and how these interact is not well understood. On the other hand, focusing on sharing classroom practice in relation to schema activation and other methods of scaffolding understanding by active learning and group work are approaches which are well established in the college (Mork and Howard 2015: 74).
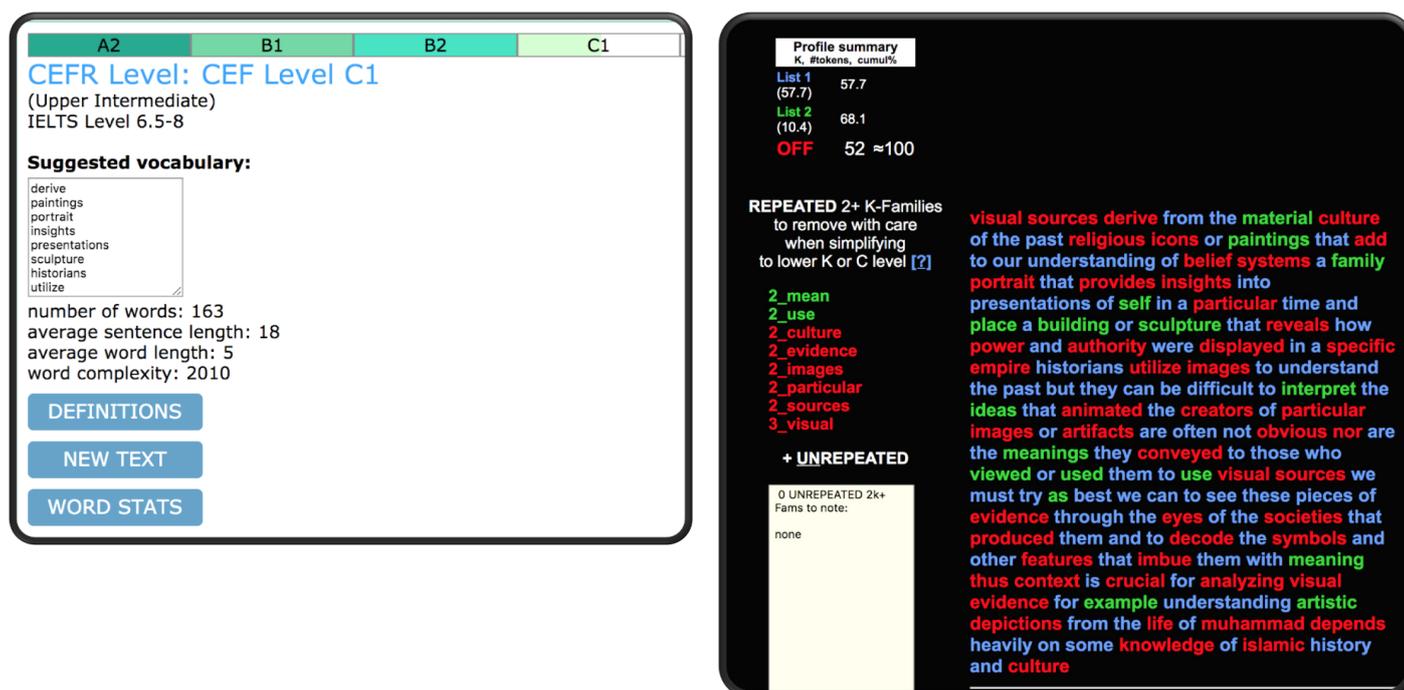
List 1 refers to KET vocabulary list A2 / List 2 –PET(B1)

Figure 1. Text Analysis of a CLIL text using text analyzers: roadtogrammar.com/textanalysis/ and Lextutor.ca.

3.5.2 Using the CEFR/CV illustrators to evaluate levels

In addition to using text analyzers to identify content level, the CEFR/CV illustrative descriptors provide ways to evaluate the necessary skills for CLIL/EAP. The analysis immediately reveals that language skills at a B1 level are essential to participate successfully in CLIL courses. Students on the Introduction to History course listen to lectures, use critical thinking to discuss the lectures in small groups and then write answers to discussion questions. The CEFR/CV descriptors are very useful for relating actual classroom performance to levels (see Appendix 4). More detailed analysis, especially in terms of mediation descriptors is planned for the future. Moreover, input from other instructors of other CLIL courses would be helpful before proceeding further in evaluating ways CLIL courses align with CEFR/CV descriptors. The descriptors also have great potential as a means for students to self-evaluate their progress, particularly in relation to mediated skills. Currently there is not much discussion of student perspectives in relation to needs analysis in CLIL classrooms at MIC. We hope that in future, implementing needs analysis through small-scale surveys of students drawing on the descriptors for mediated skills will provide triangulation for the quantitatively focused text analytics and might also reveal learner perspectives that we are not aware of.

4 Conclusion

4.1 Next steps in the action research cycle: further planning and data collection

In conducting these initial steps of individual classroom projects, the authors were able to explore how the existing program relates to the CEFR, identify problems in doing so, and provide more clarification in future planning and ways to go about data collection in their investigation of how the CEFR could be further utilized in the MIC language program and language aspect of CLIL courses. The authors found that such collaborative action research, as Burns (1999) noted in surveying teachers of their experiences, that indeed such collaboration helped them "engage more closely with their classroom practice … to explore the realities they face in the process of curriculum change," as well as "understand the reasons and need for institutional curriculum change more clearly", (Burns 1999: 14-15).

The process of initial mapping and documenting ways the Oral Communication, Reading, Writing, and CLIL courses related to the CEFR and CEFR/CV, as well as seeking more transparent and fairer assessments, allowed them to recognize they were indeed only in the planning stages of their action research project, where initial data collection served as a pilot. In relation to Oral Communication, the sequencing of the functional objectives was shared with other teachers and proposals were put forward to overhaul the in-house speaking test, moving in the direction of relating the speaking test more closely to the functional objectives. In relation to Reading, discussion with colleagues focused around the possibility of differentiated targets for reading to reduce the burden on lower level readers and allow them to build up reading fluency before undertaking more challenging work. Regarding the use of Can Do lists, both authors recognized the need to have their selection of Can Do items scrutinized by colleagues as a preliminary form of validation, and decided to seek opportunities to do this within the framework of a faculty development session. It was hoped that content-teaching colleagues might become more engaged with the task of simplifying and sequencing language within their courses as a result of exposure to our ideas. Finally, further investigations of student perceptions of the usefulness of Can Do checklists were also agreed upon as a follow-up that could be carried out by teachers on an individual basis.

Several weaknesses were identified. First, relying on only one author's experience on curriculum mapping limits the conclusions and led to subjective judgements of results. Exploration of other evidence-based cases for mapping and alignment of curriculum to the CEFR is needed especially when looking beyond the Oral Communication course and the wider university targets. A different approach for CLIL courses is needed. In particular more exploration of mediation descriptors of the CEFR/CV is needed. Mediation especially between Japanese and English is an ignored skill within the program

because of an inherited "only English" rule within the program. Second, more collaboration is needed with other language faculty to assess the need for change to generate more effective data collection and to agree on processes for more transparent and fairer assessment. Doing so would provide much more in-depth data collection of the current program at MIC. For example, reflective projects such as adding Can Do statements to the lessons and surveys of students' self-assessment of abilities and portfolios should be shared with other faculty. This should then be coupled with questionnaire feedback from colleagues and students. This would provide more evidence for the conclusions made by the authors, as well as, allow them to investigate how helpful such additions of reflective practice were for the students in raising awareness of their language learning goals. In addition, a round-table discussion and faculty development seminar on evaluating productive skills, looking at examples of output and comparing how different instructors' rate them is needed. Doing so would lead to the development of more effective rubrics in the writing course, as well as, in the Oral Communication course.

4.2 Limitations and need for further research

In addition to the need for more extensive literature review of evidence-based cases for mapping and alignment of curriculum to the CEFR descriptors, much further research needs to be conducted in the area of EMI contexts and defining the target language use domain for the wider academic EMI program at MIC. In terms of action research, the projects of analyzing vocabulary profiles for the CLIL readings and ways in which the CEFR/CV relates to these courses are in an even more preliminary stage of exploring and identifying the issues to be examined. Although this paper outlined these initial efforts, without further literature review of how to evaluate vocabulary profiles against the wider academic TLU domain, the next steps of planning and effective data collection to evaluate these courses cannot be effectively conducted. As discussed in section 4. 1, collaboration with other faculty will provide insight into the direction which needs to be taken both with content curriculum and assessment reform. Gaining a greater understanding of the TLU at MIC, which attempts to prepare students for studying authentic content in an EMI setting, is essential for any claims of true alignment to the CEFR.

4.3 Overall reflections

In conclusion, initial steps of projects discussed in this paper have been described in terms of an action research approach, where each project is part of a larger goal to discover ways the CEFR could be more explicit in the program curricular goals and individual teacher practices. We have explored the utility of CEFR descriptors and CEFR-inspired tools in our own classes, and the process has created opportunities to dialogue together and with colleagues. It can be seen that the CEFR and CEFR/CV highlight gaps in stated goals and inconsistencies in level progression not only within the English program but also in CLIL courses. The writing of this paper has offered a chance to reflect with the intention of improving what we as individual teachers offer to learners, both in teaching, assessment, and out-of-class support. Our projects are limited to our own classroom, but they have helped to keep us and our students engaged. In terms of the bigger picture, further work on the assessment and level-management system is a perceived need in our university. Since 2019 the proportion of overseas students is increasing, including those who may stay for a single semester or year. As the diversity of the student body increases it will become more important to have a framework which corresponds with those used by other institutions globally. In addition, the introduction of means-tested government scholarships in 2020 will increase pressure to standardize grading. The existing assessment system is well designed but leaves some of the target skills and abilities unaccounted for. The CEFR/CV offers a descriptive framework which has proved useful in our own classes and has been used by other universities to relate their curricular levels to those of testing bodies and national frameworks. It is worth remembering that the CEFR was always intended to be adapted to fit local contexts (North 2010) and aid dialogue between classroom teachers and those working to create externally referenced, objective standards. It remains to be seen how we

at MIC will deal with our assessment challenges but we hope that our work will prove helpful to others in our college and outside.

5 References

Bennett, Philip. 2017. Miyazaki International College English Program Handbook. School of International Liberal Arts. [unpublished]

Bennett, Philip. 2018. Vocabulary Learning in the MIC International Liberal Arts Program and the skills of speaking, listening, reading and writing. *Comparative Culture, The Journal of Miyazaki International College* 22. 3-21. [Special Issue]. https://www.academia.edu/38218977/Vocabulary_Learning_in_the_MIC_International_Liberal_Arts_Program (accessed 19 February 2020)

Bishop, Katherine. 2018. Focus on Topics in American Literature: A Love Letter to Teaching Epistolary Texts. *Comparative Culture, The Journal of Miyazaki International College* 23. 20-28. https://www.mic.ac.jp/files/uploads/Comparative_Culture2018.pdf

Bower Jack, Rutson-Griffiths, Arthur, Runnels, Judith, Schmidt, Rebecca, Cook, Gary, Lehe, Lyndon L. & Kodate, Azusa. 2017. Aligning a Japanese University's English Language Curriculum and Lesson Plans to the CEFR-J. In O'Dwyer, Fergus, Hunke, Morten, Imig, Alexander, Nagai, Noriko, Naganuma, Naoyuki & Schmidt, Gabriela (eds.). 2017. *Critical, Constructive Assessment of CEFR-informed Language Teaching in Japan and Beyond*. 176-225. Cambridge University Press. Cambridge.

British Council & Eaquals (European Association for Quality Language Services). 2015. *Core Inventory for General English.* http://englishagenda.britishcouncil.org/continuing-professional-development/cpd-teacher-trainers/british-council-eaquals-core-inventory-general-english (accessed 19 February 2020)

Broek, Simon, Inge van den Ende. 2013. *The implementation of the Common European Framework for Languages in European education systems*. Strassburg: European Union. https://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/495871/IPOL-CULT_ET%282013%29495871_EN.pdf (accessed 10 June 2020).

Brown, Howard. 2017. Current Trends in English-medium Instruction at Universities in Japan. *OnCue Journal* 10(1). 3-20. https://www.academia.edu/35562396/Current_Trends_in_English-medium_Instruction_at_Universities_in_Japan (accessed 19 February 2020)

Burns, Anne. 1999. *Collaborative Action Research for English Language Teachers.* Cambridge University Press. Cambridge.

Byram, Michael & Parmenter. Lynne (Eds). 2012. *The Common European Framework of Reference: The globalization of language education policy.* Bristol: Multilingual Matters.

Cook, Gary and Rutson-Griffiths, Yukari. 2018. Investigating First-Year Students' Perceptions and Knowledge of the CEFR at a Japanese University. *Hiroshima Bunkyo Women's University Journal* 53. 67-83.

Cook, Gary. 2019. Working in silos: A report on the coordination of course collaboration at a Japanese university. *The IAFOR International Conference on Education-Hawaii 2019 Official Conference Proceedings*. 1-15. https://papers.iafor.org/submission42228/ (accessed 12 February 2020)

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press. http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp (accessed 19 February 2020)

Council of Europe. 2018. *Common European Framework of Reference for Languages: learning, teaching, assessment . Companion volume with new descriptors.* Strasbourg, France: Council of Europe. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989 (accessed 15 October 2019)

Council of Europe. 2020. *European Language Portfolio*. Strasbourg, France: Council of Europe. https://www.coe.int/en/web/portfolio (accessed 4 July 2020).

English Profile: *Grammar Profile.* 2015. Cambridge: Cambridge University Press.   http://www.englishprofile.org/english-grammar-profile (accessed 19 October 2019)

European Association for Quality Language Services. 2015. *Eaquals Bank of Descriptors as Levels.* London, UK: Eaquals. https://www.eaquals.org/resources/revision-and-refinement-of-cefr-descriptors/ (accessed 19 February 2020)

Hamiuc, Monica & Jonathan Parker. 2016. Interdisciplinary Course Design for CLIL. *Comparative Culture, The Journal of Miyazaki International College* 21. 2-14. https://www.mic.ac.jp/files/uploads/2016.pdf (accessed 30 July 2020)

Kodate, Azusa. 2017. Developing ELP informed self-access centre learning materials to support a curriculum aligned with the CEFR. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.) (2017). *Critical, Constructive Assessment of CEFR-informed Language Teaching in Japan and Beyond*. 226-247. Cambridge: Cambridge University Press.

Little, David, Francis Goullier & Gareth Hughes. 2011. The European Language Portfolio: The story so far. https://www.coe.int/en/web/portfolio/the-european-language-portfolio-the-story-so-far-1991-2011-executive-summary (accessed 30 July 2020)

Mills, Goeffrey. 2006. *Action Research: A Guide for the Teacher Researcher (3ʳᵈ edition)* London: Prentice Hall.

Ministry of Education, Culture, Sports, Science and Technology-Japan (MEXT). (2019). *Gaikokugo (Eigo) Koakarikyuramu nit suite* [Foreign Language- The English Language Core Curriculum]. https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2019/04/04/1415122_3.pdf (accessed 19 June 2020).

Mork, Catherine Mette & Anne McLellan Howard. 2015. An Investigation into Active Learning at MIC: A Beginning and the Way Forward. *Comparative Culture, The Journal of Miyazaki International College* 20. 67-86.

Morrow, Keith. 2004. *Insights from the Common European Framework*. Oxford: Oxford University Press.

Nagai, Noriko & Fergus O'Dwyer. 2011. The actual and potential impacts of the CEFR on language education in Japan. *Synergies Europe* 6. 141–152.

Nation, Paul. 2007. The Four Strands. *Innovation in Language Learning and Teaching* 1(1). 2-13.

Nation, Paul, with Marcos Benevides, James Broadbridge & Joseph Siegal. 2018. *Reading for Speed and Fluency, 1 & 2. Second Edition.* Korea: Compass Publishing.

Negishi, Masashi, Tomoko Takada & Yukio Tono. 2013. A progress report on the development of the CEFR-J. *Exploring Language Frameworks: Proceedings of the ALTE Kraków Conference* In Evelina D. Galaczi & Cyril J. Weir (eds.). 135-163. Cambridge: Cambridge University Press.

North, Brian. 2010. Linking Certification to the CEFR: Do we need standard setting? In Mader, Judith & Zeynep Urkun (Eds). Putting the CEFR to good use. *IATEFL*. http://www.ealta.eu/documents/resources/IATEFL_EALTA_Proceedings_2010.pdf (accessed 18 February 2020).

Piccardo, Enrica, Brian North & Tim Goodier. 2019. Broadening the Scope of Language Education: Mediation, Plurilingualism, and Collaborative Learning: the CEFR Companion Volume, *Journal of e-Learning and Knowledge Society* 15 (2). 17-36.

Schmidt, Rebecca. 2018. Development and Evaluation of a CEFR-J Based Classroom Activity. *Hiroshima Bunkyo Women's University Journal* 52. 1-13.

Strong, Gregory. 2019. EAP as a Bridge to EMI: Learning from the UK. *The Language Teacher* 43(6). 3-9.

Swap, Robert. J. & Kent Wayland. 2013. Working Across Disciplines and Chipping Away at Silos with SLCE: An Interdisciplinary Approach to Educating Science and Engineering Students. *International Journal for Service Learning in Engineering Special Edition*, Fall. 120–136.

Szirmai, Monika. 2014. The Globalization of the CEFR Reconsidered in a Socio-Cultural Context. Retrieved from http://www.tufs.ac.jp/common/fs/ilr/ASIA_kaken/_userdata/23-38_Szirmai.pdf (accessed 31 January 2020).

Tannenbaum, Richard & Caroline Wylie. 2015. Mapping the TOEIC and TOEIC Bridge Tests on the Common European Framework of Reference for Languages. https://www.ets.org/s/toeic/pdf/toeic-cef-mapping-flyer.pdf (accessed 19 February 2020).

Applications referred to:

- Microsoft (n.d.). Flipgrid: Empower Every Voice https://info.flipgrid.com/
- Goldberg, Paul, Thomas Robb, Rob Waring, Willy Renandya & Charles Browne. 2020. X reading. https://xreading.com/
- Miles, Scott, Anders McCarthy & Michael Glass. 2008, 2020. Westbridge Education: Praxis Ed. http://praxised.com/
- Hock, Tanner, Romeo, Genivieve Romeo, Anne Beatty & Susan Kelly. 2020. Read Theory https://readtheory.org/

6 Biographies

Rebecca Schmidt is currently a Lecturer at Miyazaki International College. She previously worked 6 years with a team of instructors at Hiroshima Bunkyo University that aimed to align the entire university's language program with the CEFR through a three-stage project. Prior to that she taught 10 years in Japanese public schools, where she also helped establish a curriculum for elementary schools' English language programs. She has a Master's in Applied Linguistics from Macquarie University. Other research interests involve collaborative learning, classroom group dynamics, and ludic uses of language in teaching practice.

Ellen Head has worked in Japan since 2000. She has a strong interest in learner autonomy and assessment for learning. After becoming involved with the Learner Development SIG in 2001, she wrote her MA dissertation on Learner Autonomy in Language Education in Japan (2006), and worked with the Framework and Language Portfolios SIG on a chapter for Can do statements in Language Education in Japan and Beyond in 2009-10. She has also worked as a language researcher on the Cambridge International Dictionary of English (1998) and an assessment consultant for British Council China (2015-7). Currently she teaches at Miyazaki International College.

Appendix 1: MIC Functional Speaking Objectives Mapped to CEFR Levels

Functional Speaking Objective (as listed in the SILA Language Program Handbook)	Functions	Grammar
Oral Communication 1: Spring Semester (1st year)		
1. **Managing conversations** (*wh-questions / yes / no questions*)	**A1**- Giving Personal Information (family life, hobbies, leisure activities)	**A1:** Questions / To be / Verb + -ing
2. **Classroom English**	**B1**- Expressing Opinions / Expressing agreement and disagreement / Taking the Initiative in Interaction / Checking Understanding / Managing Interaction	**A2:** Modals - should

Functional Speaking Objective (as listed in the SILA Language Program Handbook)	Functions	Grammar
3. **Describing routines & habits** (*present simple tense*)	A1/ A2 - Describing habits and routines	A1 / A2: Adverbs of Frequency A2: Adverbial phrases of time A1: Present Simple Tense
4. **Describing visual images** (*present simple tense / present progressive tense / present perfect tense/*)	A2: Describing people, places, and things B1: Describing places	A1: There is / are / present simple tense / prepositions of place. A2: Present continuous B1: Present perfect
5. **Recounting past events** (*past simple tense*)	A2: Describing past experiences	A2: Past Simple / Wh- Questions in the past
6. **Giving & responding to advice** (*modals of advice*)	A2: Obligation & Necessity / Suggestions	A2: Modals - have to / should B1: Modals - must / have to
Oral Communication 2: Fall Semester (1st year)		
1. **Managing conversations** (*leading group discussions / bringing others into the discussion / offering alternative suggestions and opinions*)	B1 - Expressing Opinions / Expressing agreement and disagreement / Checking Understanding / Initiating and Closing Conversation / Managing Interaction/ **B2:** Expressing reaction / Interacting informally, reacting, expressing interest, sympathy, surprise, etc. / Taking initiative in interaction	B1 Negative questions / complex question tags
2. **Describing situations and circumstances** (*describing present situations / how they originated and how long they have lasted*)	B1: Describing places B2- Describing experiences	B1: present perfect tense / and past simple tense / present perfect progressive / used to
3. **Talking about the future** (*talking about planned future events and possibilities*)	B1 / B2- Describing experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.	A2: future simple tense / going to / present simple tense for the future / present progressive tense / first conditional - future possibilities B2: Will and going to for prediction

Functional Speaking Objective (as listed in the SILA Language Program Handbook)	Functions	Grammar
4. **Comparing and contrasting** (*compare and contrast situations and people/ Discuss changes from past to present*)	**A2:** Describing people, places, and things **B1:** Describing places	**A1:** *-er / more* with adjectives adverbs, nouns or verbs. **A2:** *as . . . as . . .* **B1:** superlative Present and past participial adjectives
5. **Expressing possibility, probability, and certainty** (*discuss present and past forms of probability*	**B2:** Expressing certainty, probability, doubt	**A2:** models-*could, can't* Adverbs of probability: *perhaps, possibly, maybe, probably* **B1:** : modals *must, may, might*
6. **Expressing rules and obligations** (*modals of advice*)	**A2:** Obligation & Necessity / Suggestions	**A2:** Modals - have to / should **B1:** Modals - must / have to / mustn't / don't have to
Oral Communication 3: Spring Semester (2nd year)		
1. **Managing conversations** (*actively engage with the ideas expressed in discussion/ justify opinions: giving examples, listing reasons, being more specific / clarify meaning*)	**B2:** Developing an argument / Encouraging and inviting another speaker to come in / Expressing agreement and disagreement / Expressing opinions/ Interacting and reacting / opinion – justification / Taking the initiative	No specified grammar targets
2. **Giving definitions** *of new words or ideas* (*adjective clauses*)	**A2:** Describing people, places, and things **B1:** Describing places	**B1:** *defining, object, with 'who/ that'* *Can use a defining relative clause with 'who' or 'that' as the object.*
3. **Telling stories** *with complex time frames* (*past simple; past progressive; past perfect*)	**B1:** Describing experiences and events / describing feelings and emotions / describing places	**A2:** simple past; past continuous- *describing background events and events in progress* **B2:** past continuous – *giving reason and explaining ongoing repeated events*
4. **Expressing the unreal past** *including expressing criticism and regret* (*past modals of advice: should have; ought to have / third conditional*)	**B1:** Describing experiences and events / describing feelings and emotions / **B2:** Critiquing and Reviewing / Describing experiences, feelings and emotions/ expressing abstract ideas	**B2/C1:** *models past: should have; might have* **B1:** if +past perfect + would have – to talk about imagined situations in the past, often with regret

Functional Speaking Objective (as listed in the SILA Language Program Handbook)	Functions	Grammar
5. **Making hypotheses or predictions** (*first conditional / second conditional*)	**B2:** Critiquing and Reviewing / Describing hopes and plans; expressing abstract ideas	**B1:** Modals of deduction and speculation; if + will to introduce a possible future condition; if + past simple + would to talk about an imagined situation **B2:** Modals of deduction and speculation; will and going to for prediction
6. **Reporting opinions & summarizing** (*reported speech*)	**B2:** Critiquing and Reviewing	**B2:** reported speech

Appendix 2: Reading 1 Can Do list items referenced to descriptors in CEFR/CV

Reading Can Do List (Reading 1 & 2)	Descriptors from the CEFR Companion Volume (CEFR/CV)
1. I can enjoy reading stories or non-fiction in English if the level of the book is right for me. (Reading as a leisure activity)	B2 Reading as a leisure activity. Can read for pleasure with a high degree of independence, adapting the style and speed of reading to different texts. (p.65) Reading for orientation B1 Can assess whether a book or article is on the required topic. (62)
2. I can read with a reading speed of about 200 words per minute.	B 2 Reading for Orientation: reading quickly. (62)
3. When I see a reading passage in English, I can predict what the article will be about by looking at the title and pictures. (Reading for orientation)	B1 Identifying cues and inferring. (67) Can make basic inferences about text content from headings, titles and headlines.
4. I can find the main idea in a reading passage and highlight it. (Reading for information and argument)	B1 Reading for information and argument. (63) Can identify main conclusions in a straightforward text.
5. I can scan quickly through a reading passage to find information which I need such as numbers or places. (Reading for orientation)	B1 Reading for orientation. (62)
6. I can use an English-English dictionary to help me understand what I read. (Reading for information; translation)	A2 Can locate specific information in lists. (62) (dictionary use is mentioned as an aid to understanding rather than as a goal)
7. I can use punctuation marks (, . ; : - "!" "?") to help me read aloud.	not mentioned

Reading Can Do List (Reading 1 & 2)	Descriptors from the CEFR Companion Volume (CEFR/CV)
8. I can re-write something in my own words after reading it (paraphrase). (Mediation)	B1 Relaying specific information in writing. (108)
	A2 to B2 Can exploit paraphrasing and simplification to make spoken and written texts more accessible. (112)
	B1 Adapting language. Can paraphrase the main points in straightforward written or spoken text. (128)
9. I can make notes on what I read. (Mediation)	A2 Processing text in writing. Can list main points as bullet points. (112)
10. I can explain what I have read to another person. (Mediation)	A2 to B2 Can exploit paraphrasing and simplification to make spoken and written texts more accessible. (112).
	B1 Adapting language. Can paraphrase the main points in straightforward written or spoken text. (128)
11. I can think about what I read and make inferences. (Mediation)	B1 Reception strategies. Identifying cues and inferring. (67)

Appendix 3: Academic Writing 1 Can Do List Correlation with CEFR/CV Descriptors

Academic Writing 1 Can Do List	Correlation with CEFR/CV Descriptors
Item	CEFR/CV
1. I can write a paragraph introducing myself and my hobbies and interests.	Overall Written Production A1 can give information in writing about matters of personal relevance. (75)
	Coherence and cohesion, using paragraphs to emphasize text structure. B1 Can link a series of shorter, discrete simple elements into a connected, linear sequence of points. (145)
2. I know how to word-process and layout and format a paragraph on the computer.	NOT in CEFR
3. I can write a paragraph describing a person or place, with a topic sentence and supporting details.	Overall Written production: B1 Can write straightforward connected text on a range of familiar subjects within his/her field of interest, by linking a series of shorter discrete elements into a linear sequence. (75)
4. I can write a story about something interesting which happened to me or something interesting which I heard.	Creative writing: B1 Can clearly signal chronological sequence in narrative text. Can narrate a story. (76)

Academic Writing 1 Can Do List	Correlation with CEFR/CV Descriptors
5. I can write a paragraph giving my opinion and using examples and reasons to support my opinion.	Written reports and essays: B1 Can write a text on a topical subject of personal interest, using simple language to list advantages and disadvantages and give his/her opinion. (from old CEFR) (77) Thematic development: B1 Can develop an argument well enough to be followed most of the time. (141) B2 Can develop a clear argument, expanding and supporting his/her point of view with relevant supporting detail and examples. (141)
6. I can use grammar accurately, for example, using the past tense for telling a story and using singular and plural verbs.	Not dealt with in this sense? Dealt with under spoken production? Grammatical accuracy: B1 Uses reasonably accurately a repertoire of frequently used routines and patterns associated with more predictable situations. (133)
7. I can plan a paragraph or essay using techniques like brainstorming, mind-mapping or free-writing.	Not dealt with specifically in relation to writing? Dealt with under spoken production? Production strategies: Planning B1. (78)
8. I can make my writing interesting by using a variety of different words, and expressions, not repeating the same word.	Communicative language competence: vocabulary range B2 Can vary formulation to avoid frequent repetition. (132)
9. I can edit my writing or my partner's writing to improve the grammar and content.	Not mentioned
Second semester: same list with additional points added in AW2	
10. I can write a five paragraph essay giving my opinion about a controversial topic such as global warming.	Written reports and essays B 2 Can write an essay or report that develops an argument systematically with appropriate highlighting of significant points and relevant supporting details. (77)
11. I can write an essay using evidence to persuade someone that my opinion is correct.	Covered under speaking Sustained monologue: putting a case (eg. Ina debate) B2 Can develop a clear argument, expanding and supporting his/her points of view at some length with subsidiary points and relevant examples. (72)
12. I can write an essay or report which describes a problem and suggests solutions to the problem.	Covered under Production It involves learning the expectations and conventions of the genre concerned. (68)
13. I can use language to explain the causes of an event (for example, Why did Sei Shonagon call her book "The Pillow Book"?)	As above, number 10.

Academic Writing 1 Can Do List	Correlation with CEFR/CV Descriptors
14. I can use language to explain the effects of an event (for example, what were the effects of the USA election in 2017?)	As above number 10.

Appendix 4: CLIL Classroom Practices Mapped to CEFR/CV

Reception Activities CEFR level	Illustrative Descriptor	Classroom practice
Understanding conversation between other speakers: **B1**	Can generally follow the main points of extended discussion around him/her, provided speech is clearly articulated in a familiar accent.	Students discuss the main points of the lecture in small groups.
Listening as a member of a live audience: **B1**	Can follow in outline straightforward short talks on familiar topics, delivered in clearly articulated standard speech.	Students follow the content instructor's lecture aided by a slideshow, which should be somewhat familiar after reading texts for homework.
Overall reading comprehension **B1**	Can read straightforward factual texts on subjects related to his or her field of interest with satisfactory levels of comprehension.	Field of interest = course topic (history)/reading should be adjusted at B1 level to accomplish 'satisfactory' comprehension.
Reading for orientation **B1**	Can scan longer texts in order to locate desired information from different parts of the text.	Vocabulary and reading comprehension tasks require students to do this.
Overall written Production **B1**	Can write straightforward connected texts on a range of familiar subjects	Students are expected to answer short essay questions on handouts and in assessments related to texts and lectures.
Overall spoken interaction Formal Discussion **B1**	Can exchange, check and confirm information. Can take part in routine formal discussion of familiar subjects	Each class, after listening to the lecture, students repeat what they understood, and check and confirm what they might be confused about.

Learner perspectives: familiarization, knowledge, and perceptions of the CEFR

Gary Cook, Hiroshima Bunkyo University

Yukari Rutson-Griffiths, Hiroshima Bunkyo University

Since 2012, various aspects of the Common European Framework of Reference for Languages (CEFR) have been utilized in language curricula at Hiroshima Bunkyo University's language learning center: the Bunkyo English Communication Center (BECC). Whilst teacher training regarding the CEFR had taken place, students had not received any direct CEFR-education. A decision was taken in 2017 to involve one cohort of students from the Global Communication Department in activities over three years to a) determine their ability to sort Can Do descriptors into their respective levels on the self-assessment grid (A1-C2), and b) establish their knowledge and perceptions of the CEFR through questionnaires. The sorting activity, termed the CEFR shuffle in this paper, was also intended to raise students' awareness of the CEFR. Results of this research comparing two groups of students streamed into low (A1-A2) and high (A2-B1) classes show that 1) their ability to correctly sort descriptors as a cohort showed no change over a period of 2 years, 2) the high class performed better than the low class in terms of making fewer sorting mistakes in the first and third years of the study but these results were not found to be statistically significant, and 3) students' knowledge of the CEFR improved slightly while perceptions received mixed results. Further results are discussed and suggestions made for the improvement of this study in future editions.

Keywords: Awareness, Self-assessment, Curriculum, Training, Knowledge, Perceptions, CEFR, Can Do descriptors

1 Introduction

The Bunkyo English Communication Center (BECC) was established in 2008, to provide English language education for students of Hiroshima Bunkyo Women's University University, currently known as Hiroshima Bunkyo University since the admission of male students in the 2019 academic year. Since 2012, staff at the BECC have been involved in major curriculum projects which have seen the employment of the Common European Framework of Reference for Languages (CEFR) (Council of Europe [CoE] 2001) as a guide for the creation of its English language courses. One example is a compulsory first-year English Communication course created in-house which has materials based on 13 language themes provided by the *Waystage* document (Ek and Trim 1991), CEFR-informed goals from the European Association for Quality Language Services (Eaquals) bank of descriptors (2015), and the Association of Language Testers in Europe (ALTE) Can Do project (2002), and assessments modeled on the CEFR aligned *Cambridge English: Key* (University of Cambridge ESOL Examinations 2012a) and *Cambridge English: Preliminary* (University of Cambridge ESOL Examinations 2012b) tests. For detailed information on this project see Bower, Runnels, Rutson-Griffiths, Schmidt, Cook, Lehde and Kodate (2017) and Bower, Rutson-Griffiths, Cook, Schmidt, Lehde, Kodate and Runnels (2017). The project required a great deal of collaboration amongst staff, including the creation and adaptation of a plan by management, organization of staff into various committees to create the curriculum, and ultimately numerous rounds of feedback from teachers as a result of the implementation of lessons. This collaborative effort inspired further projects to coordinate other BECC language courses with CEFR-informed goals and assessments (Cook 2019).

There were various motivations behind this study. Learner autonomy is synonymous with the CEFR. Indeed, one of the main functions of the CEFR-linked European Language Portfolio (Council of Europe 2000) is to support the development of learner autonomy. Due to one of the authors being a learning advisor in the BECC's Self-Access Language Center (SALC), one of the motivating factors was to empower students with knowledge regarding their own language learning. The SALC is an English-only space which primarily exists to encourage students to develop their autonomous language learning skills. Goal setting and reflection are two such skills that are fostered through advising sessions with learning advisors and in-house materials including SALC activities, which are a core component of the aforementioned first-year English communication course. Students select these activities based on what they believe their CEFR-levels to be in the skills of reading, listening, speaking or writing. Pre- and post Can Do descriptors within activities help guide these choices (for more information on the SALC activities, see Kodate 2017). Can Do descriptors are also utilized as individual lesson and overall curriculum goals in a variety of the BECC's language courses. Therefore, we thought it beneficial for students to engage them in a Can Do descriptor familiarization task followed by goal setting and reflection, based on interacting with the CEFR self-assessment grid at all levels.

Furthermore, with BECC staff having undergone various training regarding the CEFR, such as workshops on the history of the use of the CEFR in Japan, how to write Can Do descriptors, and more recently the introduction of the CEFR Companion Volume, familiarizing students considering their role as important stakeholders was a practical step in the continued implementation of the CEFR at the BECC. As stated above, the BECC provides its students with English language education based on the CEFR. Students are presented with CEFR levels and Can Do descriptors in class and the SALC, and are given chances to reflect on their language proficiency before and after each task and assessment. However, those tasks and assessments often give a reference to their performances within a limited range, and the students do not receive many opportunities to engage with Can Do descriptors at all levels. Having stated that one of the intended uses of the CEFR is to support self-directed learning including "raising the learner's awareness of his or her present state of knowledge" and "self-assessment" (Council of Europe [CoE] 2001: 6), it is argued "the main potential for self-assessment . . . is in its use as a tool for motivation and awareness raising: helping learners to appreciate their strengths, recognise their weakness and orient their learning more effectively" (CoE 2001: 192).

As it is stated that the accuracy of self-assessment is enhanced if some training is given to the learners (CoE 2001), the study had first-year students engage in a sorting activity with Can Do descriptors from the CEFR self-assessment grid, termed the CEFR shuffle in this paper. The activity had students in groups read and closely study the CEFR Can Do descriptors from all levels and place them in order from the lowest level, A1, to the highest level, C2, by analyzing what qualities or components differentiate the Can Do descriptors. This was repeated in students' second and third years of study.

Another motivation factor for this research came from the growing popularity of the CEFR in Japan. This is evident from the development of a Japanese version of the CEFR; the CEFR-J (see Tono and Negishi 2012), in addition to the Ministry of Education, Culture, Sports, Science and Technology's (MEXT) decision in 2013 to propose CEFR-referenced goals for all junior and senior high schools (MEXT 2013). Most of the students in this study were attending high schools in Japan from the 2014 academic year, so we were curious as to what CEFR-knowledge our students had obtained before admission to the university in 2017, and how that knowledge may have changed in their first three years of education at the BECC.

The integration of the CEFR in Japan has been well documented (Nagai and O'Dwyer 2011; O'Dwyer 2015; Fennelly 2016; Schmidt et al. 2017); however, there is a lack of research that has a focus on what Japanese learners of English know about the CEFR. Therefore, this article has a dual purpose: firstly, to show the results of our attempt to shine light on the extent of our learners' knowledge of the CEFR, and secondly, to publish research with a hope that other educators may follow suit and carry out similar studies with a focus on the language learner and the CEFR in practice. In addition to the CEFR shuffle,

students were required to answer a questionnaire to establish the extent of students' knowledge and perceptions of the CEFR. The questionnaire was only conducted in the first and third years of the study. The following three research questions were investigated:

1. How well can students sort Can Do descriptors into their respective levels on the self-assessment grid?
2. What effect does language ability have on the Can Do descriptors sorting exercise?
3. How has students' knowledge and perceptions of the CEFR changed in three years of English language education?

2 Methodology

2.1 Participants

This study was conducted from the academic year of 2017 until the end of the 2019 academic year among 31 students enrolled in the Global Communication Department at Hiroshima Bunkyo University. The students received compulsory English lessons every year: six lessons per week in the first year, four lessons per week in the second year, and one lesson per week in the third year, with extra elective classes also available in the third year. Each lesson was 90 minutes. The students' TOEIC scores varied from 180 to 515 in the first year of the study. At the beginning of each year, all the students were streamed into two classes. The overall TOEIC score averages across the three years of the study are shown in Table 1.

Table 1. *Overall TOEIC score averages across three years*

	April, 2017	July, 2018	July, 2019
High-streamed	370.4	477.1	574.1
Low-streamed	282.9	379.3	398.8

It should be noted here that there were a few changes to the number of participants across the three years due to absences on the days of the study and leaves for long-term study abroad programmes. In addition, three students initially placed in the high-streamed class in 2017 exchanged places with three students from the low stream in the third year of this study.

2.2 CEFR shuffle

The Manual provided by the Council of Europe for language educators who wish to incorporate the CEFR into examinations suggests that they prepare for the implementation through some familiarization tasks (Council of Europe 2009). One of the activities proposed in the document entails reconstructing the CEFR self-assessment grid. For this activity, the participants place the provided Can Do descriptors in the correct empty cells of the CEFR table. This study adapted a similar task as a learner training exercise. Every year, the students in each class were separated into groups of three or four and were given a set of the CEFR self-assessment grid table and 30 Can Do descriptor cards: five skills and six levels from A1 to C2. After receiving a brief explanation on what the CEFR is and how it is utilized in the BECC, they were asked to place the Can Do descriptors onto the correct places of the table. The descriptors were color-coded according to the skills, so students only needed to determine the levels of the descriptors. The descriptors were written in Japanese, and students collaborated to complete the task speaking Japanese. After checking the answers with the researchers, photos were taken of the results, and the students were asked to map three TOEIC scores (225, 550, and 785) and three Eiken grades (pre-2, 2, and pre-1) onto equivalent CEFR levels. The TOEIC score average of each group is shown in Tables 2 and 3[1].

---

1. It is noted here that although the groups in the first and second years consisted of the same members, new groups were formed in the third year due to the exchange of students in the high- and low-streamed classes.

Table 2. *TOEIC score average of each CEFR shuffle group in 2017 and 2018*

	Groups	Number of students	April, 2017	July, 2018
High	1	4	347.5	400.0
	2	3	401.7	493.3
	3	3	398.3	563.3
	4	4	348.8	477.5
Low	5	4	300.0	373.8
	6	4	231.7	355.0
	7	4	315.0	356.3
	8	3	268.3	441.7

Table 3. *TOEIC score average of each CEFR shuffle group in 2019*

	Groups	Number of Students	July, 2019
High	1	4	536.3
	2	4	633.8
	3	4	587.5
	4	4	538.8
Low	5	3	357.5
	6	3	425.0
	7	4	406.3
	8	4	392.5

After the CEFR Shuffle in the first and third years, students were given time to reflect on their English learning and set learning goals based on the CEFR self-assessment grid.

2.3 Survey

Approximately three months after the CEFR shuffle in 2017 and 2019, the participants were asked to take a short survey to find out their knowledge and perceptions of the CEFR (see Appendix). In the following section, the results of the survey from 2019 will be compared with those of 2017, reported in Cook and Rutson-Griffiths (2018).

3 Results

3.1 CEFR shuffle results

The following results are reported from one cohort of students across three years (2017-2019). The low-streamed class results can be seen in Table 4. Students made 128 mistakes in total across three consecutive years with 90 mistakes coming from the upper half of the self-assessment grid (B2 to C2), contrasting with 38 from the lower half (A1 to B1). Students made mistakes in all five skills with reading receiving the highest number of mistakes (32), and spoken production the fewest (14). The level that was least accurately categorized was C1, which drew 29% of total mistakes. This can be seen in contrast with A1 which only drew 8% of total mistakes.

Table 4. *CEFR self-assessment grid mistakes of lower-streamed class for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	1	1	1	8	7	5	23
Reading	1	3	3	3	11	11	32
Spoken Interaction	2	3	2	9	8	5	29
Spoken Production	1	1	0	0	6	6	14
Writing	5	7	7	5	5	1	30
Total	10	15	13	25	37	28	128

   Whilst the results from the high-streamed class in Table 5 show students made 34 fewer total mistakes than the low-streamed class, similar patterns can be seen. The upper-level descriptors (B2-C2) drew more mistakes, 67 in total, whereas 27 mistakes occurred with the lower-level descriptors (A1-B1). All skills received mistakes with reading once again topping the count (25 mistakes), and spoken production incurring the fewest mistakes (12). The C1 level was again categorized least accurately with 31% of total mistakes, and the A1 level the most accurate with only 4% of the total mistakes.

Table 5. *CEFR self-assessment grid mistakes of higher-streamed class for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	0	1	1	5	7	3	17
Reading	2	4	1	4	8	6	25
Spoken Interaction	0	1	2	7	7	5	22
Spoken Production	0	0	1	1	5	5	12
Writing	2	6	6	2	2	0	18
Total	4	12	11	19	29	19	94

   In Table 6 the low- and high-streamed class results are combined to show how this cohort fared with classification errors in each category across the three years of this study. As noted previously with Tables 4 and 5, the upper-level descriptors (B2-C2) were more problematic for students to categorize than the lower-level descriptors (A1-B1). 71% of all mistakes came from the upper-level of the self-assessment grid. In particular, the C1 level proved most challenging to categorize correctly with 30% of all mistakes, followed by its neighboring grids, C2 (21%) and B2 (20%). The A1 level was the most accurately categorized for this cohort of students with only 6% of total mistakes coming from the lowest level of the CEFR self-assessment grid. In terms of mistakes according to the five separate skills, reading incurred most mistakes (26%), in contrast with spoken production incurring the fewest mistakes (12%). Indeed, the spoken production skill only had one mistake per level from A1 to B2. There were only two other Can Do descriptors that received just one mistake across three years, with the A1 listening and C2 writing descriptors being the most accurately categorized, which are interestingly at opposite ends of the scale of levels.

Table 6. *CEFR self-assessment grid mistakes of combined high- and low-streamed classes for 2017-19*

	A1	A2	B1	B2	C1	C2	Total
Listening	1	2	2	13	14	8	40
Reading	3	7	4	7	19	17	57
Spoken Interaction	2	4	4	16	15	10	51
Spoken Production	1	1	1	1	11	11	26
Writing	7	13	13	7	7	1	48
Total	14	27	24	44	66	47	222

To determine how each class performed on a yearly basis, we can see in Figure 1 that the low-streamed class made 41 mistakes on their first attempt at sorting Can Do descriptors on the self-assessment grid. They improved slightly in 2018 by collectively making 2 fewer mistakes, however, in 2019 their overall performance deteriorated to make a total of 48 mistakes, which was worse than their initial attempt in 2017. Conversely, the high-streamed class started in 2017 with 31 mistakes, 10 fewer than the low stream, but then matched the low-streamed class in 2018 with 39 mistakes, and finally improved with their best result in 2019 of 24 mistakes, exactly half the number of mistakes of their low-streamed counterparts. A point of interest here is the corresponding average TOEIC scores per class for each year. In 2017, the low-streamed average TOEIC score was 282, versus the high-streamed class' average of 370. Both classes improved their average by approximately 100 points in 2018, however, in 2019 whereas the high stream continued to improve their average by approximately a further 100 points to 574, the low-streamed class improved only slightly by 20 points, to an average of 399 points. The role of language ability in sorting Can Do descriptors has piqued the interest of the authors, especially due to the fact that this CEFR shuffle task was completed in the students' native language of Japanese.
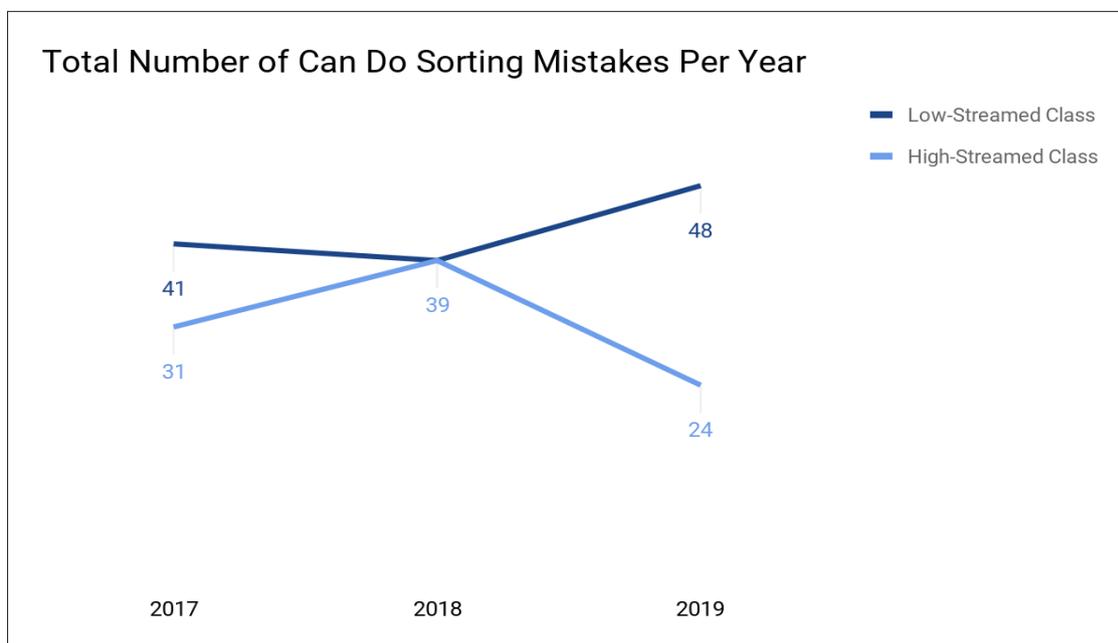
Figure 1. Can Do descriptor sorting mistakes according to class and year.

In order to examine the correlation between language proficiency and ability to correctly sort Can Do descriptors, each group's TOEIC average was compared with the number of mistakes they made each year, making 24 cases in total: four low-streamed groups and four high-streamed groups across

three years. According to the regression analysis on this data set, there was no statistically significant correlation between the students' TOEIC scores and the number of sorting mistakes (r^2=.11, p=.11).

3.2 Survey results

As can be seen in Figure 2, in 2019, when asked whether they can assess their language proficiency based on the CEFR scale, just over 30% of the students reported that they can, which showed some improvement from approximately 13% in the first year of the study. Also, no students reported that they couldn't remember the scales this time, while about 13% of the participants in the first year couldn't remember them.
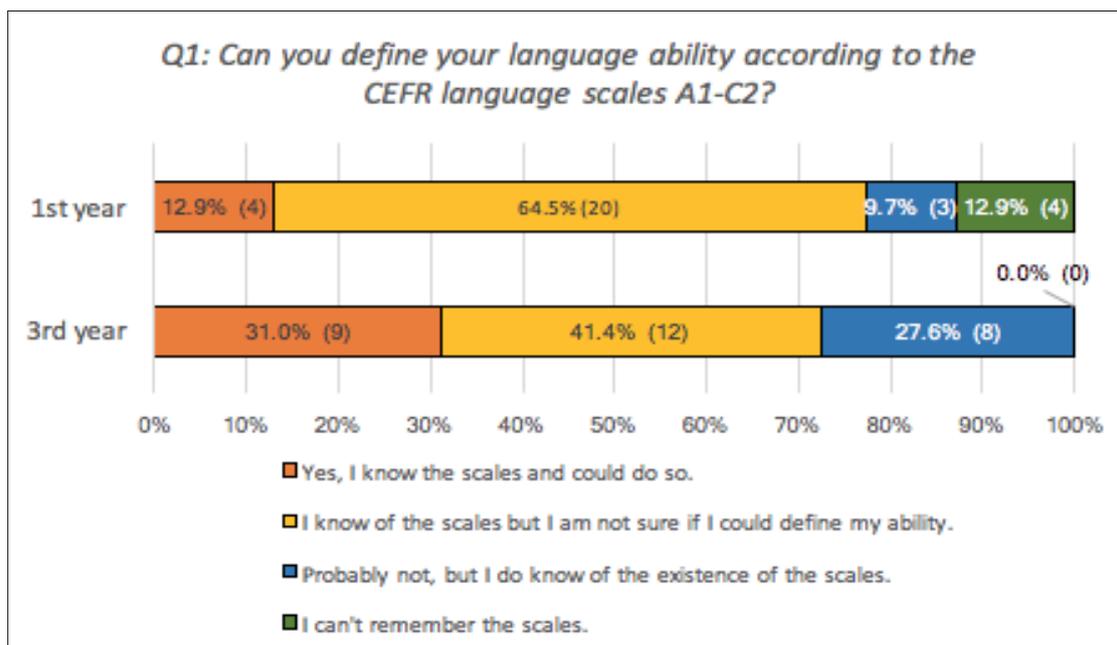
Figure 2. Students' ability in self-assessing their CEFR levels.

In 2019, more participants reported that they think the CEFR levels are useful to some degree. Although fewer participants chose "absolutely," when the two positive answer choices "absolutely" and "somewhat" are combined, the percentage of the participants who recognized the usefulness of the CEFR levels went up to 93.1% in the third year from 74.2% in the first year (Figure 3). In a similar vein, more students felt that Can Do descriptors are useful for them in the third year when comparing the two positive responses "absolutely" and "somewhat" in 2017 and 2019 (61.3% and 72.4% respectively). 22.6% and 38.7% of students did not know whether the CEFR levels and Can Do descriptors were useful for them in the first year, and these percentages came down to 3.5% and 24.1% of students in the third year.
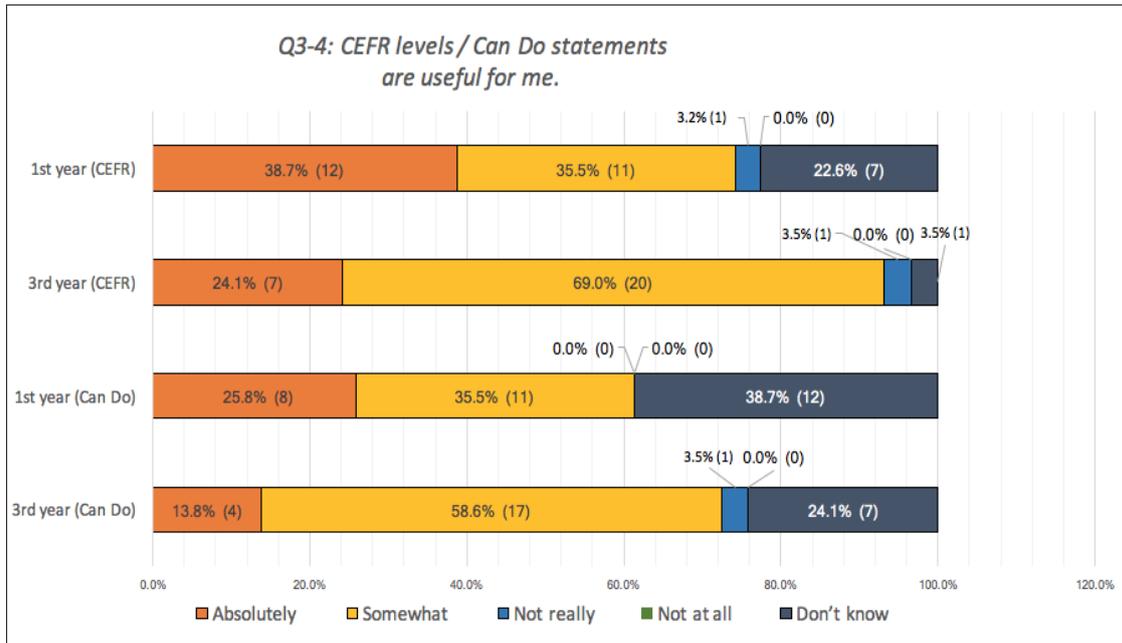
Figure 3. Usefulness of CEFR levels and Can Do descriptors.

When asked about the usefulness of the CEFR shuffle, Figure 4[2] shows that 82.7% of students reported positively in the first year of this study, and 75.9% in the third year. Three students in the first year said it was useful because it helped them set goals in their language learning. Eight students in the first year and seven students in the third year said it helped them better understand the CEFR levels. However, the percentage of the positive responses reduced from 82.7% to 75.9%, with three participants reporting that it was not really useful to them in 2019. Two of these three students' reasons were related to forgetting the descriptors soon after the activity. The other student did not give a specific reason for their choice.
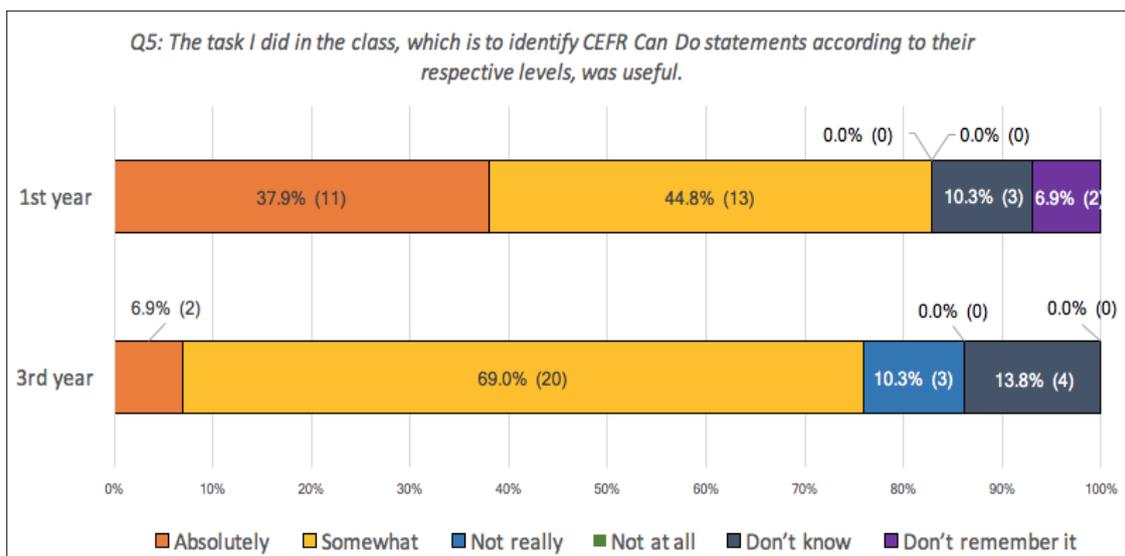
Figure 4. Usefulness of the CEFR shuffle.

---

2. It is noted here that two students who were absent on the CEFR shuffle day in the first year are excluded from the first year data.

Figure 5 shows the students' preference among three references to English proficiency. TOEIC was the most popular answer, followed by CEFR levels and Eiken grades in both the first and third years of the study. Slightly more students reported that CEFR levels are important to them in the third year while fewer students chose TOEIC. The Eiken grade received a similar score in 2019 compared to 2017.
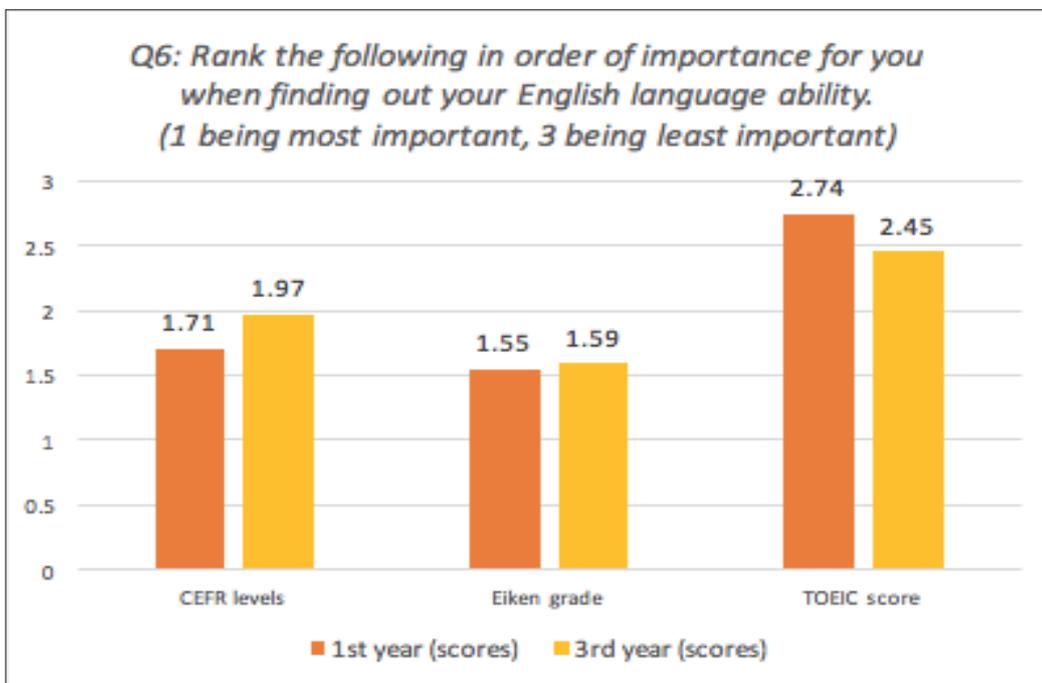
Figure 5. Importance of three language measurements.

Question 7 investigated how familiar the participants are with CEFR levels, TOEIC scores, and Eiken grades. Compared to the first year, the participants' knowledge of these language proficiency levels increased in the third year; nearly 80% of the students were able to place all the items in the correct order while 58.1% of them were able to do so in their first year (Figure 6).
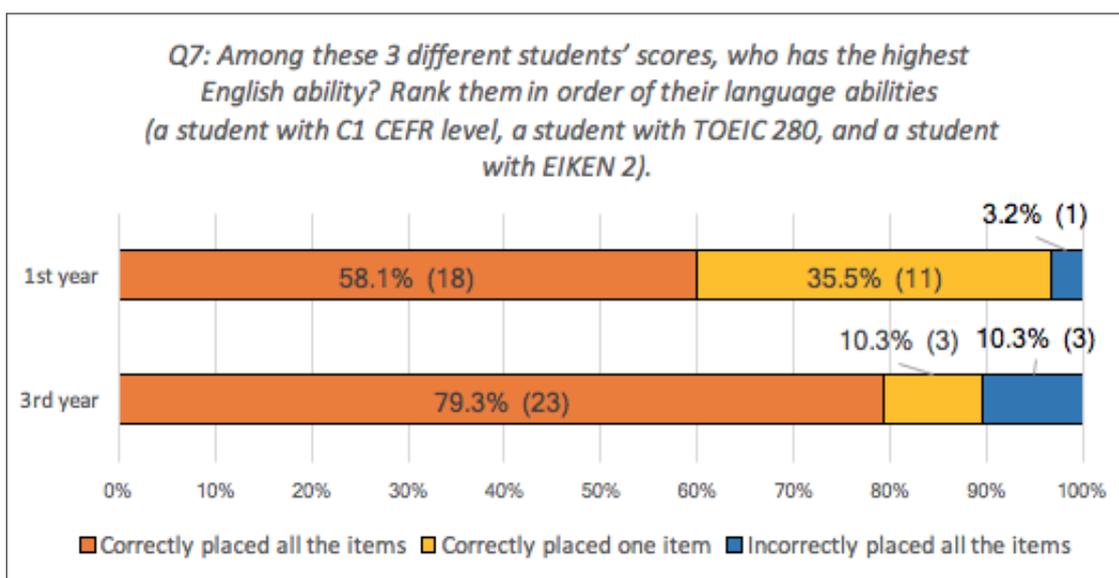
Figure 6. Students' knowledge of ranking CEFR, TOEIC and Eiken language ability.

For Questions 9 to 13, students were asked to self-assess their five language skills based on the CEFR scale (A1-C2). Displayed in Table 7 are the results of reading and listening, along with whether their self-assessment CEFR levels are supported by the results of an external English test conducted by an international education company which gives CEFR feedback: EF Education First. The results of this online test only give students an indication of reading and listening levels, hence we have not included spoken production, spoken interaction, and writing here. In terms of accuracy, more than half of the students who self-assessed their reading were correct in year 1 (58.3%), versus only 15% of students being correct in year 3. Conversely for listening, more students were correctly able to self-assess their levels in year 3 (42.1%) when compared with year 1 (16.7%). Students who responded that they could self-assess their reading and/or listening skills, but did not receive test results from EF Education First, were excluded from this table.

Table 7. *Students' accuracy in self-assessing their CEFR reading and listening levels*

	Reading		Listening	
	1st year	3rd year	1st year	3rd year
Students who assessed correctly	7	3	2	8
Students who assessed incorrectly	5	17	10	11
Total number of students	12	20	12	19

Questions 14 to 16 asked to what extent knowing their own language levels based on the three references: Eiken, CEFR, and TOEIC was useful when they graduate university. It is observable from these questions that more students appreciated the importance of these language assessments in the first year compared to the third year. Although TOEIC received the most positive responses among the three items in both years, fewer students were certain about its usefulness in the third year. Four students (approximately 14%) each reported in the third year that neither the CEFR nor Eiken results were useful for their future career or studies.
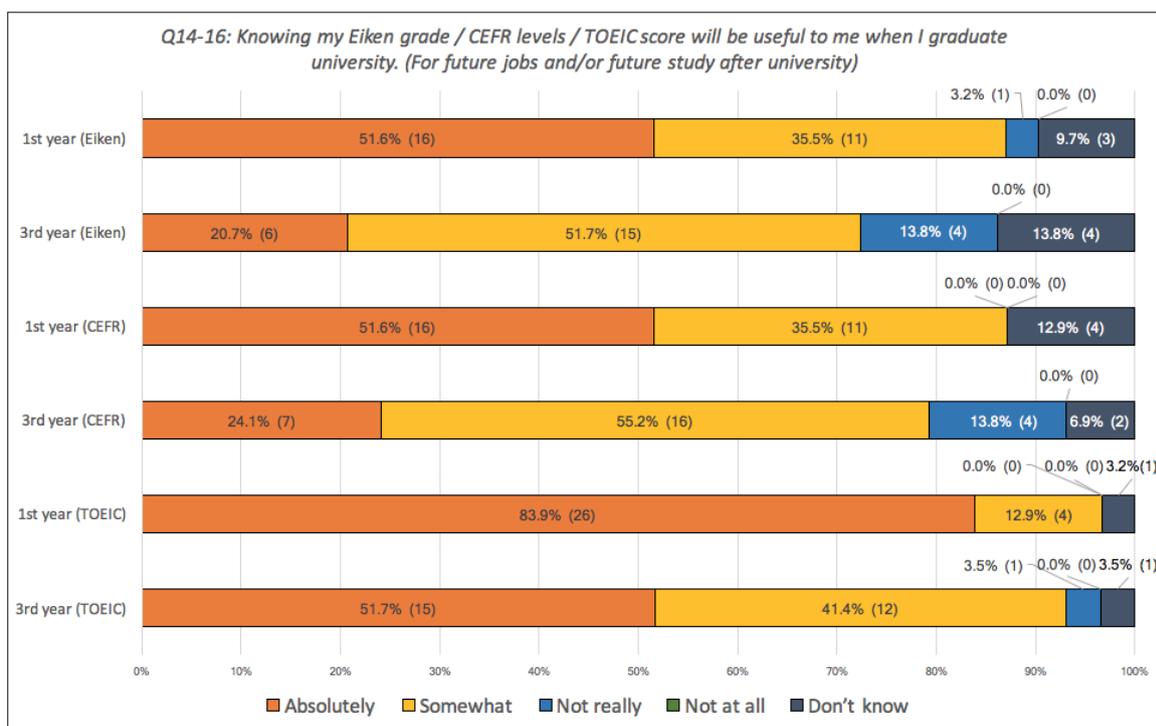
Figure 7. Popularity of three language measurements for future career and studies.

4 Discussion

The purpose of this research was to investigate how well students could sort Can Do descriptors into their respective levels of the self-assessment grid, examine what effect language ability may have had on the Can Do descriptor sorting exercise, and gain an understanding as to how students' knowledge and perceptions of the CEFR changed in their first three years of English language education at the BECC. The results from three consecutive years of conducting the CEFR shuffle show that as a cohort, students made no improvement in being able to correctly place Can Do descriptors onto the self-assessment grid. In fact, the total number of mistakes in years 1 and 3 were exactly the same. The results gave an indication that the levels of the CEFR had some effect as to where mistakes were more likely to occur. In particular, the upper half of the grid, levels B2 to C2, were most often incorrectly categorized with C1 consistently topping the mistake count each year. We suggest two reasons for this occurrence. First, the average level of our students in their third year of study is on the border of A2/B1, which is below where most mistakes occurred. Students may have had more problems conceptualising tasks above their level of English compared to their current or previously attained levels, hence confusion with which descriptor to place in B2, C1 or C2. However, we note here that this was not the case for all skills at these higher levels, with the exception being writing which had more mistakes in the A2-B1 range than the B2-C2 range. In fact, the C2 level writing descriptor was one of the most accurately placed descriptors with only one mistake in three years. The second reason we offer as to why our students were more consistent with accurately placing lower-level descriptors over higher ones is in the wording of the descriptors and abstractness of tasks. Lower-level descriptors often describe more concrete tasks using simple words, such as understanding familiar words on posters (reading A1), whereas higher-level descriptors utilize more difficult vocabulary, for example "contemporary literary prose" (reading B2), and abstract ideas: "appreciating distinctions of style" (reading C1) (CoE 2001: 27).

At first, language ability appeared to have some effect on students' ability to sort descriptors accurately. In both years 1 and 3 the high-streamed class outperformed the low stream, with year three being particularly of interest due not only to the high stream making the fewest mistakes of the study (24), but the result from the low stream making twice as many mistakes (48) as their counterparts. When comparing each group's average TOEIC scores, it is apparent the high-streamed class continued to improve year-by-year, with an average gain of 100 points per year, whereas the low group only managed to improve their third year TOEIC average by 20 points. However, as discussed earlier, the results of a regression analysis on the correlation between each groups' TOEIC average scores and the number of mistakes showed that there is no statistically significant correlation between those two factors. These mixed results may warrant a further study as students worked in groups and individual opinions may not have been reflected in the results. One possible change to the CEFR shuffle in the future is to add a step after the initial group discussion whereby students are individually asked to determine where Can Do descriptors are placed and examine the correlation between the results and their individual language proficiency.

The post-activity surveys conducted in the first and third years of the study suggest that the students' knowledge of the CEFR improved to some extent. When asked to define their CEFR levels, more students reported that they were able to do so (Question 1). However, when accuracy is considered with the results of an externally validated English test, students were not very successful or consistent at estimating their CEFR levels by skill. Reading levels were self-assessed more accurately in Year 1 (58.3%) than Year 3 (15%). Conversely, listening levels were self-assessed more accurately in Year 3 (42.1%) than Year 1 (16.7%) (Table 7). We consider the result that more students gained confidence to self-assess their language ability by skills in Year 3 a positive, however, and are currently trying to determine ways of improving accuracy of self-assessment for future students. The results of Question 7 showed that more students were aware of the CEFR levels in relation to the two major language proficiency tests in Japan: TOEIC and Eiken. As stated earlier, CEFR levels and Can Do descriptors are used in various forms in the BECC, from lesson materials and self-access resources to assessment and feedback. It is natural

to assume that the students had many chances to be exposed to the CEFR's level system in the last three years. Cook and Rutson-Griffiths (2018: 78) argue "students should become more aware of their CEFR levels as teachers give feedback on assignments utilizing CEFR-informed rubrics." This is probably the main reason for the improvement of their CEFR-knowledge. Although it is difficult to argue with certainty as there is no supporting data, it is hoped that our annual CEFR shuffle also contributed to this positive change.

Regarding the students' perceptions, the results of Questions 3 and 4 show the number of students who find CEFR levels and Can Do descriptors useful to some degree increased in two years. Additionally, Question 5 saw the majority of students agreeing that the CEFR shuffle task was useful to them with reasons being attributed to its contribution towards goal-setting and understanding of the CEFR levels. We believe the results of these three questions align favourably with a promotion of self-directed learning as was originally intended with the foundation of the CEFR (CoE 2001: 6), and one of motivating factors behind this study. However, this was not the case for all students. There were still several students who did not recognize the usefulness of CEFR levels or Can Do descriptors in the third year. This combined with two students' comments that the CEFR shuffle was not really useful to them due to the descriptors being quickly forgotten highlight the need for continued training to promote a more effective learning experience.

According to the results of Question 6, which was to find out the popularity of three major language proficiency measurements in Japan, TOEIC was the most popular, although its popularity slightly declined, and in turn, the CEFR attracted a little more favor in the third year. Our interpretation of this result is students' knowledge of the CEFR increased as they saw more examples of its use in the BECC over two years. The outcome was a better awareness of how the CEFR may be beneficial for their futures. A possible reason for the minimal change in the language measurement popularity is that these students are required to take TOEIC by their department at regular intervals, and it is still the most widely used language test for job hunting in Japan. It is interesting, however, to see that the popularity for all three language measurements declined in the third year (Figure 7). This could be the result of some of the students realizing or deciding that they would not need English in their future career or studies, leading to the somewhat negative perception towards these measurements of language ability.

5 Conclusion

We carried out this study to investigate students' familiarity, knowledge, and perceptions towards the CEFR and to raise their awareness of this framework upon which their English language education is based. Through our observation we discovered that students did not show improvement sorting Can Do descriptors over two years and that most mistakes were consistently made in the upper-level of the self-assessment grid, from B2 to C2 levels. However, this study is not without its limitations. Although the high-streamed group made half the number of mistakes as the low-streamed group in the third year of this study, the results of analysis on the correlation between students' language proficiency and descriptor sorting ability suggest there is no statistically significant correlation between those two factors. We suggest a further study that would have students engaging in the CEFR shuffle in small groups but making their final decisions individually to investigate the role between language proficiency and descriptor sorting ability. Feedback is one area that can be improved on in future editions. There was little discussion as to why Can Do descriptors were incorrectly placed. With better feedback, students may increase their accuracy of the placement of descriptors, which could contribute to improved results with students' ability to self-assess their language levels. Further limitations to this study were the small number of participants and the use of a single language test as a reference to compare the participants' task performances. A possible change to future research may be involving a larger sample size and employing a different language proficiency measurement. The contrasting results of students' CEFR knowledge improvement and a small decline in their interest in language proficiency references may

suggest that there are other factors to consider such as students' motivation and needs for English studies for their future. Nevertheless, this three-year study helped us to better understand our students' prior knowledge as well as initial perceptions of the CEFR and how these may change over a few years. The CEFR shuffle was viewed positively by most of our students, and although we hope this study also served to contribute towards self-directed learning we suggest continuous education about the CEFR and training on self-assessment.

6 References

Association of Language Testers in Europe. 2002. *The ALTE Can Do project: Articles and Can Do statements produced by the members of ALTE 1992-2002.* https://www.alte.org/Materials (accessed 19 February 2020)

Bower, Jack, Judith Runnels, Arthur Rutson-Griffiths, Rebecca Schmidt, Gary Cook, Lyndon Lusk Lehde & Azusa Kodate. 2017. Aligning a Japanese university's English language curriculum and lesson plans to the CEFR-J. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 176-225. Cambridge: Cambridge University Press.

Bower, Jack, Arthur Rutson-Griffiths, Gary Cook, Rebecca Schmidt, Lyndon Lusk Lehde, Azusa Kodate & Judith Runnels. 2017. The key questions in Bunkyo. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 247-266. Cambridge: Cambridge University Press.

Cook, Gary. 2019. Working in silos: A report on the coordination of course collaboration at a Japanese university. *The IAFOR International Conference on Education-Hawaii 2019 Official Conference Proceedings*. https://papers.iafor.org/submission42228/ (accessed 12 February 2020)

Cook, Gary & Yukari Rutson-Griffiths. 2018. Investigating first-year students' perceptions and knowledge of the CEFR at a Japanese university. *Hiroshima Bunkyo Joshi Daigaku Kiyo* [Bulletin of Hiroshima Bunkyo Women's University] 53. 67-83.

Council of Europe. 2000. *European Language Portfolio (ELP): Principles and guidelines.* Strasbourg: Language Policy Division. https://rm.coe.int/16804586ba (accessed 19 February 2020)

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. 2009. *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): A manual.* Strasbourg: Language Policy Division. https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr (accessed 19 February 2020)

Ek, Jan Ate van & John Leslie Melville Trim. 1991. *Waystage 1990*. Cambridge: Cambridge University Press.

European Association for Quality Language Services. 2015. *Eaquals bank of descriptors.* https://www.eaquals.org/resources/revision-and-refinement-of-cefr-descriptors/ (accessed 19 February 2020)

Fennelly, Mark Graham. 2016. The influence of CEFR on English language education in Japan. *Shikoku Daigaku Kiyo* [Bulletin of Shikoku University] 46. 109-122. http://id.nii.ac.jp/1456/00000378/ (accessed 19 February 2020)

Kodate, Azusa. 2017. Developing ELP-informed self-access centre learning materials to support a curriculum aligned to the CEFR. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 226-247. Cambridge: Cambridge University Press.

Ministry of Education, Culture, Sports, Science and Technology. 2013. *Gurobaru ka ni taio shita eigo kyoiku kaikaku jisshi keikaku* [English education reform plan corresponding to globalization]. https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1343704.htm (accessed 19 February 2020)

Nagai, Noriko & Fergus O'Dwyer. 2011. The actual and potential impacts of the CEFR on language education in Japan. *Synergies Europe* 6. 141-152. https://gerflint.fr/Base/Europe6/noriko.pdf (accessed 19 February 2020)

O'Dwyer, Fergus. 2015. Toward critical, constructive assessments of CEFR-based language teaching in Japan and beyond. *Gengo bunka kenkyu* [Studies in Language and Culture] 41. 191-204. https://ir.library.osaka-u.ac.jp/repo/ouka/all/51427/slc_41-191.pdf (accessed 19 February 2020)

Schmidt, Maria Gabriela, Judith Runnels & Noriko Nagai. 2017. The past, present and future of the CEFR in Japan. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English Profile Studies 6), 18-48. Cambridge: Cambridge University Press.

Tono, Yukio & Masashi Negishi. 2012. The CEFR-J: Adapting the CEFR for English language teaching in Japan. *JALT Framework & Language Portfolio SIG Newsletter.* 5-12.

University of Cambridge ESOL Examinations. 2012a. *Cambridge English Key: Key English Test (KET) CEFR level A2 handbook for teachers*. Cambridge: UCLES.

University of Cambridge ESOL Examinations. 2012b. *Cambridge English Preliminary: Preliminary English Test (PET) CEFR level B1 handbook for teachers*. Cambridge: UCLES.

7 Biographies

Gary Cook has been a lecturer at the Bunkyo English Communication Center in Hiroshima Bunkyo University since 2011, and is currently the Global Communication Department coordinator. He has previous teaching experience in Spain, France, England, Georgia, and his native New Zealand. His research interests are in the areas of curriculum development, analysing the needs of language learners, and the CEFR.

Yukari Rutson-Griffiths has been a learning advisor at Hiroshima Bunkyo University since 2015 and is the coordinator at the self-access learning center at the university. She has also been involved in creating self-access learning materials using the CEFR and CEFR-J. Her research interests include self-access learning, learner autonomy, and the CEFR.

Appendix

Survey regarding students' knowledge and perceptions of the CEFR

Q1: Can you define your language ability according to the CEFR language scales A1-C2?

- *Yes, I know the scales and could do so.*
- *I know of the scales but I am not sure if I could define my ability.*
- *Probably not, but I do know of the existence of the scales.*
- *I can't remember the scales.*

Q2: Where have you seen Can Do statements and/or CEFR levels? Write as many ideas as you can.

Q3: CEFR levels are useful for me.

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q4: Can Do statements are useful for me.

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q5: The task I did in the class, which is to identify CEFR Can Do statements according to their respective levels, was useful.

- *Absolutely / Somewhat / Not really / Not at all / Don't know / Don't remember it*

Q6: Rank the following in order of importance for you when finding out your English language ability. (1 being most important, 3 being least important)

- *CEFR levels / Eiken grade / TOEIC score*

Q7: Among these 3 different students' scores, who has the highest English ability? Rank them in order of their language abilities.

- *A student with C1 CEFR level / A student with TOEIC 280 / A student with Eiken 2*

Q8: What is your most recent TOEIC score? If you can't remember it please write "I don't remember."

Q9: What is your reading CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q10: What is your listening CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q11: What is your writing CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q12: What is your spoken interaction CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q13: What is your spoken production CEFR level?

- *C2 / C1 / B2 / B1 / A2 / A1 / I don't know*

Q14: Knowing my Eiken grade will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q15: Knowing my CEFR level will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Q16: Knowing my TOEIC score will be useful to me when I graduate university. (For future jobs and/or future study after university)

- *Absolutely / Somewhat / Not really / Not at all / Don't know*

Investigating the difficulties for university learners of English in Japan of CEFR B1-level phrases

Takeshi Matsuzaki, Meiji University

Kevin Mark, Meiji University

This study examines the variations in difficulty encountered by university learners of English in Japan with regard to English phrases that are classified as CEFR B1 level by the English Vocabulary Profile (EVP). Of the 332 English phrases categorized as B1 level at the time of investigation, 60 were identified as worthy of close examination for this research. An English phrase test was created, comprising two sections: one testing recognition and the other productive ability. Each section consisted of 60 test items embedded within 11 short written passages that were devised to provide context for the items used in this study. The test was administered to 360 university students in Japan, with the recognition section given first and the production section following immediately after. The results obtained show that there was a wide variation in the difficulty measures of the 60 phrases, and that recognition and production showed a discrepancy in terms of level of difficulty. The latter finding suggests that determination of difficulty based on a single processing mode is unlikely to provide an adequate indication of the difficulty of phrases, and that use should therefore be made of measurement instruments that assess both recognition and production ability. Implications for learning and pedagogy and for future directions for this line of research are discussed.

Keywords: English Vocabulary Profile (EVP), B1 level, phrases, difficulty, recognition, production

1 Introduction

The English Vocabulary Profile or EVP (http://vocabulary.englishprofile.org), a service provided by Cambridge University Press, describes a number of English phrases, in addition to different meanings of individual words, that are estimated to be 'typically known and used' by learners at different proficiency levels designated by the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001; Harrison and Barker 2015). Because everyday language use teems with multi-word expressions (e.g., Erman and Warren 2000) and deviant use of such expressions by L2 learners can result in an increased and sustained processing burden for proficient speakers of the target language (Millar 2010; Stengers et al. 2011), the EVP serves as a valuable online resource for both L2 learners and teachers of English, enabling them to make more informed decisions as to what phrases to focus on at a given stage of learning English as a second language.

The primary objective of this study was to identify the difficulty measures of the selected B1 phrases that the authors identified as worthy of scrutiny, in the expectation that this will eventually lead to well-informed guidance for L2 learners of English in higher education in Japan and also for professionals working with them. The choice to target B1 phrases was made in consideration of the proficiency levels of college learners of English in Japan. We estimated, in accordance with the results of two surveys reported on in Negishi (2012), that the CEFR levels of the great majority of college learners of English

in Japan are likely to fall within A1 and A2 levels, and therefore phrases categorized under the B1 level were selected as items that represent reasonable target items for such learners to work on in order to advance to a higher level.

As far as the authors are aware, the only study to date that has looked into the difficulties for Japanese learners of English phrases sorted by the EVP is the investigation by Negishi, Tono and Fujita (2012). In their study, 100 English 'phrasal verbs' categorized by the EVP under CEFR A1 to B2 levels were examined. The researchers developed a fill-in-blanks phrasal verb test and administered it to some 1,600 Japanese learners, approximately 95% of whom were senior high school students. Negishi et al.'s study demonstrated that the difficulty measures of the phrasal verbs in each of the four levels examined vary to a large extent. Likewise, we anticipated that markedly differing difficulty measures would be found for our target B1 phrases.

One unique characteristic of our study is that it represents an attempt to delve into learners' depth of knowledge of the target phrases, which is not a matter that can be ticked off as simply present or absent: a learner at a certain learning stage may be able to recognize a phrase when it comes to comprehending input, while being unable to produce the same phrase as output. Unlike Negishi et al.'s study, our study sought a more textured understanding of the knowledge of B1 phrases possessed by college learners of English in Japan, by means of conducting an English phrase test that contains both recognition and production components.

2 Method

2.1 Participants

The participants in the study were undergraduate students in Japan taking courses taught by one of the co-authors. A total of 360 students took the two-section phrase test developed for the study, details of which are given below. The data for 59 students, however, were determined to be 'misfits' by the Rasch model, which we decided to adopt for our measurement due to the small sample at hand. The data for these misfit participants were then excluded from further difficulty analyses of the target phrases, as the primary objective of this study was to identify the difficulty measures of the target phrases, not to assess the learners' ability *per se.* This is a relatively large number of discounted misfits, but we were still able to ensure a sample size that is more than sufficient for Rasch measurement. The dispersion of the ability measures obtained of the 301 'fit' participants can be found in Table 2 and Figure 1 below.

2.2 Target B1 phrases

At the outset of the study all phrases assigned B1 level were extracted from the overall list of phrases on the EVP website, which gave the authors 332 candidates from which to choose for the investigation. Each one of these phrases was then carefully reviewed, albeit intuitively, in terms of whether the college students we teach use them in speaking or writing, and whether we think their meaning can be easily expressed by the same students in a circumlocutory way. Although this classification was solely based on intuition, we felt confident, as experienced teachers, in our ability to identify those phrases that we rarely find in our students' speech or written production, and which would be relatively difficult for these learners to bypass if they were to express the core meanings and functions of these phrases with their current English ability. On the basis of these criteria the 332 candidate phrases were narrowed down to 66.

In order to further reduce the total number, three major corpora (BNC, TenTen, and SkELL 3.8) were consulted to see the frequencies with which these phrases occur in them (see Table 1). However desirable it might have been to examine all 66 phrases, we wanted to also suppress the risk of negatively impacting performance on the test, as the test-takers would likely become increasingly fatigued by the sheer number of items. Since our ultimate goal with this study, as researchers who are also educators, was to acquire data that will be helpful to learners and practitioners, the most frequent phrases in the corpora were not

considered crucial candidates for this study, on the grounds that learners have a better chance of learning such phrases independently of direct instruction, by virtue of the fact that they will naturally encounter them more often than others. The top six high-frequency phrases on the list were thus excluded from our investigation, and the target phrases for this study were finalized with the remaining 60 phrases.

Table 1. *The sixty-six candidate B1 English phrases for the present study*

Corpus frequency order	Phrase	Average frequency per million words	Corpus frequency order	Phrase	Average frequency per million words
1	at least	211.7	34	go wrong	7.9
2	you know	205.6	35	would rather	7.8
3	rather than	151.9	36	feel like/as if	7.8
4	I know	140.5	37	I bet (you)	7.0
5	at all	124.4	38	change your mind	6.6
6	used to do/be sth	72.9	39	fall asleep	5.8
7	up to 10, 20, etc.	61.4	40	either way	5.7
8	at the same time	54.6	41	again and again	5.7
9	keep doing sth	53.8	42	be up to sb	5.6
10	so far	43.8	43	it/that depends	5.5
11	after all	41.7	44	get worse	5.0
12	get down/into/off, etc.	41.3	45	things like that	5.0
13	make sb do sth	41.0	46	get sth wrong	5.0
14	not until	38.6	47	(just) in case	4.4
15	as long as	37.2	48	tell sb how/what/when to do sth	3.9
16	at first	35.6	49	as you know	3.9
17	not really	33.5	50	that sort of thing	3.9
18	in time	32.2	51	have sth in common	3.5
19	be supposed to do sth	29.0	52	for fun or for the fun of it	3.3
20	be worth sth/doing sth	27.4	53	feel sorry for	2.8
21	be willing (to do sth)	27.1	54	can't/couldn't help doing sth	2.8
22	take advantage of sth	23.1	55	tired of doing sth	2.8
23	no way	20.9	56	take a break/rest, etc.	2.3
24	get rid of sth	19.4	57	wait a minute	2.0
25	in advance	19.4	58	keep sb waiting	1.6
26	can afford	18.8	59	on purpose	1.6
27	do badly/well	14.2	60	be just about to do sth	1.1
28	ever since	13.5	61	if I were you	.8
29	get to know sb/sth	9.6	62	get cold/ill/late, etc.	.6
30	up to date	8.9	63	miss a chance/opportunity	.6
31	What if ...?	8.5	64	feel bad about sth/doing sth	.5
32	go badly/well, etc.	8.1	65	have been meaning to do sth	.4
33	get/become used to sb/sth/doing sth	7.9	66	be up to sth	.3

2.3 Measurement instrument

Using the English phrase test designed in the study by Schmitt, Dörnyei, Adolphs and Durow (2004) as a model, a test consisting of two distinct sections, the combined results of which would be analyzed by means of the Rasch model, was developed for this study: a multiple-choice recognition section (henceforth, the R-section) and a fill-in-blanks production section (henceforth, the P-section).[1] As with Schmitt et al's study, the decision was made that all test items be embedded in some kind of meaningful context that would allow test-takers to process the target phrases in as naturalistic a way as possible. To this end, we wrote 11 short dialog or monolog contexts (seven and four each), which respectively incorporated five or six of the target phrases. The identical contexts were used for both sections, and 60 test items, each addressing one of the 60 target phrases, were prepared for each section.

As this study was going to be conducted with university students in Japan, the great majority of whom are estimated to be working towards the B1 level, we consulted the EVP website to ensure that none of the words used in the test would exceed B1 – including the directions for taking the test, the multiple-choice distractors in the R-section, and the synonymous expressions for the phrases in the P-section (see below for details). Limiting the test's vocabulary level to B1 level or below served to maintain the unidimensionality of the test items and the attribute of equal discriminative power – two of the prerequisites for using the Rasch model. This enabled us to assume that the determining factor for whether the participants could figure out the right answer to each item was restricted to their knowledge of the B1 phrase in question, and that the whole test would not be too difficult for them (cf. Shimada 2006).

With regard to the R-section, three distractors were generated for each multiple-choice item, all devised in such a way that they would be as similar in meaning and length to the correct form as possible. It was thus expected that the ability to make the correct choice for an item would be based on knowledge of the phrase in question. This would be the case even if, importantly, participants chose distractors for other items in the same context (as the way in which the story was unfolding should have been clear to them). In other words, local independence – another condition to be met when applying the Rasch model – was secured in this section. Also, a fifth option ("I DON'T KNOW") was prepared for each item, the purpose of which was to minimize wild guesses – yet another condition to be satisfied for using the Rasch model – when participants had little or no clue about the target item in question. Below is an example item from the R-section.

Example:
Learning English is boring and it is also hard work. I **[1]** _____ lists of words.
1.   (A) stay repeating
        (B) hold repeating
        (C) remain repeating
        (D) keep repeating
        (E) I DON'T KNOW
                [Answer: (D)]

Moving on to the description of the P-section, the primary motivation for including this component as well as the recognition section was to add a further dimension to our understanding of the participants' depth of knowledge of the target phrases. Inclusion of this section was also meant to further weaken the influence of any wild guesses in the R-section when evaluating the participants' ability measures (see Section 2.5 for scoring details).

---

1.   The purpose of Schmitt et al.'s study was not to examine the difficulty measures of English phrases but to investigate the influence of learner characteristics, such as attitudes toward L2 learning, on the learning of multi-word expressions.

In devising the fill-in-blanks items in the P-section, we followed, although not entirely, the test design in the study by Schmitt et al. (2004), adopting blended elements of cloze and C-test techniques. In our study, each target phrase was first categorized in terms of how many 'keywords' (or lexically strong words) it contains, with lexically weak words deliberately excluded from that counting. For example, the phrase *be worth sth/doing sth* was classified as containing one keyword, and both *be* and *sth/doing sth* were discounted. There were, then, eight one-keyword phrases, 30 two-keyword phrases, 17 three-keyword phrases, four four-keyword phrases, and one five-keyword phrase.

In the case of one-keyword phrases, the keyword was substituted by a blank followed by the word's final letter.[2] It should be noted that we ensured that the length of each blank corresponded to that of the word in question, which constituted a hint for the respondents.[3] So the phrase *be just about to do sth*, for example, would appear as *was just ____t to say* (italics are used here but not in the test). Two-keyword phrases were displayed with the first word replaced by a blank with its final letter remaining, and the second word with a complete blank. So *get off* (registered as *get down/into/off, etc.* on the EVP website) appeared as *___t ____*. Phrases of three keywords such as *get rid of sth* appeared as *___t ___d of*, with the third keyword left untouched. Phrases containing four or five keywords were presented with two words being completely blanked (e.g., *at the ____ ____* for *at the same time*).

Variations on this treatment were applied in a number of cases. When a word to be blanked with its final letter remaining ended with a plural or third-person singular *-s*, *-ing*, *-ed*, *-l*, *-ll*, *-ly*, *-e* or *-h*, one extra letter was left. Adjustments were also made, elaborated on below, when the blanks for an item were filled in an unanticipated way by one or more of the native speakers participating in the pilot. Also elaborated on below are some particular cases, such as where *__ ____se* (for *on purpose*) was used instead of *__n _____*.

We wanted to see in this section whether the participants would be able to come up with the targeted phrase for each test item, given the context and the hints, rather than whether they could simply figure out the meaning of that phrase. A gloss of the meaning expressed by the item in question was therefore given to the participants in the right margin. With the provision of synonymous expressions in the P-section, local independence of the items – again, a prerequisite of the Rasch model – was expected to be protected.

An additional note regarding those synonymous expressions is that the parameter of each paraphrase was tailored so that the phrase in question would be neither too obvious nor too much of a riddle for the test-takers, and that creating the alternative wording, which had to adhere to such restrictions as using words up to B1 level, would be feasible on our part. Thus, for instance, the alternative expression prepared for the phrase *make sb do sth*, which was used in the sentence *Don't make me do that again!*, was *I don't want to have to do* (for the underlined words).

Each set of words corresponding to a paraphrase (i.e., each test item in the P-section) was shown in bold font with shaded background, as shown in the following example (with the paraphrase appearing in italics on the right).

Example:

WOMAN:	What do you do if you see a student sleeping in your class?	
MAN:	**It _____ds.** Sometimes I just make a joke.	(*I can't give the same answer in every situation*)

[Answer: It depends]

Draft items were piloted on three native-speaker university teachers of English in Japan.[4] Although the finalized test was administered with the R-section first and the P-section second, the native speakers

---

2. In Schmitt et al.'s study, the word's initial letter(s) was/were left and the rest of it were blanked.
3. No such adjustment in length was made in Schmitt et al.
4. One of them is no longer in Japan.

were, for this piloting, asked to complete the P-section before moving to the R-section. The main reason for doing so was that we wanted them to think about the blanks without the possibility of any inhibitory priming effects arising from their taking the R-section first: we assumed that they might come up with some words that we did not anticipate, and that they might even be unable to fill in some of the blanks. There were in fact some such cases, and there was also one item where one of the native speakers failed to fill in the blanks; the number of items in the P-section for which the native speakers performed unexpectedly was three, six, and seven respectively. There were also two cases in which the native speaker answers made us aware of multiple possible answers. In the case of the R-section, one of the native speakers made all the choices that we expected, while the two others made what we deemed to be mistakes with one and two items respectively. There were a further seven test items for which we discovered one of the distractors to also be an acceptable choice.

The pilot test thus led to revisions to each of the test items in question. The finalized recognition and production sections of the test are available in Appendices A and B, along with a list of the notes regarding the cases where spelling hints in the P-section were prepared in specific ways in Appendix C.

2.4 Procedure

The finalized test was administered in the following manner. First, all instructions were read to participants in Japanese in order to ensure that there would be no misunderstanding about the English instructions provided in the test booklets. A copy of the R-section booklet was then given to each participant. The participants were instructed to take as much time as they wished on this section. Upon completing the R-section, each student submitted their completed R-section booklet, received a copy of the P-section booklet, and then spent as much time as they wished on this second section, after which the test booklet was submitted.[5]

It should be kept in mind that the recognition section was conducted before the production section, which is the reverse of how the native speakers took the test in the pilot stage.[6] The results reported on below would thus have been different, quite possibly to a large extent, had the P-section been administered first. It can be assumed that items in the P-section would have been far more difficult to answer correctly without the learners' residual memory of the R-section. Indeed, it seems very likely that the main reason that the three native speakers were unable to fill in some blanks in the ways we anticipated is that they were asked to work on the P-section first. The R-section, on the other hand, would have been much less difficult had it followed the P-section. Either way, the memory trace from the section implemented initially would affect the test-takers' performance on the second section, and the judgment made for this study was that more informative data would be obtained by giving the R-section first.

It should also be noted about the procedure that we divided the participants into two groups of approximately the same number. Each group was given the contexts in reverse order to the other group. The first context for one group, entitled 'Learning English', thus appeared last for the other group, and so on. This reversal was adopted in order to average out the effect of cognitive fatigue on performance, as the total number of test items was quite large (60 x 2 = 120) although the same 11 contexts were repeated in the latter half of the test.

2.5 Scoring

For each P-section item for which the blanks were filled in correctly, 1 point was allotted, whereas 2 points were given for each correctly chosen item in the R-section.[7] This scoring method was adopted

---

5. It took about twenty minutes for the fastest participants to finish the entire test, and about fifty minutes for the slowest ones.
6. In Schmitt et al. (2004), the production test was administered before the recognition test, too.
7. In the scoring for the P-section, intelligible spelling errors were not penalized.

on the assumption, explained above, that the P-section must have been less difficult than it would have been had the R-section not been implemented first. That is, the participants' memory trace from taking the R-section can be assumed to have helped them fill in the blanks in the P-section, and thus there was a need to compensate for this priming effect on scoring.[8]

A further adjustment was that for each item that was answered correctly in both sections, 4 points were given instead of 3 points (i.e., the sum of 2 and 1) – the full score for this test was therefore 240 (4 x 60). The extra point was allotted largely as a way to diminish the effect of wild guesses on scoring. Wild guessing is a persistent issue in multiple choice tests that test givers have strived to eliminate, but one that is yet to be adequately addressed (Choi 1992); even though the participants in this study were encouraged to choose the option "I DON'T KNOW" when uncertain about an item in the R-section, there was no guarantee that they actually did this. Because the Rasch model presupposes a minimum level of wild guessing (which will also affect the degree to which equal discriminative power among the test items will be established), a participant in this study getting an item right in both sections was interpreted as evidence of not having guessed wildly for the phrase in question, and a bonus point was therefore justified. A further case for this bonus point can be made if one regards the ability to both correctly produce and recognize a phrase as evidence that the test-taker's knowledge of it goes beyond mere recognition level.

3 Results and discussion

3.1 Results of the phrase test

Table 2 summarizes the statistics for the measured persons and items, derived from Rasch measurement using the software *Winsteps*. Starting with the measured persons, the table only summarizes the data for the 301 fit participants, whose infit mean-square (MnSq)[9] figures range within .75 and 1.30 (cf. Bond and Fox 2007; McNamara 1996); as mentioned above, the data for 59 participants were counted out as misfits. The average θ (person ability measure) of the fit participants is .21, the standard deviation (SD) is .33, and the reliability of their estimated θ is sufficiently high (α = .91). Turning to the measured items, the average δ (item difficulty measure) of the items is .00, the SD is .35, and the reliability of the items' estimated δ is .98. With the average score of the participants being 133.6 out of 240 (about 56%) and this high level of reliability, this phrase test should be regarded as a highly reliable measurement instrument for the target B1 phrases in this study (although, as will be seen below, there were two misfit items). That is, similar item difficulty measures will be derived if the same test is administered to other learners, especially learners whose English proficiency level is about the same as or not too far off from that of the fit participants in this study.

Figure 1, the distribution map of the diverse ability measures of the 301 fit participants and the varying difficulty measures among the 60 B1 phrases, illustrates one possibility to consider about the level of the phrases. While the difficulty measures for all items falling within a range of -1 to 1 should be interpreted as a corroboration of the EVP's CEFR level assignments of the phrases (cf. Negishi, Tono and Fujita, 2012), the gap between the most and least difficult phrases is arguably wide. The variances may thus suggest that certain phrases should perhaps be assigned two successive CEFR levels as opposed to a single distinct level. The variances could also suggest, if one wishes to adhere to a 'one level for one phrase' categorization, the validity of adopting the CEFR's newly-proposed 11 levels (Council of Europe 2018) in place of its previous six levels.[10] This is a tentative proposal, however, because the current study focused solely on the phrases on the B1 list, instead of including A2 and B2 phrases in the test.

---

8. Schmitt et al. (2004) argue that priming effects were minimized in their study, as they sandwiched three other language tests and a questionnaire between their productive and receptive phrase tests.

9. See: https://www.winsteps.com/winman/misfitdiagnosis.htm

10. So the least difficult phrases investigated in this study such as *if I were you* and *get to know sb/th* (see Table 3 below) can be labeled as A2-B1 (or more simply 'A2+'), and the most difficult ones such as *either way* and *for fun or for the fun of it* (also see Table 3) as B1-B2 (or 'B1+').

Table 2. *Test statistics*

```
      SUMMARY OF 301 MEASURED Persons
-------------------------------------------------------------------------------
|           TOTAL                        MODEL      INFIT        OUTFIT      |
|           SCORE     COUNT     MEASURE   S.E.    MNSQ   ZSTD    MNSQ   ZSTD  |
|---------------------------------------------------------------------------|
| MEAN     133.6      60.0       .21      .09     1.00    .0     1.01    .1  |
| P.SD      37.5       .0        .33      .01      .11    .8      .14    .7  |
| S.SD      37.5       .0        .33      .01      .11    .8      .14    .7  |
| MAX.     231.0      60.0      1.47      .21     1.30   1.9     1.30   1.8  |
| MIN.      31.0      60.0      −.92      .09      .77  −1.8      .75  −1.7  |
|---------------------------------------------------------------------------|
| REAL RMSE    .10 TRUE SD    .32  SEPARATION  3.26  Person RELIABILITY  .91 |
|MODEL RMSE    .10 TRUE SD    .32  SEPARATION  3.35  Person RELIABILITY  .92 |
| S.E. OF Persons MEAN = .02                                                 |
-------------------------------------------------------------------------------
Persons RAW SCORE−TO−MEASURE CORRELATION = .99
CRONBACH ALPHA (KR−20) Persons RAW SCORE "TEST" RELIABILITY = .92  SEM = 10.86

      SUMMARY OF 60 MEASURED Items
-------------------------------------------------------------------------------
|           TOTAL                        MODEL      INFIT        OUTFIT      |
|           SCORE     COUNT     MEASURE   S.E.    MNSQ   ZSTD    MNSQ   ZSTD  |
|---------------------------------------------------------------------------|
| MEAN     670.4     301.0       .00      .04     1.00    .0     1.01    .1  |
| P.SD     207.5       .0        .35      .00      .11   1.8      .15   1.8  |
| S.SD     209.2       .0        .35      .00      .12   1.9      .15   1.8  |
| MAX.    1097.0     301.0       .66      .06     1.46   7.3     1.65   8.0  |
| MIN.     294.0     301.0      −.89      .04      .71  −5.4      .70  −4.5  |
|---------------------------------------------------------------------------|
| REAL RMSE    .04 TRUE SD    .35  SEPARATION  8.06  Items  RELIABILITY  .98 |
|MODEL RMSE    .04 TRUE SD    .35  SEPARATION  8.22  Items  RELIABILITY  .99 |
| S.E. OF Items MEAN = .05                                                   |
-------------------------------------------------------------------------------
Items RAW SCORE−TO−MEASURE CORRELATION = −1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

    MEASURE                                      |                            MEASURE
      <more> −−−−−−−−−−−−−−−−−−− Persons −+− Items  −−−−−−−−−−−−−−−− <rare>
        2                                         +                            2
                                                  |
                                                  |
                                                  |
                                                  |
                                                  |
                                                  |
                                                  |
                                            .     |
                                                  |
                                                  |
                                            .     |
                                                  |
        1                                   .   + |                            1
                                           .#   |
                                          .  T|
                                            #   |
                                           ##  |T
                                         ####  |   X
                                        .##### |   XXX
                                       ####### S|  XX
                                       .######### |   X
                                    ############# |   XXX
                                      ##########  |S  XXXX
                                       .######### |   XXXXX
                                   ############# M|  XX
                                     ############ |   X
                                       ######### |   XXX
        0                              .######### +M XXXX                       0
                                        ######## |   XXXXX
                                       .##### S|   XXXX
                                        ####   |   XXXXXXX
                                       .#####  |   XXX
                                        .###  |S XXX
                                        .##   |   XX
                                        .  T|   X
                                        .     |
                                        #     |   XXXX
                                        .     |
                                        .     |T
                                        .     |
                                        .     |   X
                                        .     |
       −1                               .   + |                            −1
          <less> −−−−−−−−−−−−−−−−−−− Persons −+− Items  −−−−−−−−−−−−−−−− <freq>
          EACH "#" IN THE Persons COLUMN IS 2 Persons: EACH "." IS 1
```
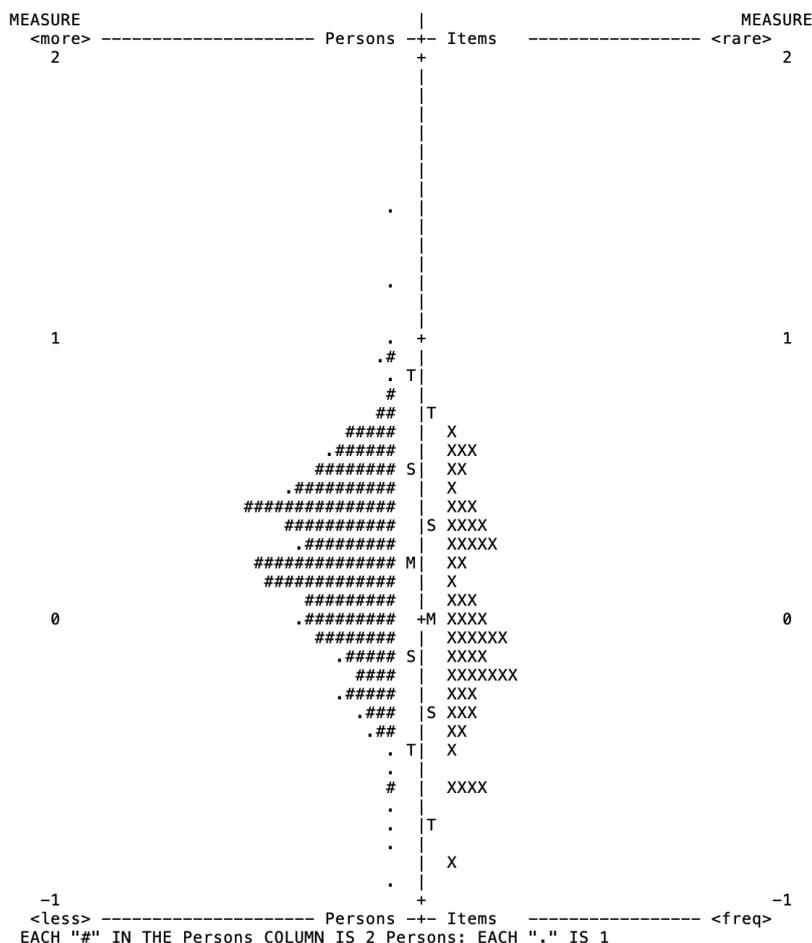
Figure 1. The person by item distribution map of the test.

Finally, the statistics for the individual items (Table 3) reveal the difficulty measures for the top 27 phrases (from *either way* to *feel bad about sth/doing sth*) as over 0.00. It is probably safe then to argue that learners whose proficiency level is either already at or approaching B1 level would benefit by focusing more on these phrases than the remaining 33 phrases. For learners stagnating at A2 (or even A1) level, the opposite may be the case, although of those 33 phrases, the test items created in this study for *keep sb waiting* (entry no. 41) and *in advance* (entry no. 8) were calculated by *Winsteps* as misfit items.

Table 3. *Test item statistics*

Items STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Items
44	294	301	.66	.05	.94	−.6	1.00	.0	.39	.35	30.9	34.0	either way
22	313	301	.62	.05	.94	−.6	1.06	.5	.23	.36	40.9	32.9	for fun or for the fun of it
54	313	301	.62	.05	1.11	1.2	1.22	1.9	.29	.36	18.3	32.9	be just about to do sth
10	318	301	.60	.05	.89	−1.3	1.06	.6	.28	.36	45.8	32.6	ever since
34	341	301	.56	.05	1.16	1.8	1.11	1.0	.35	.37	28.9	31.0	go badly/well, etc.
53	341	301	.56	.05	.91	−1.1	.98	−.2	.30	.37	39.2	31.0	have been meaning to do sth
49	392	301	.45	.04	1.15	2.0	1.21	2.1	.38	.38	16.3	27.2	do badly/well
56	404	301	.43	.04	1.19	2.6	1.39	3.8	.10	.39	27.2	26.7	be up to sth
57	422	301	.40	.04	.98	−.3	1.01	.2	.38	.39	19.3	25.8	that sort of thing
40	428	301	.39	.04	1.02	.3	1.03	.4	.36	.39	16.9	25.6	not until
24	442	301	.37	.04	.88	−1.8	.97	−.4	.38	.40	31.9	24.7	be worth sth/doing sth
37	456	301	.34	.04	1.03	.5	.99	.0	.45	.40	16.3	24.5	go wrong
20	469	301	.32	.04	.95	−.8	.91	−1.1	.52	.40	13.6	23.7	(just) in case
30	478	301	.30	.04	.85	−2.5	.84	−2.1	.54	.41	17.9	23.2	on purpose
60	487	301	.29	.04	1.22	3.3	1.21	2.6	.41	.41	11.0	22.9	would rather
33	488	301	.29	.04	1.03	.4	.98	−.2	.52	.41	11.0	22.9	be supposed to do sth
13	500	301	.27	.04	1.01	.2	1.02	.2	.43	.41	17.3	22.2	so far
35	506	301	.26	.04	.97	−.4	.96	−.5	.48	.41	13.6	22.0	What if ...?
59	520	301	.24	.04	.96	−.7	.94	−.8	.52	.41	15.0	21.0	can afford
5	527	301	.23	.04	1.01	.2	1.01	.2	.34	.42	17.9	20.8	after all
45	563	301	.17	.04	1.12	2.1	1.24	3.2	.22	.42	14.0	19.4	take a break/rest, etc.
47	597	301	.12	.04	.94	−1.0	.94	−.9	.41	.43	16.3	17.1	feel sorry for
29	612	301	.09	.04	1.03	.6	1.10	1.5	.23	.43	15.6	16.5	feel like/as if
32	618	301	.09	.04	.79	−4.1	.78	−3.5	.54	.43	18.3	16.4	get/become used to sb/sth/doing sth
28	624	301	.08	.04	1.02	.4	1.00	.1	.39	.43	11.6	16.4	not really
58	659	301	.02	.04	.98	−.4	.96	−.5	.40	.43	17.3	14.8	get sth wrong
46	663	301	.02	.04	1.01	.3	1.03	.4	.36	.43	12.3	14.8	feel bad about sth/doing sth
51	685	301	−.02	.04	1.15	2.6	1.17	2.4	.39	.43	10.3	14.0	tired of doing sth
7	692	301	−.03	.04	1.09	1.6	1.08	1.1	.42	.43	11.6	13.7	have sth in common
55	697	301	−.03	.04	1.04	.7	1.01	.1	.44	.43	9.6	13.6	can't/couldn't help doing sth
41	702	301	−.04	.04	1.46	7.3	1.65	8.0	.21	.43	9.0	13.4	keep sb waiting
15	722	301	−.07	.04	.98	−.3	1.00	.1	.46	.43	9.0	13.2	as long as
17	729	301	−.08	.04	.97	−.5	.94	−.9	.56	.43	6.0	12.4	it/that depends
50	732	301	−.09	.04	.86	−2.6	.86	−2.1	.52	.43	15.0	12.7	up to date
42	736	301	−.09	.04	.99	−.1	1.00	.0	.41	.43	9.0	12.4	up to 10, 20, etc.
25	766	301	−.14	.04	.83	−3.0	.81	−2.7	.59	.43	13.6	12.1	miss a chance/opportunity
8	774	301	−.15	.04	.71	−5.4	.70	−4.5	.58	.43	19.6	13.3	in advance
43	775	301	−.16	.04	1.08	1.3	1.05	.7	.39	.43	14.3	13.3	get down/into/off, etc.
19	776	301	−.16	.04	1.14	2.3	1.15	1.9	.37	.43	12.3	13.2	tell sb how/what/when to do sth
27	803	301	−.20	.04	.96	−.6	.97	−.3	.41	.43	13.3	13.3	at the same time
48	804	301	−.20	.04	.94	−1.0	.98	−.2	.47	.43	17.9	13.2	take advantage of sth
9	805	301	−.20	.04	1.00	.1	1.02	.3	.37	.43	12.3	14.0	at first
52	814	301	−.22	.04	.87	−2.3	.87	−1.6	.43	.43	19.6	13.8	get worse
4	816	301	−.22	.04	1.09	1.4	1.09	1.1	.38	.42	12.6	13.8	be willing (to do sth)
14	817	301	−.22	.04	1.01	.3	1.02	.3	.44	.42	14.3	14.0	no way
31	817	301	−.22	.04	1.09	1.4	1.17	2.0	.43	.42	10.6	14.0	get rid of sth
23	827	301	−.24	.04	.86	−2.3	.89	−1.3	.48	.42	16.6	14.2	things like that
38	843	301	−.27	.04	1.09	1.4	1.06	.7	.39	.42	16.9	15.5	make sb do sth
39	852	301	−.28	.04	.94	−.9	1.00	.0	.48	.42	17.6	16.9	in time
3	865	301	−.31	.04	.97	−.5	.90	−1.1	.50	.42	17.6	17.1	be up to sb
36	867	301	−.31	.04	.95	−.7	.96	−.4	.38	.42	21.3	18.2	wait a minute
16	873	301	−.32	.04	.92	−1.2	.90	−1.1	.47	.42	23.9	18.3	I bet (you)
26	905	301	−.38	.04	1.00	.1	.93	−.6	.44	.41	23.3	21.6	change your mind
1	920	301	−.41	.04	1.04	.5	1.08	.7	.33	.40	23.3	25.7	keep doing sth
11	947	301	−.46	.05	1.10	1.2	1.17	1.4	.34	.40	30.9	32.3	get cold/ill/late, etc.
2	994	301	−.57	.05	.92	−.9	.92	−.5	.42	.38	43.9	41.6	again and again
18	996	301	−.57	.05	1.08	.8	.93	−.5	.46	.38	44.5	41.6	fall asleep
12	997	301	−.58	.05	.95	−.5	.92	−.5	.44	.38	43.9	42.4	as you know
6	1006	301	−.60	.05	.95	−.4	.88	−.8	.37	.37	44.5	43.6	get to know sb/sth
21	1097	301	−.89	.06	.91	−.6	.71	−1.4	.41	.31	74.8	73.8	if I were you
MEAN	670.4	301.0	.00	.04	1.00	.0	1.01	.1			20.6	22.0	
P.SD	207.5	.0	.35	.00	.11	1.8	.15	1.8			12.4	10.8	

3.2 Implications for learning and teaching

Several further implications for learning and teaching emerge from the participants' performance on each section of the test. At this point, we only consider the relatively difficult phrases. Table 4 is a detailed description of the participants' performance on the 27 phrases rated as having a difficulty measure above 0.00. Phrases for which the accuracy rate is low in both sections (such as *either way*, *be just about to do sth*, *go badly/well, etc.*, *do badly/well*, *would rather*) are probably the biggest challenges to learners who are approximating to or currently at the B1 level, which suggests that they deserve more focused study and instruction than other phrases.

In the case of those phrases for which a relatively large percentage of the participants were able to choose the correct forms in the R-section but unable to successfully fill in the blanks in the P-section (such as *that sort of thing*, *(just) in case*, *on purpose*, *feel sorry for*, *get/become used to sb/sth/doing sth*), output training alone could be very effective for A2-B1 learners.

Where a large proportion of the participants performed correctly only in the P-section (such as *for fun or for the fun of it*, *ever since*, *have been meaning to do sth*, *be worth sth/doing sth*, *feel like/as if*), college learners are likely to be able to recognize the phrases (or their constituent words) without yet being sufficiently familiar with their collocational attributes. Such linguistic features might be best handled by explicit instruction that draws learners' attention to them.

The results as a whole, which reflect a wide range of performance, appear to us to indicate that the idea, mentioned above, that certain phrases be assigned two successive levels (or different levels using the CEFR's new 11 levels) would be more helpful if complemented by the potentially equally important proposal that distinction between production and recognition be made for level assignments.

Lastly, the participants' performance aside, focused study may be very effective if directed to low-frequency phrases (such as *be just about to do sth*, *have been meaning to do sth*, *be up to sth*, *on purpose*, *feel bad about sth/doing sth*), simply because learners, including those in the Japanese context, appear to have fewer opportunities to encounter and learn them in natural input.[11]

3.3 Caveats in interpreting the data

There are some caveats in interpreting the data gained in this study. To begin with, although measures were taken to minimize the influence of wild guesses, which would help to maintain parity of the discriminative power of the test items, it is possible that better test design and scoring could perhaps have further reduced that influence. In addition, priming effects were inevitable, given the content and procedural structure of the test. Thus, while the overall difficulty measures may have been roughly the same even if the administration of the R-section and the P-section had been reversed, the distribution of the accuracy rates for the two sections would probably have been somewhat different. Another issue is to do with how the blanks in the P-section were constructed: whether the rather complicated criteria governing their design may have affected the participants' performance. Issues with priming effects and the construction of the blanks weakened the integrative quality of the P-section and therefore the overall validity of the test as a tool to measure the actual difficulties of the target phrases. Last but not least, the range of the average difficulty measures found for the contexts, shown in Table 5, may not be negligible. It is important to note that this was the case even though the contexts were presented to the participants in two orders (see Section 2.4). There thus remains room for doubt as to whether local independence was sufficiently secured in this test.

---

11. The correlation between the phrases' difficulty measures and their average frequency figures derived from the three corpora consulted (BNC, TenTen, and SkELL 3.8) is virtually non-existent (r = -.072, p = .584), suggesting that a phrase's difficulty has little to do with its frequency.

Table 4. *The participants' performance on each section of the test*

δ		Item		Distribution of the participants			
Ranking	Measure	Phrase	Average frequency per million words	Correct in both R- & P-sections (M = 44%)	Correct only in R-section (M = 13%)	Correct only in P-section (M = 21%)	Incorrect in both sections (M = 22%)
1	0.66	either way	5.7	11%	15%	25%	50%
2	0.62	for fun or for the fun of it	3.3	10%	9%	45%	36%
	0.62	be just about to do sth	1.1	12%	23%	12%	53%
4	0.60	ever since	13.5	12%	3%	52%	33%
5	0.56	have been meaning to do sth	.4	13%	6%	48%	33%
	0.56	go badly/well, etc.	8.1	18%	8%	27%	48%
7	0.45	do badly/well	14.2	20%	22%	7%	51%
8	0.43	be up to sth	.3	19%	9%	39%	33%
9	0.40	that sort of thing	3.9	18%	26%	15%	41%
10	0.39	not until	38.6	20%	19%	23%	38%
11	0.37	be worth sth/ doing sth	27.4	22%	2%	53%	22%
12	0.34	go wrong	7.9	26%	8%	31%	35%
13	0.32	(just) in case	4.4	25%	23%	10%	42%
14	0.30	on purpose	1.6	23%	31%	4%	42%
15	0.29	be supposed to do sth	29.0	30%	11%	19%	40%
	0.29	would rather	7.8	31%	17%	6%	47%
17	0.27	so far	43.8	30%	5%	37%	29%
18	0.26	What if …?	8.5	29%	13%	25%	33%
19	0.24	can afford	18.8	30%	21%	11%	38%
20	0.23	after all	41.7	29%	13%	34%	24%
21	0.17	take a break/ rest, etc.	2.3	33%	7%	44%	17%
22	0.12	feel sorry for	2.8	32%	27%	16%	25%
23	0.09	feel like/as if	7.8	36%	1%	58%	5%
	0.09	get/become used to sb/sth/ doing sth	7.9	34%	28%	13%	25%
25	0.08	not really	33.5	39%	7%	36%	18%
26	0.02	get sth wrong	5.0	39%	22%	18%	21%
	0.02	feel bad about sth/doing sth	.5	42%	9%	36%	14%

Note. Average frequencies per million words were derived from BNC, TenTen, and SkELL 3.8.

Table 5. *Context-by-context average difficulty measures*

Context no.	Item no.	Theme of the context	Average δ
1	1-5	Learning English	-0.26
2	6-10	Romance	-0.08
3	11-16	Health	-0.23
4	17-21	Teaching	-0.28
5	22-26	Travel	0.05
6	27-32	Watching TV	0.02
7	33-38	Asking for help, but not in a direct way	0.15
8	39-44	Late for the test	0.08
9	45-50	Absent from school	0.08
10	51-55	A difficult relationship	0.18
11	56-60	Parents and marriage	0.28

4 Conclusion

This study has examined the variations in difficulty encountered by university learners of English in Japan with regard to English phrases that the EVP classifies as CEFR B1 level. It has demonstrated that while the B1 level indeed seems to be valid in a broad sense, Japanese university learners do also seem to encounter within this level considerable variations in difficulty. It can be presumed that this range of difficulty also applies to other phrases that were not investigated in this study. The most important assertion that can be confidently made on the basis of the data and analysis presented here is that the inherent difficulty of a phrase differs depending on whether the mode of language processing is production or recognition. The results thus suggest that while it may be seemingly helpful to assign a single level to a phrase, it may be more realistic and ultimately helpful to take account of the processing mode. More detailed accounts of the global difficulty of a phrase, as suggested in this paper, may well be helpful to learners and teachers, although defining global difficulty is a tremendously challenging task.

This study points to a number of future directions. The test developed for this study seems, despite its inherent limitations, worthy of replication to see if its findings are validated with learners of English in different contexts who have non-Japanese L1 backgrounds. Insights gained from the present study can also help in the design of new tests for measuring difficulties of other English phrases. Whatever measurement tool is developed, a phrase ought to be examined from at least the two aspects of recognition and production. The way this notion was handled in this study can easily be extended to listening. While applying it to speaking will be more challenging, it is certainly not out of the question to do so. Also, scientifically more rigorous data could be acquired if, say, a two-section test were to use two different sets of contexts so as to minimize priming effects. Testing involving a very large number of participants would help to resolve the issues associated with wild guessing and equal discriminative power, as such testing would allow for the adoption of a three-parameter logistic model.

In conclusion, it is hoped that, all caveats and limitations considered, this study's findings can lead in the direction of more information for L2 learners of English at the tertiary level within and outside of Japan, for classroom practitioners teaching them and for material developers. It is also hoped that the study has shed light on the complexities inherent in the comprehension and production of multi-word expressions in English, and that future research projects may benefit from these insights.

5 Acknowledgements

6 References

Bond, Trevor & Christine Fox. 2015. *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Choi, Inn-Chull. 1992. *An application of item response theory to language testing.* New York: Peter Lang.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors.* Strasbourg: Council of Europe.

Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open-choice principle. *Text, 20,* 29-62.

Harrison, Julia & Fiona Barker. 2015. *English Profile in Practice,* English Profile Studies volume 5. Cambridge: Cambridge University Press.

McNamara, Tim. 1996. *Measuring second language performance.* London: Longman.

Millar, Neil. 2010. The processing of malformed formulaic language. *Applied Linguistics, 32*(2), 1-21.

Negishi, Masashi. 2012. The development of the CEFR-J: Where we are, where we are going. In N. Tomimori, M. Furihata, K. Haida, N. Kurosawa, & M. Negishi (Eds.), *New Perspectives for Foreign Language Teaching in Higher Education: Exploring the Possibilities of Application of CEFR*, 105-116. Tokyo: WOLSEC, Tokyo University of Foreign Studies.

Negishi, Masashi, Yukio Tono & Yoshihito Fujita. 2012. A validation study of the CEFR levels of phrasal verbs in the English Vocabulary Profile. *English Profile Journal, 3*(1), 1-16.

Shimada, Takuji. 2006. On IRT-based language testing: Aren't there any blind spots? *Foreign language education: Theory and practice, 32,* 29-42.

Schmitt, Norbert, Zoltán Dörnyei, Svenja Adolphs & Valerie Durow. 2004. Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use*, 55-86. Amsterdam: John Benjamins Publishing Company.

Stengers, Helene, Frank Boers, Alex Housen & June Eyckmans. 2011. Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *IRAL, 49*(4), 321-343.

7 Biographies

Takeshi Matsuzaki is an Associate Professor at Bunkyo Gakuin University, Japan. He has taught English at the university level in Japan for over 15 years. He holds an MS in TESOL from University of Pennsylvania and a PhD in Humanities from Tokyo University of Foreign Studies. His research interests include the learning and teaching of multi-word expressions in foreign language contexts and the complex and dynamic nature of second language learning.

Kevin Mark is a Professor at Meiji University, where he has been working since 1991. He argues that, for language teaching research and theory to be relevant, the specialized strands of applied linguistics need to be integrated by broad humanistic and educational principles. His innovative work in autonomy, global education, curriculum, CALL, materials writing and learner corpus development as a part of teaching addresses this apparently paradoxical and overarching question: How can mass education be made a more congenial and human process for teachers and learners alike?

Appendix A
B1 phrase recognition section of the test

Vocabulary Phrases (multiple choices)

NAME _____

Directions:

Each of the following pieces of language is spoken by one or two speakers. Each one contains five or six missing phrases. Choose from (A), (B), (C) or (D) and circle the letter for the phrase which fits best. If you are not sure, circle (E) for "I don't know."

Example

I'm **[1]** ___ a team of twenty people.

[1]   (A) responsible of
           (B) in responsibility for
           (C) the charge of
           (D) in charge of
           (E) I DON'T KNOW

When you are told to, go on to the next page and start taking the test.

Learning English

In the following conversation between two friends, the man complains about how boring and hard learning English is, and the woman gives him some advice.

MAN:	Learning English is boring and it is also hard work. I **[1]** ___ lists of words. Every week I'm doing the same things, **[2]** ___.
WOMAN:	It doesn't have to be like that.
MAN:	What do you mean?
WOMAN:	Whether it's interesting to you or not **[3]** ___. There are actually many interesting ways to study. You only need to **[4]** ___ look for them. I'm sure you will discover that learning English does not have to be boring **[5]** ___.

[1]   (A) stay repeating
       (B) hold repeating
       (C) remain repeating
       (D) keep repeating
       (E) I DON'T KNOW

[2]   (A) in a repeating way
       (B) in a frequent way
       (C) again and again
       (D) again and over
       (E) I DON'T KNOW

[3]   (A) is after you
       (B) is based on you
       (C) is on you
       (D) is up to you
       (E) I DON'T KNOW

[4]   (A) be wanting to
       (B) be meaning to
       (C) be intending to
       (D) be willing to
       (E) I DON'T KNOW

[5]   (A) at the end
       (B) all over
       (C) at last
       (D) after all
       (E) I DON'T KNOW

Romance

In the following conversation between two friends, the woman asks the man about how his relationship with his girlfriend started.

WOMAN:	How did you and your girlfriend **[6]** ___ each other?
MAN:	We met through a friend of ours. She told us we **[7]** ___.
WOMAN:	So you knew **[8]** ___ that you would like each other?
MAN:	No, but we hoped we would.
WOMAN:	How did you feel when you met her?
MAN:	**[9]** ___ I was a little embarrassed. But we very quickly started to feel comfortable with each other.
WOMAN:	And you've been together **[10]** ___?
MAN:	That's right!

[6] (A) get to know
     (B) become knowing
     (C) grow into knowing
     (D) turn into knowing
     (E) I DON'T KNOW

[7] (A) shared many things
     (B) held many things in common
     (C) had a lot in common
     (D) got lots of common things
     (E) I DON'T KNOW

[8] (A) in advance
     (B) before advance
     (C) at the advance
     (D) for advance
     (E) I DON'T KNOW

[9] (A) In the first time
     (B) To begin with it
     (C) First of all
     (D) At first
     (E) I DON'T KNOW

[10] (A) ever after
     (B) after that
     (C) since that
     (D) ever since
     (E) I DON'T KNOW

Health

In the following conversation between a couple, the woman gives the man some advice about how he can lose weight.

MAN:	Do you think I'm **[11]** ___?
WOMAN:	Well, **[12]** ___, a man of your height should be under 70 kilos. **[13]** ___ you've been able to reach 80, right? Let me ask you something. Don't you think you are still eating too much pasta?
MAN:	There **[14]** ___ I am going to eat less pasta. I love it!
WOMAN:	Well, **[15]** ___ you eat so much, **[16]** ___ you won't be able to lose those 10 kilos.

[11] (A) going thinner
     (B) getting thinner
     (C) turning thinner
     (D) being thinner
     (E) I DON'T KNOW

[12] (A) since you know
     (B) for you to know
     (C) since it's known
     (D) as you know
     (E) I DON'T KNOW

[13] (A) For now
     (B) So far
     (C) Before now
     (D) For the past
     (E) I DON'T KNOW

[14] (A) is no way
     (B) are no possibilities
     (C) is not the possibility
     (D) is not a way
     (E) I DON'T KNOW

[15] (A) as far as
     (B) so far as
     (C) as long as
     (D) as if
     (E) I DON'T KNOW

[16] (A) you bet
     (B) the bet is
     (C) I bet
     (D) it's a bet
     (E) I DON'T KNOW

Teaching

In the following conversation between two teachers and colleagues, the woman asks the man how he deals with students sleeping in his class.

WOMAN:	What do you do if you see a student sleeping in your class?
MAN:	**[17]** ___. Sometimes I just make a joke. Once, when I was giving a private lesson, the girl I was teaching **[18]** ___ in front of me.
WOMAN:	What did you do?
MAN:	Well, I wondered for a few minutes, but then she woke up. So, I **[19]** ___ before coming to the next lesson.
WOMAN:	What did you tell her?
MAN:	I told her that she should drink at least three cups of coffee. She smiled, thanked me and said she would, **[20]** ___.
WOMAN:	**[21]** ___, I would have said at least four cups.

[17] (A) The case is different
    (B) It depends
    (C) Things change
    (D) The choices are different
    (E) I DON'T KNOW

[18] (A) went sleeping
    (B) fell asleep
    (C) just slept
    (D) was sleepy
    (E) I DON'T KNOW

[19] (A) told her what to do
    (B) told her what should she do
    (C) told her to do what
    (D) told her what she does
    (E) I DON'T KNOW

[20] (A) just in the case
    (B) just in a case
    (C) just in cases
    (D) just in case
    (E) I DON'T KNOW

[21] (A) If you were me
    (B) If I am you
    (C) If I could be you
    (D) If I were you
    (E) I DON'T KNOW

Travel

In the following conversation between a couple, the man wants the woman to go with him on a very cheap tour to Hawaii.

MAN:	I just heard about a really cheap three-night tour to Hawaii. The flight leaves on Friday afternoon and gets back on Monday at lunchtime.
WOMAN:	That's tomorrow afternoon. Are you crazy?
MAN:	It only costs 50,000 yen for everything. I think we should go, just **[22]** ___. We could go swimming and shopping and do **[23]** ___.
WOMAN:	You're right. It **[24]** ___ the price. It would be crazy to **[25]** ___ when it's so cheap.
MAN:	I'm glad you've **[26]** ___!

[22] (A) for the fun of it
    (B) for the fun
    (C) for its fun
    (D) of the fun for it
    (E) I DON'T KNOW

[23] (A) things like that
    (B) a thing like that
    (C) the stuff like that
    (D) stuff such as that
    (E) I DON'T KNOW

[24] (A) is worth
    (B) is fair by
    (C) is valuable for
    (D) is right for
    (E) I DON'T KNOW

[25] (A) fail an opportunity to go
    (B) escape opportunities to go
    (C) miss a chance to go
    (D) give away chances to go
    (E) I DON'T KNOW

[26] (A) decided to change minds
    (B) changed your mind
    (C) changed your decision
    (D) decided to change
    (E) I DON'T KNOW

Watching TV

In the following conversation, the mother wants her son to study harder.

MOTHER:	I told you not to watch TV and do your homework **[27]** ___. I know that you aren't doing a good job with your homework.
SON:	I am doing a good job!
MOTHER:	Most of your attention is going to the TV, isn't it?
SON:	**[28]** ___. Only a little.
MOTHER:	**[29]** ___ you're doing this just to annoy me. You're doing it **[30]** ___.
SON:	No, I'm not!
MOTHER:	I'm going to **[31]** ___ this TV tomorrow.
SON:	Don't do that!
MOTHER:	You'll **[32]** ___ it.

[27] (A) in the same time
(B) at the same time
(C) in the same moment
(D) at the same moment
(E) I DON'T KNOW

[28] (A) Not really
(B) Not much so
(C) Not really so
(D) Not very
(E) I DON'T KNOW

[29] (A) I'm likely to think
(B) I feel I think
(C) I think it's likely
(D) I feel like
(E) I DON'T KNOW

[30] (A) on purpose
(B) on your purpose
(C) for its purpose
(D) for the purpose of it
(E) I DON'T KNOW

[31] (A) take off
(B) get rid of
(C) put off
(D) hand out
(E) I DON'T KNOW

[32] (A) become used for
(B) become used with
(C) get used for
(D) get used to
(E) I DON'T KNOW

Asking for help, but not in a direct way

In the following conversation, the older sister wants her younger brother to help her with her smartphone.

SISTER:	Why are you watching TV? I thought **[33]** ___ doing your English homework.
BROTHER:	Don't bother me!
SISTER:	Did your English test **[34]** ___ today?
BROTHER:	No!
SISTER:	**[35]** ___ you fail again? Maybe you need some help. Shall I help you?
BROTHER:	**[36]** ___. What are you trying to do here? You're not really interested in my English, are you? Has something **[37]** ___ with your smartphone again? I spent an hour helping you with it yesterday. Don't **[38]** ___ that again!

[33] (A) there was a need for you to be
(B) you had the need for
(C) you were supposed to be
(D) it was the importance of your
(E) I DON'T KNOW

[34] (A) go well
(B) do well
(C) get well
(D) become well
(E) I DON'T KNOW

[35] (A) What's possible if
(B) How possible is it that
(C) How can it be if
(D) What if
(E) I DON'T KNOW

[36] (A) Wait the second
(B) Wait a minute
(C) Hold the second
(D) Hold a minute
(E) I DON'T KNOW

[37] (A) gone wrong
(B) got wrong
(C) come bad
(D) become bad
(E) I DON'T KNOW

[38] (A) make me do
(B) get me
(C) want me for
(D) request me for
(E) I DON'T KNOW

Late for the test

In the following, the speaker is upset because she doesn't think she can get to school before her test starts.

With this train delay I'm not going to get to school **[39]** _____ for the test, even though it **[40]** _____ . I just sent an email to the professor, telling him what's happened. He answered that he can **[41]** ___ for a little while, but only **[42]** ___ ten minutes. It might be quicker if I **[43]** ___ the train and take a bus. But **[44]** ___, I'm going to be too late for the test.

[39] (A) in the time
(B) in time
(C) within the time
(D) within time
(E) I DON'T KNOW

[40] (A) is going to start not by ten
(B) is starting not until ten
(C) isn't starting by ten
(D) isn't going to start until ten
(E) I DON'T KNOW

[41] (A) keep everyone waiting
(B) keep everyone waited
(C) get everyone waiting
(D) get everyone waited
(E) I DON'T KNOW

[42] (A) before
(B) to
(C) up until
(D) up to
(E) I DON'T KNOW

[43] (A) take out
(B) take off
(C) get out
(D) get off
(E) I DON'T KNOW

[44] (A) in both ways
(B) either way
(C) each way
(D) in each way
(E) I DON'T KNOW

Absent from school

In the following, the speaker had an accident and has not gone to school for a while. His classmates have been really kind to him. He is feeling guilty, because he has actually been enjoying himself.

I've had to **[45]** ___ from school for the past three weeks, because of my accident. I **[46]** ___ about the fact that everyone at school has been **[47]** ___ me, because actually I've been enjoying myself. I've been able to **[48]** ___ the break to do a lot of studying. I'm sure that I am going to **[49]** ___ in my classes when I go back to school. I will be **[50]** ___, even though I have been absent for three weeks.

[45] (A) be at home
(B) be on vacation
(C) have time resting
(D) take a break
(E) I DON'T KNOW

[46] (A) have the wrong feeling
(B) have the bad feeling
(C) feel wrong
(D) feel bad
(E) I DON'T KNOW

[47] (A) having a sorry feeling for
(B) being sorry to
(C) feeling sorry for
(D) being sorry thinking about
(E) I DON'T KNOW

[48] (A) take advantage of
(B) make advantage of
(C) take advantage with
(D) make advantage with
(E) I DON'T KNOW

[49] (A) get well
(B) do well
(C) go well
(D) be well
(E) I DON'T KNOW

[50] (A) to date
(B) to the date
(C) up to date
(D) up to the date
(E) I DON'T KNOW

A difficult relationship

In the following, the speaker talks about a difficult classmate. Because of this classmate, she has been having a hard time.

I have a classmate who is always saying negative things about other people. I'm really **[51]** ___ to her. I'm afraid she is **[52]** ___, and I **[53]** ___ say something to her about it. The other day **[54]** ___ something to her when she asked me for advice about a problem she was having. She said she **[55]** ___ saying negative things about other people.

[51] (A) tired to listen
(B) tired of listening
(C) tiring listening
(D) tiring to listen
(E) I DON'T KNOW

[52] (A) getting worse
(B) going worse
(C) getting bad
(D) going bad
(E) I DON'T KNOW

[53] (A) mean to have had to
(B) have had the meaning to
(C) have meant that I
(D) have been meaning to
(E) I DON'T KNOW

[54] (A) it was about the time to say
(B) it was the time for saying
(C) I was just close to saying
(D) I was just about to say
(E) I DON'T KNOW

[55] (A) couldn't stop from
(B) wasn't able to stop from
(C) couldn't help
(D) didn't help
(E) I DON'T KNOW

Parents and marriage

In the following, the speaker talks about his parents. He is worried about how they will accept the news about his marriage.

My Mom and Dad want to know what **[56]** ___ each week. Usually we talk about my part-time job, my friends, my studies and **[57]** ___. They usually think I make decisions without enough thought, which isn't true. So I want to be sure that they won't **[58]** ___ it wrong this time. I am going to tell them that I am getting married next week! I am sure that they will tell me to wait until after I finish my college education. But she comes from a very rich family, so they don't need to worry that **[59]** ___ this. And they **[60]** ___ have a rich son than a poor one!

[56] (A) stuff I have been doing
(B) thoughts have been in me
(C) I have been up to
(D) I have been about
(E) I DON'T KNOW

[57] (A) the thing of that sort
(B) that sort of thing
(C) things that are similar
(D) similar things of that sort
(E) I DON'T KNOW

[58] (A) get
(B) understand
(C) have
(D) hold
(E) I DON'T KNOW

[59] (A) I can't have the money for
(B) my money is not enough for
(C) I can't afford to do
(D) my money cannot do
(E) I DON'T KNOW

[60] (A) will want to
(B) will like to
(C) would like to
(D) would rather
(E) I DON'T KNOW

This is the end of the test. Close this booklet and follow the instructions.

Appendix B
B1 phrase production section of the test

Vocabulary Phrases (filling in blanks)

NAME _____

Directions:

Each of the following pieces of language is spoken by one or two speakers. Each one contains five or six phrases with missing letters or words. You can see the meaning of the phrase on the right side.

Here is an example:

Phrase	Meaning
I'm ___ _____ge of a team of twenty people.	*(I'm responsible for)*

Answer

I'm *in charge* of a team of twenty people.

Shorter or longer lines such as "___" or "_____" give you an idea about how many letters are missing. Try to fill in the missing letters or words. If it is too difficult, just go on quickly to the next phrase with missing letters or words.

When you are told to, go on to the next page and start taking the test.

Learning English

In the following conversation between two friends, the man complains about how boring and hard learning English is, and the woman gives him some advice.

MAN:	Learning English is boring and it is also hard work. I **___p repeating** lists of words. Every week I'm doing the same things **a___ and ___n**.	*(continue to repeat)* *(feeling annoyed about doing the same thing so many times)*
WOMAN:	It doesn't have to be like that.	
MAN:	What do you mean?	
WOMAN:	Whether it's interesting to you or not **is __p __ you**. There are actually many interesting ways to study. You only need to **be wi___g to look for** them. I'm sure you will discover that learning English does not have to be boring **a___r a__**.	*(is something you are responsible for)* *(be positive about looking for)* *(even if you have not thought so until now)*

Romance

In the following conversation between two friends, the woman asks the man about how his relationship with his girlfriend started.

WOMAN:	How did you and your girlfriend **___t to kn___** each other?	*(become familiar with)*
MAN:	We met through a friend of ours. She told us **we had a lot __ _____**.	*(many of our interests and attitudes were the same)*
WOMAN:	So you knew **__ _____ce** that you would like each other?_	*(before meeting each other)*
MAN:	No, but we hoped we would.	
WOMAN:	How did you feel when you met her?	
MAN:	**__t ___t** I was a little embarrassed. But we very quickly started to feel comfortable with each other.	*(In the beginning)*
WOMAN:	And you've been together **e___ s____**?	*(since then)*
MAN:	That's right!	

Health

In the following conversation between a couple, the woman gives the man some advice about how he can lose weight.

MAN:	Do you think I'm **___ting thinner**?	*(becoming thinner)*
WOMAN:	Well, **___s you ___w**, a man of your height should be under 70 kilos. **S___ ___r** you've been able to reach 80, right? Let me ask you something. Don't you think you are still eating too much pasta?	*(although I know you already realize this)* *(Until now)*
MAN:	**There is __ __y** I am going to eat less pasta. I love it!	*(It is impossible)*
WOMAN:	Well, **___s ___g as you eat** so much, **I ___t** you won't be able to lose those 10 kilos.	*(if you continue to eat) (I am sure)*

Teaching

In the following conversation between two teachers and colleagues, the woman asks the man how he deals with students sleeping in his class.

WOMAN:	What do you do if you see a student sleeping in your class?	
MAN:	It _____ds. Sometimes I just make a joke. Once, when I was giving a private lesson, the girl I was teaching __ell ____ in front of me.	*(I can't give the same answer in every situation)* *(started to sleep)*
WOMAN:	What did you do?	
MAN:	Well, I wondered for a few minutes, but then she woke up. So, I just **told her what __ __** before the next lesson.	*(told her what she should do)*
WOMAN:	What did you tell her?	
MAN:	I told her that she should drink at least three cups of coffee. She smiled, thanked me and said she would, ____t in ___se.	*(to make sure that she would not start to sleep during a lesson)*
WOMAN:	__ I ____ **you**, I would have said at least four cups.	*(In your position)*

Travel

In the following conversation between a couple, the man wants the woman to go with him on a very cheap tour to Hawaii.

MAN:	I just heard about a really cheap three-night tour to Hawaii. The flight leaves on Friday afternoon and gets back on Monday at lunchtime.	
WOMAN:	That's tomorrow afternoon. Are you crazy?	
MAN:	It only costs 50,000 yen for everything. I think we should go, just **for the ___n ___ it**. We could go swimming and shopping and do ___gs ___ke that.	*(because it will be enjoyable)* *(other similar activities)*
WOMAN:	You're right. **It is ____th the price**. It would be crazy to m___ a ___ce to go when it's so cheap.	*(For that price, we should join the tour)* *(not take an opportunity to go)*
MAN:	I'm glad **you've ____ged your ___d**!	*(you are not thinking the way you were)*

Watching TV

In the following conversation, the mother wants her son to study harder.

MOTHER:	I told you not to watch TV and do your homework **at the ____ ____**. I know that you aren't doing a good job with your homework.	*(together)*
SON:	I am doing a good job!	
MOTHER:	Most of your attention is going to the TV, isn't it?	
SON:	**__t ____lly**. Only a little.	*(That's not completely true)*
MOTHER:	**I __el ___ke** you're doing this just to annoy me. You're doing it **___ ____se**.	*(My feeling is that)* *(because you want to annoy me)*
SON:	No, I'm not!	
MOTHER:	I'm going to **___t ___d of this TV** tomorrow.	*(take this TV out of the house)*
SON:	Don't do that!	
MOTHER:	You'll **___t __sed to it**.	*(soon start to think that you are fine without a TV)*

Asking for help, but not in a direct way

In the following conversation, the older sister wants her younger brother to help her with her smartphone.

SISTER:	Why are you watching TV? I thought you **were _____sed to be** doing your English homework.	*(should be)*
BROTHER:	Don't bother me!	
SISTER:	**Did your English test __o w___** today?	*(Were you successful in your English test)*
BROTHER:	No!	
SISTER:	**___t __** you fail again? Maybe you need some help. Shall I help you?	*(What will happen if)*
BROTHER:	**___t a ___te**. What are you trying to do here? You're not really interested in my English, are you? Has something **___ne ____g** with your smartphone again? I spent an hour helping you with it yesterday. **Don't ___ke me do** that again!	*(Stop, because I have a question)* *(stopped working properly)* *(I don't want to have to do)*

Late for the test

In the following, the speaker is upset because she doesn't think she can get to school before her test starts.

With this train delay I'm not going to get to school **i___ ___me for the test**, even though **it's ___ going to start ___il** ten. I just sent an email to the professor, telling him what's happened. He answered that he can **___p everyone ____** for a little while, but only **__p ___** ten minutes. It might be quicker if I **__t ___** the train and take a bus. But **___r ___**, I'm going to be too late for the test.	*(early enough to take the test) (it is going to start at)* *(make everyone wait)* *(a maximum of)* *(move out of)* *(whether I stay on the train or take a bus)*

Absent from school

In the following, the speaker had an accident and has not gone to school for a while. His classmates have been really kind to him. He is feeling guilty, because he has actually been enjoying himself.

I've had to **__ke a ___k** from school for the past three weeks, because of my accident. I **__el ___d about** the fact that everyone at school has been **__ling ___y for** me, because actually I've been enjoying myself. I've been able to **__ke a_____ge of** the break to do a lot of studying. I'm sure that I am going to **__o w___** in my classes when I go back to school. I will **be ___p to ___te**, even though I have been absent for three weeks.	*(be absent for a short period)* *(have a guilty feeling about)* *(giving kind attention to)* *(use and not waste)* *(be successful)* *(have the latest information)*

A difficult relationship

In the following, the speaker talks about a difficult classmate. Because of this classmate, she has been having a hard time.

I have a classmate who is always saying negative things about other people. **I'm really __red __ listening to her**. I'm afraid she is **__ting __se**, and I **have been m__ing to say** something to her about it. The other day I **was just __t to say** something to her when she asked me for advice about a problem she was having. She said she **__dn't ___lp saying** negative things about other people.	*(I don't like listening to her anymore)* *(saying more and more negative things about people)* *(have been thinking that I should say)* *(was on the point of saying)* *(was unable to stop herself from saying)*

Parents and marriage

In the following, the speaker talks about his parents. He is worried about how they will accept the news about his marriage.

My Mom and Dad want to know what I have **been ___p ___** each week. Usually we talk about my part-time job, my friends, my studies and **___t ___t of thing**. They usually think I make decisions without enough thought, which isn't true. So I want to be sure that they won't **___t it ____g** this time. I am going to tell them that I am getting married next week! I am sure that they will tell me to wait until after I finish my college education. But she comes from a very rich family, so they don't need to worry that I **___n't ____d to do this**. And they **___d ___** have a rich son than a poor one!	*(been doing and thinking about)* *(things like these)* *(make a mistake about what I am saying)* *(am not able to do this because I don't have enough money) (prefer to)*

This is the end of the test. Close this booklet and follow the instructions.

Appendix C

Notes regarding the cases where spelling hints in the P-section were prepared in specific ways.

No.	Entry on the EVP website	No. of keywords	Keyword(s)	Test item	Notes on the creation and revision of items
1	keep doing sth	1	keep	___p repeating	
2	again and again	3	again and again	a____ and ____n	In the pilot, one native speaker wrote *on and on* for the draft item ____*n and* ____*n*; the hints were adjusted accordingly.
3	be up to sb	2	up to	is __p ___ you	
4	be willing (to do sth)	1	willing	be wi____g to look for	In the pilot, one native speaker wrote *trying* for the draft item _____*g*; the hints were adjusted accordingly.
5	after all	2	after all	a___r a__	In the pilot, two native speakers wrote *for ever* and *for you* respectively for the draft item ____*r* __; the hints were adjusted accordingly.
6	get to know sb/sth	3	get to know	___t to kn___	In the pilot, one native speaker did not write anything for the draft item __*t to* ___*w;* the hints were adjusted accordingly.
7	have sth in common	3	have, in common	we had a lot ___ _____	For a three-keyword phrase with a sandwiched filler (*sth*), both the last two keywords were made complete blanks.
8	in advance	2	in advance	__ _____ce	(1) As the first word (*in*) is a preposition, it was totally blanked instead of the second, content word (*advance*). (2) In the pilot, one native speaker wrote *at once* for the draft item __ _____ *ce*; however, no change was made.

No.	Entry on the EVP website	No. of keywords	Keyword(s)	Test item	Notes on the creation and revision of items
9	at first	2	at first	__t ____t	(1) As the first word (*at*) is a preposition, it was totally blanked instead of the second, content word (*first*). (2) In the pilot, one native speaker wrote *I admit* for the draft item __ ____*t*; the hints were adjusted accordingly.
10	ever since	2	ever since	e___ s____	In the pilot, one native speaker wrote *after that* for the draft item ____*r* ____; the hints were adjusted accordingly.
11	get cold/ill/late, etc.	1	get	__ting thinner	
12	as you know	3	as you know	__s you ___w	
13	so far	2	so far	S__ ___r	In the pilot, two native speakers wrote *To date* and *To now* respectively for the draft item __o ___; the hints were adjusted accordingly.
14	no way	2	no way	There is ___ ___y	If the second word (*way*) were totally blanked, there could be an alternative word to fill (*chance*); the final letter was therefore left.
15	as long as	3	as long as	__s ___g as you eat	
16	I bet (you)	2	I bet	I ___t	As the first word is a pronoun, it was kept as is, and the second word was replaced by a blank with the final letter remaining.
17	it/that depends	2	it depends	It _____ds	As the first word is a pronoun, it was kept as is, and the second word was replaced with a blank with its final letters remaining.
18	fall asleep	2	fall asleep	__ell _____	
19	tell sb how/ what/when to do sth	4	tell, what to do	told her what ___ ___	
20	(just) in case	3	just in case	___t in ___se	
21	if I were you	4	if I were you	__ I ____ you	
22	for fun or for the fun of it	5	for the fun of it	for the ___n ___ it	With the paraphrase created for this phrase (*because it will be enjoyable*), it was expected that the word *fun* would be too difficult, if there were no hint, for test-takers to come up with; the final letter was therefore left.
23	things like that	3	things like that	____gs ___ke that	
24	be worth sth/ doing sth	1	worth	It is ____th the price	
25	miss a chance/ opportunity	3	miss a chance/ opportunity	m___ a ____ce to go	In the pilot, one native speaker wrote *pass (up) a chance to go* for the draft item __ss a ____ce to go; the hints were adjusted accordingly.
26	change your mind	3	change your mind	you've ____ged your ___d	
27	at the same time	4	at the same time	at the ____ ____	

No.	Entry on the EVP website	No. of keywords	Keyword(s)	Test item	Notes on the creation and revision of items
28	not really	2	not really	___t ____lly	If the second word (*really*) were totally blanked, there could be an alternative word to fill (e.g., *most*); the final letters were therefore left.
29	feel like/as if	2	feel like	I ___el ___ke	If the second word (*like*) were totally blanked, there could be an alternative word to fill (*that*); the final letters were therefore left.
30	on purpose	2	on purpose	___ _____se	As the first word (*on*) is a preposition, it was totally blanked instead of the second, content word (*purpose*).
31	get rid of sth	3	get rid of	___t ___d of this TV	
32	get/become used to sb/sth/ doing sth	3	get used to	___t __sed to it	
33	be supposed to do sth	1	supposed	were _____sed to be	
34	go badly/well, etc.	2	go well	Did your English test __o w____	In the pilot, one native speaker wrote *go well/okay/fine* for the draft item *__o ____*; the hints were adjusted accordingly.
35	What if …?	2	what if	____t ___	
36	wait a minute	3	wait a minute	____t a ____te	
37	go wrong	2	go wrong	___ne ____g	If the second word (*wrong*) were totally blanked, there could be an alternative word to fill (*bad*); the final letter was therefore left.
38	make sb do sth	1	make	Don't ___ke me do	
39	in time	2	in time	i_ ___me for the test	In the pilot, all three native speakers wrote *on time for the test* for the draft item *_ ___me for the test*; the hints were adjusted accordingly.
40	not until	2	not until	it's ___ going to start ___il	If the second word (*until*) were totally blanked, there could be an alternative word to fill (*till*); the final letters were therefore left.
41	keep sb waiting	2	keep, waiting	____p everyone _____	
42	up to 10, 20, etc.	2	up to	___p ___	
43	get down/into/ off, etc.	2	get off	___t ____	
44	either way	2	either way	_____r ___	
45	take a break/ rest, etc.	3	take a break	___ke a ____k	
46	feel bad about sth/doing sth	3	feel bad about	__el __d about	
47	feel sorry for	3	feel sorry for	___ling ____y for	
48	take advantage of sth	3	take advantage of	___e a_____ge of	In the pilot, one native speaker wrote *take charge of* for the draft item *__ke _____ge of*; the hints were adjusted accordingly.

No.	Entry on the EVP website	No. of keywords	Keyword(s)	Test item	Notes on the creation and revision of items
49	do badly/well	2	do well	__o w____	In the pilot, one native speaker wrote *do well/ great* for the draft item __o ____; the hints were adjusted accordingly.
50	up to date	3	up to date	be __p to ___te	
51	tired of doing sth	2	tired of	I'm really __red __ listening to her	
52	get worse	2	get worse	___ting ____se	With the paraphrase created for this phrase (*saying more and more negative things about people*), it was expected that the word *worse* would be too difficult, if there were no hint, for test-takers to come up with; the final letters were therefore left.
53	have been meaning to do sth	1	meaning	have been m____ing to say	In the pilot, one native speaker wrote *meaning/ planning* for the draft item ___ning; the hints were adjusted accordingly.
54	be just about to do sth	1	about	was just ___t to say	
55	can't/couldn't help doing sth	2	couldn't help	___dn't ___lp saying	(1) For the first word (*couldn't*), the third letter from the last (*d*) was left because tense was not tested. (2) If the second word (*help*) were totally blanked, there could be an alternative word to fill (*stop*); the final letters were therefore left.
56	be up to sth	2	up to	been __p ___	
57	that sort of thing	4	that sort of thing	___t ___t of thing	If both *that* and *sort* were totally blanked, an alternative phrase (*this kind*) would become possible; both words' final letter was therefore left.
58	get sth wrong	2	get, wrong	___t it ____g	With the paraphrase created for this phrase (*make a mistake about what I am saying*), it was expected that the word *wrong* would be too difficult, if there were no hint, for test-takers to come up with; the final letter was therefore left.
59	can afford	2	can afford	___n't _____d to do this	In the pilot, one native speaker wrote *don't want* for the draft item __n't _____; the hints were adjusted accordingly.
60	would rather	2	would rather	_____d _____	

Interpretation of the CEFR Companion Volume for developing rating scales in Cuban higher education

Claudia Harsch, University of Bremen

Ivonne de la Caridad Collada Peña, University of Informatics Sciences, Havana

Tamara Gutiérrez Baffil, University of Pinar del Río

Pedro Castro Álvarez, University of Informatics Sciences, Havana

Ioani García Fernández, University of Cienfuegos

Considering the need for improving English proficiency among Cuban university graduates, the Ministry of Higher Education (MES) implemented a new policy for teaching, learning and assessment of English proficiency. The policy adopted the CEFR (Council of Europe [CoE] 2001) as a proficiency framework, with the level B1 as the targeted attainment level. The CEFR needed to be adapted to suit the local context while operating within an internationally recognised framework. In 2017, the development of a valid and reliable proficiency exam was initiated. This work has been carried out by a network of Cuban teachers of English within the MES, coordinated by the University of Informatics Sciences, Havana, in collaboration with the University of Bremen, Germany. This article is a practice report of the process of developing rating scales for writing as part of the new exam. We explore the feasibility of using the CEFR Companion Volume (CEFR/CV; CoE 2018) descriptors as a basis for developing localised rating scales. Moreover, we describe the challenges faced during the process, which included creating more specific descriptors for the CEFR 'plus' levels (CoE 2001: 32, 181). Our insights show how the CEFR/CV (CoE 2018) descriptors can be adapted and how adaptation challenges can be overcome.

Keywords: rating scales, CEFR-based assessment, standardised testing, descriptor development, adaptation of the CEFR/CV for rating purposes

1 Introduction and background

The Ministry of Higher Education in Cuba (MES) introduced a national policy for English education in 2015 that considered the CEFR (CoE 2001) as the main proficiency framework, with B1 as the target level for university exit requirement. The attainment level (B1) was selected for a variety of reasons:

- the low proficiency level displayed by the majority of the new enrolments at university level who, in spite of the efforts made by Cuban general education, complete upper secondary education with poor English skills;
- the limited number of hours allotted to English in university undergraduate curricula, which makes it impossible to go beyond level B1 if "Below A1" is the starting point for many students; and finally,

- the fact that B1, as the "Threshold level" (CoE 2001: 34), in which language learners have acquired the beginning of an independence as users of the language, is the lowest level for university graduates to be able to start their professional lives with a possibility of continuing their training in English for academic and professional purposes through postgraduate education. This is considered a temporary phase since general education is also developing an improvement policy and will eventually upgrade the exit level of upper secondary schools.

One of the main issues when starting to implement the new policy was the lack of a proficiency exam available for certifying the exit requirement, given the impossibility of financial means to access international tests due to the budgetary and free nature of the education system in Cuba, which is subsidised by the state. In order to develop such an exam for Cuban higher education, a project was implemented in July 2017, the main goal of which is to develop a teaching and certification system for English so that Cuban language centres can reliably and validly certify students' English proficiency. The certification aims at international recognition through alignment to the CEFR (CoE 2001) proficiency levels. Partners in the endeavour are the MES, the University of Informatics Sciences (UCI) representing all Cuban universities, the University of Bremen, Germany, and the VLIR ICT for Development Network University Cooperation Program[1]. This way, we bring together local and global expertise to reflect local requirements while striving to adhere to international standards. The project included setting up a network of representatives of all Cuban universities (Cuban Language Assessment Network in Higher Education, abbreviated CLAN, which is part of LAALTA[2]).

The project encompasses three important objectives: first, developing assessment literacy among the CLAN teachers and preparing them for cascading this literacy in all universities; second, the development, validation, and implementation of the exam through a sustainable system; and third, research on assessment to support the first and second objectives.

So far, the first and second objectives have been addressed by means of six workshops, during which training and hands-on sessions for test development were provided, using the CEFR/CV (CoE 2018) as a framework for the CLAN members. In the workshops, all areas of assessment literacy in theory and practice have been covered. The members have had online working phases after each workshop, where they have collaboratively developed assessment materials and received feedback from each other and from the international trainers. The CLAN members have also been cascading their knowledge to other teachers in their institutions. Outcomes obtained so far include test specifications and item writer guidelines for the skills of listening, reading, speaking, and writing, along with the development of a bank of tasks for the four skills, as well as interlocutor guides for speaking. Based on this work, a small group of seven researchers (five of whom are the authors of this article) undertook the initial drafting of the rating scales for writing. In the next phases, speaking will be addressed, and the CLAN members will contribute to further refining the scales, following Harsch and Martin's (2012) development and validation approach, and Holzknecht et al.'s (2018) as well as Harsch and Seyferth's (2019) approach of involving teachers in developing tests.

This progress report describes the process of the initial rating scale development, focusing on the applicability of the CEFR/CV as a cornerstone, with a specific focus on the challenges faced and how we addressed them.

2 Rating scales development

Before outlining the actual development process, we will describe the basis, i.e. the assessment criteria and levels defined in the test specifications, as well as task characteristics that are relevant for the rating scale development.

---

1. The project was also supported financially by the British Council Cuba and UK, and ILTA.
2. LAALTA: Latin American Association of Language Testing and Assessment.

2.1 Basis: levels and criteria

The targeted level of the final exam is B1, as explained above; yet, in the first years, the exam should allow certifying students who can only demonstrate an A2 level. As part of the change management in the initial stage of the implementation of the policy, the Ministry decided to accept level A2 as exit requirement for a transitional period (2015-2021), until universities have been able to adjust to the new policy by creating all necessary human and material resources.

That is why the rating scale encompasses descriptors from A1+ to B1+. The decision to incorporate the so-called 'plus levels' in the scale is derived from the fact that the CEFR (CoE 2001) criterion levels (i.e. the main six levels) are too broad (Deygers and Van Gorp 2013: 4; Fulcher 2004: 258-259; Martyniuk and Noijons 2007: 6), even more so considering the narrow range of levels targeted in this project. We thus followed the CEFR's "branching approach" which suggests "cut[ting descriptors] into practical local levels" (CoE 2001: 32), i.e. adjusting the number of level subdivisions and hence the CEFR descriptors defining these sublevels to local needs. This way, we took into account the local context: teachers in Cuban higher education lack experience in working with analytic scales that span several levels. Accordingly, we introduced the plus levels A1+, A2+ and B1+ in order to provide more guidance and precision without making the scale too granular.

In the test specifications, the CLAN members defined the targeted skills, task characteristics, expected attributes of student performances, and an initial version of relevant assessment criteria. In order to decide which criteria to choose for rating written performances, the members considered the terms and concepts that teachers have traditionally used in Cuban teaching practice, in order to minimise the negative impact of change resistance amongst teachers when introducing the new system. The following criteria for assessing writing skills emerged:

- task fulfilment (TF, for interactive and productive tasks)
- coherence and cohesion (CC)
- vocabulary (VO, covering range and appropriateness)
- grammar (GR, covering range and accuracy)
- orthography (OR, covering spelling and mechanics).

We adapted the categorisations of the CEFR/CV to our local needs. With regard to the categories of interaction and production, for instance, we followed the CEFR/CV differentiation and developed productive and interactive writing tasks. Each exam includes one interactive and one productive task. These two aspects are also reflected in the rating scale category of task fulfilment, as will be explained in more detail below.

2.2 Methodology

The approach taken for development, validation and revision of the rating scales is an iterative one (Piccardo et al. 2019: 28), which was modelled on the research reported by Harsch and Martin (2012) and Harsch and Seyferth (2019). We are employing intuitive, qualitative and quantitative stages (CoE 2001; Fulcher, Davidson and Kemp 2011). Intuitive methods refer to approaches that "do not require any structured data collection, just the principled interpretation of experience", as the CEFR states (CoE 2001: 208).

We took existing descriptors, i.e. relevant descriptors from the CEFR/CV and from assessment scales in the context of CEFR-aligned exams, as a starting point. During the initial intuitive phase, a group of seven researchers selected existing descriptors for the targeted criteria/levels and then adapted formulations to avoid repetition or vagueness and to account for the local context (i.e. teaching styles, most common mistakes, as well as positive and negative transfer from native language).

The researchers tried out the initial scales with a few student samples, discussed reasons for digressions and revised the wording of the descriptors accordingly. In the next phase[3], the descriptors of the scale drafts will be qualitatively sorted into their targeted levels/criteria by the CLAN members. Then, the members will try the scales with student samples in a combined training and trial approach, in which qualitative and quantitative data will be analysed. Again, reasons for digressions will be discussed and descriptors adapted where necessary.

The focus of this contribution lies on the initial intuitive phase, as the main work with the CEFR/CV (CoE 2018), its adaptations and descriptor revisions took place during this phase.

2.3 Working with the CEFR/CV during the intuitive phase

The starting point for the rating scale development was the proficiency descriptors and the additional materials in the appendix of the CEFR/CV. Other scales consulted were the Aptis Speaking rating scale (O'Sullivan and Dunlea 2015), the IELTS speaking and writing band descriptors (IELTS 2013; IELTS 2016; and IELTS 2018) and the Pearson Global Scale of English Learning Objectives for Academic English (Pearson Education 2015). These scales were chosen because they have been widely valued and consulted by most of the faculty bodies in Cuban universities since the new policy was introduced. Appendix A shows the final draft of the rating scale (all sources are color-coded), with which we will go into training and validation with the CLAN members.

2.4 Compiling existing descriptors

In a first step, we considered the writing assessment grid in the CEFR/CV (CoE 2018: 173-174), which includes the following categories: Overall, Range, Coherence, Accuracy, Description and Argument. This categorisation, however, does not match our assessment criteria (see above). Hence, we selected relevant descriptors from the grid but placed them into the best fitting criterion in our assessment criteria system. As we do not use an Overall criterion in our analytic approach, we dropped this category. Instead, we focused on the criterion TF with a close reference to our test specifications and task demands; here, we mostly added our own descriptors regarding the message conveyed, the relevance of ideas, the language functions performed and genre requirements, as well as register and politeness conventions. The CEFR/CV scale on socio-linguistic appropriateness (CoE 2018: 138) contains some descriptors that we included (see appendix A, phrases in red); we also selected some of the IELTS (IELTS 2013 and 2016) descriptors (phrases in green in Appendix A). We dropped the CEFR/CV assessment grid categories of Description and Argument (CoE 2018: 173) since their content is already included in our TF criterion. Furthermore, we consulted the CEFR/CV scales on productive and interactive writing (CoE 2018: 75-80; 93-102); while they had provided helpful input for the test specifications, we found their descriptors too generic and abstract to be directly used in the rating scale.

We used the CEFR/CV's assessment grid category of Coherence (CoE 2018: 173) but inserted descriptors from the CEFR/CV scale Coherence and Cohesion (2018: 142) as well as our own additions regarding organisation, sequencing and topic progression. With regard to the CEFR/CV grids' categories of Range and Accuracy (CoE 2018: 173), we followed the local tradition in Cuba, i.e. treating them as sub-aspects of the wider categories of grammar and vocabulary, which was also laid down in the test specifications. Hence, we arranged the aspects of linguistic range and accuracy under our criteria Vocabulary (VO) and Grammar (GR). In VO, the term *accuracy* was replaced by *appropriateness*, to account for terminological use in the Cuban context, i.e. teachers here would regard students' vocabulary choice as a matter of socio-linguistic appropriateness rather than accuracy, which is strongly associated with grammar. For VO and GR, we also used the CEFR/CV scales Vocabulary Range, Grammatical Accuracy and Vocabulary

---

3. This work actually took place in a workshop in February 2020, just after the deadline for this article. We will publish the results elsewhere.

Control (2018: 132-134), as well as the occasional IELTS descriptor wording (IELTS 2013 and 2016). For our criterion Orthography (OR), we used the CEFR/CV scale Orthographic Control (2018: 137) and some IELTS descriptor wordings (IELTS 2013 and 2016). For all our criteria, we added statements on how to treat errors (in *italics*); following Harsch and Martin's insights (2012). These statements are intended to further guide the raters, because teachers in Cuba are traditionally used to focusing on error correction.

One of the challenges we found was adapting existing descriptors to the local context (see Appendix A where we color-coded all the different sources as well as the adaptations we undertook). Another major challenge was to describe the plus levels, as the CEFR/CV scales do not consistently provide them. Thus, we had to compare the existing descriptors of the CEFR criterion levels and formulate descriptors that would enable enough differentiation between them. We will discuss below (in Section 3) a detailed example of these challenges and how we overcame them.

2.5 Pre-trial

In the informal pre-trial, the researchers/authors used the initial rating scale drafts for the analysis of three student performances, each for an interactive and a productive writing task. The performances were elicited informally in the classroom by one of the researchers, who is also an active language teacher. The aim of the pre-trial was to evaluate the usability of the descriptors: they were evaluated for "clarity, [context-related] pedagogical usefulness" (North and Docherty 2016: 25), possibilities for constructive alignment and practicality, as well as consistency across the levels and the assessment criteria. In the pre-trial, we compared students' performances with the descriptors in the rating scale (Pollitt and Murray 1996) to place performances at levels, and we qualitatively discussed digressions and underlying reasons; i.e. we each gave explanations of our decisions, justified reasons why we placed a performance at a certain level and exchanged our justifications. After careful considerations of the different viewpoints, and careful re-analysis of student performance and descriptor wording, we revised the descriptors where necessary. Appendix A shows these revisions in blue; all deletions indicated in Appendix A also took place after this pre-trial. Most revisions happened in the criteria OR and GR, some in CC, and a few in VO.

We will use this draft of our rating scale for the next qualitative phase (see Section 4 below).

3 Discussion of the challenges with the CEFR/CV

We will now summarise the main challenges we faced and how we dealt with them when using the CEFR/CV and its proficiency scales/descriptors for developing rating scales.

Abundance of scales at different places: We found the fact that the CEFR/CV contains a wealth of scales for the productive/interactive skills, strategies and linguistic competences that may be quite overwhelming. This was exacerbated by the challenge of locating relevant scales (including the writing assessment grid in the Appendix) at different places in the CEFR/CV during the actual work with the CEFR/CV[4]. Appendix B gives an overview of the scales we consulted and their location in the CEFR/CV. Even when simultaneously working on several laptops, it was a constant search for relevant descriptors and scales. Here we would recommend a searchable online data bank of all CEFR/CV descriptors, where relevant ones could be compiled (along with a transparent source reference) to facilitate working with the CEFR/CV.

Different categorisations: As described above, the categorisations in the Writing Assessment Grid and other CEFR/CV scales differed from our assessment criteria. Moreover, the CEFR/CV's assessment grid categorisation also differs from the CEFR/CV scale system: the Assessment Grid differentiates range, coherence, accuracy, description and argument, while the CV scale system shows a much wider differentiation of language activities (written production, of which description and argument are sub-aspects, and written interaction, as well as strategies) and linguistic competencies (which subsume

---

4.   One has to bear in mind that it is difficult in Cuba to print such large documents as the CEFR/CV.

range, accuracy and coherence, amongst many more aspects which are not covered in the Assessment Grid). This may be a natural phenomenon given the complexity of the construct of communicative competence, yet it does pose a challenge when the task is to compile relevant descriptors for a given set of writing assessment criteria.

Plus levels not always provided: Not all CEFR/CV scales consistently describe the plus levels. It proved difficult to develop suitable descriptors for these levels. We will provide an example in the next paragraph. It would also help to analyse actual student performances to fill the plus levels appropriately. We are planning to address this issue in the next step when we have a solid basis of student performances.

Inconsistent wording across scales and/or across levels: We found that some scales/materials (at different places) in the CEFR/CV address similar aspects but use different wording in descriptors that target the same level. Some descriptors (that appear in different scales) contain aspects that seem incoherent when comparing these aspects across different scales and levels. It was challenging to reach consistent interpretations of a given aspect (such as the nature and impact of errors) within one level and across the levels when comparing different scales (e.g., Grammatical Accuracy, Vocabulary Control, and the Writing Assessment Grid: Accuracy). Compare the following examples:

- CEFR/CV scale Grammatical Accuracy (2018: 133) states for level A2: "… still systematically makes basic mistakes …; nevertheless, it is **clear** what he/she is trying to say"; for level B1+: "Errors occur, but it is clear what he/she is trying to express." This aspect is not mentioned at B1, and there is no A2+ descriptor.
- CEFR/CV scale Vocabulary Control (2018: 134): no mention of the aspect of clarity of expression.
- CEFR/CV Writing Assessment Grid (2018: 174), criterion Accuracy, level A2: "… errors may sometimes cause **misunderstandings**"; level B1: "Occasionally makes errors that the reader usually can interpret correctly on the basis of the context." No plus levels are defined.

When comparing these statements, we found the aspects in bold (describing A2) contradictory (i.e. when there is a misunderstanding, it is *not* clear what one is trying to say). Moreover, we regarded the demand for clarity of what one wants to say too high for A2. When working on the target level A2+, we found it unfortunate that there are no A2+ descriptors in these scales. Our resolution was to make use of the IELTS (IELTS 2013 and 2016) band 4 descriptor[5]: "errors may cause strain on the reader" (IELTS 2013). We added this qualification at A2+ for our criteria VO, GR, and OR after it became clear in the pre-trial that we needed to qualify the kinds of errors we would expect and 'allow' at the different levels (for example, there are minor, non-impeding errors that are 'allowed' at B1+, while we would not expect systematic errors in basic sentence structures at this level; see the blue additions in Appendix A).

These issues were the main challenges we faced when working with the CEFR/CV (CoE 2018). In order to address these challenges, we resorted to different means, which can be summed up as follows:

- Reorganising CEFR/CV descriptors into the local assessment criteria.
- Adapting CEFR/CV descriptors (i.e. changing wording) to make levels coherent.
- Adding descriptors from other sources, particularly for the plus levels.
- Adding and adapting descriptors to account for the local context, both for criterion levels and plus levels.

4 Conclusions and outlook

Undoubtedly, the CEFR/CV provides a rich and informative source and starting point for rating scale development. Yet, one has to take into account the complexity of the CEFR/CV, its limitations and the

---

5. IELTS band 4 is actually targeting B1, which again seems in contradiction to the CEFR/CV descriptors on clarity of expression in the Grammatical Accuracy scale.

requirements of the local context. Each descriptor in the CEFR/CV has to be checked against the local test specifications (i.e. to see whether its content matches the content of the test specifications) and adapted accordingly to fit the local requirements. When adapting descriptors or writing additional ones, it is important to consult local experts and to take additional sources into consideration, such as assessment scales from other exams that are aligned to the CEFR. Particularly when the local rating scales require a finer granularity than the CEFR criterion levels, measures need to be taken to fill the plus levels with appropriate descriptors.

Based on our experiences with the CEFR/CV descriptors, we found it challenging to deal with the abundance of scales in the CEFR/CV, with differing categorisations across the CEFR/CV, with inconsistent wording within and across scales and levels, and with the fact that plus levels are not always provided. In order to overcome these challenges and to account for the local context, we reorganised CEFR/CV descriptors into our local assessment criteria, adapted CEFR/descriptors for more coherence, and added descriptors from other sources for the plus levels.

Any rating scale development is an iterative process with several rounds of revisions. It is advisable to use different methods to gain information on the validity and applicability of the new scale. In our case, we have covered the initial intuitive phase, using experts to compile, draft and trial the first version, leading to the first round of revisions. With the thus revised rating scales, we are entering the next phase, which includes a qualitative sorting exercise, i.e., the CLAN members will sort the descriptors into levels/criteria in order to validate the content and levels of the rating scales. Then, a benchmarking exercise will follow where the CLAN members will be trained to use the scales so that they can pass on this knowledge to their colleagues and roll out the new assessment approach at a national level.

5 References

Council of Europe. 2001. The Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Language Policy Division. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989 (accessed 27 November 2019).

Deygers, Bart & Koen Van Gorp. 2013. The influence of the CEFR on rating scale design. In: Jozef Colpaert, Mathea Simons, Ann Aerts, Margret Oberhofer (eds.) *Book of Abstracts, "Language Testing in Europe: Time for a New Framework?"*, Antwerp: University of Antwerp., 57-59. https://www.academia.edu/3549147/The_influence_of_the_CEFR_on_rating_scale_design (accessed 12 February 2020).

Fulcher, Glenn. 2004. Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly 1*(4), 253-266.

Fulcher, Glenn, Fred Davidson & Jenny Kemp. 2011. Effective rating scale development for speaking tests: Performance decision trees. *Language Testing 28* (1), 5-29.

Harsch, Claudia & Guido Martin. 2012: Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. In: *Assessing Writing* 17, 228–250.

Harsch, Claudia & Sibylle Seyferth. 2019: Marrying achievement with proficiency - developing and validating a local CEFR-based writing checklist. *Assessing Writing*. Online first: https://www.sciencedirect.com/science/article/pii/S1075293519301746 (accessed 20 March 2020).

Holzknecht, Franz, Benjamin Kremmel, Carmen Konzett, Kathrin Eberharter, Eva Konrad & Carol Spöttl. 2018. Potentials and challenges of teacher involvement in rating scale design for high-stakes exams. In Xerri, D., & Vella Briffa, P. (eds.), *Teacher involvement in high stakes language testing* (pp. 47-66). Springer.

IELTS. 2013. *IELTS TASK 1 Writing band descriptors* (public version). https://takeielts.britishcouncil.org/sites/default/files/ielts_task_1_writing_band_descriptors.pdf (accessed 12 February 2020).

IELTS 2016. *IELTS TASK 2 Writing band descriptors* (public version). https://takeielts.britishcouncil.org/sites/default/files/ielts_task_2_writing_band_descriptors.pdf (accessed 12 February 2020).

IELTS 2018. *IELTS Speaking: band descriptors* (public version). https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en (accessed 12 February 2020).

Martyniuk, Waldemar & José Noijons. 2007. *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe Member States*. Strasbourg: Council of Europe.

North, Brian and Coreen Docherty. 2016. Validating a set of CEFR illustrative descriptors for mediation. *Research Notes*, *63*, 24-30.

O'Sullivan, Barry & Jamie Dunlea. 2015. *Aptis General technical manual version 1.0*. Technical Report TR/2015/005. British Council: London.

Pearson Education. 2015. *Global Scale of English Learning Objectives for Academic English*. http://www.b-li.ir/ar/06CEFR/4.GSE_LO_Academic_English.pdf (accessed 12 February 2020).

Piccardo, Enrica, Brian North & Tim Goodier. 2019. Broadening the scope of language education: Mediation, plurilingualism, and collaborative learning: The CEFR companion volume. *Journal of e-Learning and Knowledge Society*, *15*(1).

Pollitt, Alastair & Neil Murray. 1996. What raters really pay attention to. *Studies in language testing*, *3*, 74-9.

6 Biographies

Claudia Harsch is a full professor at the University of Bremen, specialising in language learning, teaching and assessment. She is also the Director of the Languages Centre of the universities in the Land Bremen. She has worked in Germany and in the UK and is active in teacher training worldwide. Her research interests focus on areas such as language assessment, language and migration, intercultural communication and the implementation of the CEFR. Claudia was president of the European Association of Language Testing and Assessment from 2016-2019.

Ivonne de la Caridad Collada Peña has been a teacher and teacher trainer for more than 30 years and has a master's degree in management. She has been the head of English departments, faculty dean, and is currently the director of the Language Centre at UCI. She has taught courses on linguistics, lexicology, ESP and EFL, among other subjects. Her research interests include TESOL methods, curriculum development, lexicology, grammar and language teaching and assessment. She is currently the Cuban coordinator of the project that is presented in this contribution.

Tamara Gutiérrez Baffil has been teaching for 22 years in higher education. She has a master Degree in Educational Science. Associate Professor and Language Centre Director. Currently researching on Language Centre Management as PhD dissertation and other searches in CEFR application in the Cuban context. Granted an International Diploma in Educational Planning and Administration and a given a Professional award in L2 teacher training. She has published in several journals. Teaches ESP, EAP, EPP. Participates in international and national scientific projects.

Pedro Castro Álvarez is a full professor at the University of Informatics Sciences, Havana, Cuba, where he has been deputy director of the Language Center. He has a doctorate in pedagogy from the Universidad Central de Las Villas in 2006 and a master's degree from the same institution in 2000. The professor has a 28-year experience as a teacher in Cuban higher education and has also taught in foreign universities. His research is related to the fields of educational technologies, curricular development and language assessment. He has published several journal articles and attended multiple conferences in Cuba and abroad.

Ioani García Fernández has been teaching for 18 years, 16 in Higher Education. She is an associate

professor and has a Master's Degree in Education from the University of Cienfuegos. She was certified with the Teaching Knowledge Test (TKT) and also received a Professional Award in L2 Teacher Training. She has participated in several international scientific events and published various articles related to the teaching of English as a Foreign Language. She is currently working as a professor at the Language Centre of the University of Cienfuegos.

Appendix A
Initial draft rating scale writing, after first trial

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
B1+	The message is clearly and appropriately conveyed. (CLAN) All ideas/content are relevant to the topic of the task (CLAN) Performs all the language functions required by the task (e.g., comparing, describing, explaining, justifying etc.) (Test specs page 8 and adapted from CV page 138). ~~Shows the required length.~~ Follows the conventions of the text type required by the task (CLAN). Uses an appropriate register (adapted from CV page 138) Shows salient politeness conventions (adapted from CV 138)	Uses a meaningful sequence of linked ideas, with adequate topic progression (TS, GE). Makes logical paragraph breaks, if required by task. (adapted CV p. 142) Uses various cohesive devices to establish cohesion throughout the text. (CLAN) Establishes more complex relations between ideas, e.g., Can introduce a counter-argument with 'however', cause and consequence, cause and effect (adapted form CV p. 142).	Uses a good range of topic-specific vocabulary related to the task (CV p 132-174). Uses vocabulary with reasonable precision. (adapted from CV page131) *May show occasional inaccurate word choices and collocations (adapted from IELTS band 7 and 8).* *Errors may occur when expressing more complex thoughts. (adapted CV 134)*	Uses a good range of simple structures and features with generally good control *though mother tongue influence may be noticeable.* Shows some complex grammatical features and syntactical structures, although not always correctly. *Errors may occur, but it is clear what he/she is trying to express* (CV p 133).	Spelling is accurate enough to not strain the reader. Punctuation generally follows conventions. *Spelling and punctuation may show mother tongue influence.* (adapted from CV 137).

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
B1	The message is generally clearly conveyed. (CLAN) The ideas/content are generally relevant to the topic of the task. (CLAN) Performs most of the language functions required by the task (e.g., comparing, describing, explaining, etc.) (Test specs page 8 and adapted from CV page 138). Shows the required length. Mostly follows the conventions of the text type/format required by the task (CLAN), *but the format may be inappropriate in places* (IELTS band 5). Shows awareness of the required register, *but may still be inconsistent in tone* (IELTS band 6). Generally follows salient politeness conventions, but not always appropriately (adapted from CV 138)	Mostly organizes ideas into a meaningful sequence, with adequate topic progression (TS, GE). *May occasionally use unrelated or off-topic ideas (CLAN).* Makes simple, logical paragraph breaks if required by task. (adapted CV p. 142) Links a series of shorter, discrete simple elements into a connected, linear sequence of points by using a limited number of cohesive devices (adapted CV p. 142)	Uses sufficient topic-specific vocabulary to express themselves on familiar topics. (CV page 132) Shows appropriate use of a wide range of simple basic, frequent vocabulary. (adapted from CV page 134) *Major errors may still occur when expressing more complex thoughts.* (CV page 134) *May use circumlocution and occasionally unclear expressions.* (adapted from CV page 131, 174)	Uses a range of simple grammatical features and sentence structures with reasonable accuracy. (adapted CV p. 133) Attempts a limited range of complex sentence structures or complex grammatical features, *though they may usually be incorrect.* (adapted from IELTS band 5) *In general*, the reader can interpret the errors correctly based on the context. (adapted from CV p. 174)	Produces generally intelligible spelling for most common words, *mother tongue influence is likely with less common words.* Spelling, Punctuation is and layout are accurate enough to be followed most of the time, but *mother tongue is likely to influence punctuation.* (adapted from CV p. 137)
A2+	The message gets across but with some limitations. In general, the ideas/content are related to the topic of the task. (CLAN) Performs basic language functions required by the task (e.g., describing, explaining, narrating); *may attempt the more complex ones, but not always successfully (e.g., comparing/contrasting ideas)* (Test specs p. 8 and adapted from CV p. 138). *May use an inappropriate format* (adapted from IELTS band 4). *May use an inappropriate tone (adapted from IELTS Band 4).*	Shows some organization of ideas and a clear attempt at topic progression (TS). *May still show some limitations in sequencing and text structure. also off-topic ideas (CLAN)* *Paragraph breaks may be missing.* Uses the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points (CV p 142). *May use less frequent cohesive devices inappropriately. (CLAN)*	Uses basic, frequent vocabulary to express themselves in routine everyday situations (CV p. 132). Shows inaccuracies in word choice and collocation *that may occasionally cause strain for the reader.* (CLAN and adapted from *IELTS)* May *have to compromise the message and may use repetitions and circumlocutions (adapted from CV* 131 and CLAN).	Uses simple sentence structures and basic grammatical features (such as present perfect, continuous forms, modals). Systematic mistakes may still occur; *errors may sometimes cause strain on the reader* (adapted from IELTS Band 4), *but it is usually clear what s/he is trying to say.* (adapted from CV p. 133, 174). *May show attempts at more complex structures, but usually these are erroneous.*	Writes with reasonable phonetic accuracy, *but mother tongue is likely to be noticeable.* Punctuation is still *likely to be influenced by mother tongue.* (adapted from CV p. 137). *Errors may cause occasional strain on the reader. (adapted from IELTS band 4)*

	Task Fulfilment	Coherence/cohesion	Vocabulary (range and appropriateness)	Grammar (range and accuracy)	Orthography (spelling and mechanics)
A2	The message gets across but with some strain on the reader. The ideas/content are *not necessarily all* related to the topic of the task. (CLAN) Performs the more concrete language functions required by the task (e.g., social exchanges, invitations etc.). (Test specs p. 8). *Generally, the format may not yet be appropriate* (adapted from IELTS band 4). Apart from everyday polite forms of greeting and address, the tone *may be inappropriate* (adapted from CV page 138 and IELTS band 4).	~~Makes an attempt at organization and topic progression~~ (TS). Produces a list of points that are mostly in a logical sequence; not all are necessarily connected. *May show limitations in sequencing and text structure, ~~also off-topic ideas~~ (CLAN)* Links groups of words with simple connectors like 'and', 'but' and 'because' (CV p 142). *May overuse connectors, may use other cohesive devices unsuccessfully. (CLAN)*	Shows ~~sufficient~~ limited basic vocabulary and memorized phrases to express basic communicative needs and to communicate limited information (adapted from CV p. 132 and 174). Shows frequent inaccuracies in word choice and collocation *that may cause strain for the reader.* (CLAN and adapted from IELTS)	Shows simple sentence structures, with memorized grammatical phrases and formulae. Still systematically makes basic grammar and syntax mistakes – for example tends to mix up tenses and forget to mark agreement, *which the reader may misunderstand* (adapted from CV p. 133, 174).	Writes with reasonable phonetic accuracy the most common words, *but not necessarily following standard spelling.* (adapted from CV. p. 137) Uses punctuation such as full stop, commas, question marks, *but not necessarily accurately.* *Errors in spelling and punctuation may cause strain for the reader. (adapted from IELTS band 5)*
A1+	The message only partly gets across and usually requires a sympathetic reader. (CLAN) Shows awareness of the required topic but the ideas are very limited. (CLAN) Performs only the most concrete language functions (e.g., establish social contact) (CLAN, adapted CV 138) Format and tone are mostly inappropriate. (CLAN)	Links words or groups of words with very basic linear connectors like 'and' or '~~then~~ because' (CV p. 142). Texts longer than short notes and messages generally show coherence problems that make them very hard or impossible to understand. (adapted from CV p. 174).	Shows a very basic range of simple vocabulary and memorized expressions related to particular concrete situations (CV p. 131-132) *May overuse certain words (CLAN)*	Shows only a few simple grammatical features and sentence patterns in a learnt repertoire (CV p. 133). *Errors are likely to be frequent and common.* (CLAN)	Writes only familiar words and short phrases used regularly with reasonable accuracy. Spells his/her address, nationality and other personal details correctly. Uses only basic punctuation (full stops and question marks (adapted from CV. p. 137)

Notes: Sources used by colour code:

CEFR Companion volume/relevant scales and level | *IELTS band descriptors* | own additions CLAN and test specs | revisions after first trial in small group

Appendix B
Overview of relevant scales in CEFR/CV (CoE 2018)

Writing	Number of scales	Pages in the CEFR/CV
Production activities and strategies	5	75-80
Interaction activities and strategies	7	93-102
Communicative language competences	12	133-143
Appendix 4: Written assessment grid	2	173-174

Utilising pupils' plurilingual skills: a whole-school approach to language learning in a linguistically diverse Irish primary school

Déirdre Kirwan, former principal of Scoil Bhríde Cailíní

Since the mid-1990s, schools in many parts of Ireland have experienced an unprecedented increase in the level of linguistic and cultural diversity among pupils. This paper describes an innovative approach to integrated language learning that was developed in a primary school in West Dublin in response to this phenomenon. To ensure inclusion of all pupils and to support them in reaching their full potential, pupils' plurilingual repertoires are welcomed. Two overarching goals to language teaching and learning inform the whole-school language policy that seeks to:

- *ensure that **all** pupils become proficient[1] in the language of schooling*
- *exploit the linguistic diversity of the school for the benefit of **all** pupils (Council of Europe [CoE] 2001: 4; Garcia 2017: 18).*

Classroom procedures that facilitate inclusion of home languages in curriculum delivery and the needs of pupils who are endeavouring to learn English as an additional language are described. The importance of literacy is highlighted as is teacher, pupil, and parent cooperation. In addition to high levels of achievement in standardised tests of English and Maths, additional outcomes are identified including enhancement of the Irish language, a developing culture of learner autonomy, and the cultivation of pupil confidence and social cohesion.

Keywords: education, social cohesion, learner autonomy, linguistic diversity, inclusive, whole-school approach, plurilingual

1 Introduction

In the 1990s, unprecedented levels of immigration to Ireland resulted in major change to the linguistic landscape of primary and post-primary schools (Central Statistics Office 2017: 8; 46). This presented a major challenge for pupils, teachers and parents. As neither linguistic nor cultural diversity were issues addressed in teachers' pre- or in-service courses, there were few answers as to how pupils might best be served in multilingual educational milieu.

This article describes the innovative approach taken in response to this changed demographic by Scoil Bhríde Cailíní (SBC) – St. Brigid's School for Girls – a primary school in west Dublin.[2] In Ireland, primary education consists of an eight-year programme. Children are normally enrolled in Junior Infants in the September following their fourth birthday and progress to Senior Infants the following year. A further six years of primary education ensues. In English-medium schools, Irish is a compulsory subject and is taught from the beginning of schooling. In SBC, French is introduced in the penultimate year.

---

1. Proficiency here is not indicated in terms of a CEFR level. From an assessment perspective please see mention of standardised tests on p. 2.
2. A more detailed account of all these issues can be found in Little and Kirwan (2019) *Engaging with Linguistic Diversity: A Study of Educational Inclusion* in an Irish Primary School. Bloomsbury Academic.

In 1994, a young Bosnian speaker, whose family had been given refugee status, was registered in SBC along with 290 children who were English language speakers and native to the locality. Within 20 years, 80% of the then 322 pupils were from backgrounds where English was not the language of the home. Most of the 80% had little or no English when they started school and more than fifty languages had been identified in addition to English and Irish[3].

2 Whole-school language policy

In order to ensure that all pupils were fully included in the school and to support them in reaching their full potential, a whole-school language policy was formulated and endorsed by the Board of Management which includes parent representatives in its membership. Starting from the child-centred ethos of the Primary School Curriculum (Government of Ireland 1999), using the work of Integrate Ireland Language and Training (IILT 2006)[4], and qualitative research undertaken by the principal in the school year 2005-06 (Kirwan 2009), SBC developed an approach to language education that sought to include the plurilingual repertoires of all its pupils in the teaching and learning process (CoE 2001: 4-5). This approach is in accord with the human rights basis of the Council of Europe's language education policy, with particular reference to plurilingual education (Beacco & Byram 2007; Beacco et al. 2015).

Two overarching educational goals were agreed:

- To ensure that *all* pupils gain full access to education, which means helping them to become proficient in the language of schooling
- To exploit linguistic diversity for the benefit of all pupils by implementing an integrated approach to language education that embraces the language of schooling, languages of the curriculum (Irish and French), and home languages.

Four principles informed the policy. The first was an inclusive ethos that welcomed the diversity of the pupil population, acknowledging that each pupil had much to contribute to her own education. Second was an open language policy that encouraged use of home languages in class and throughout the school. Third was a strong emphasis on language awareness that involved drawing on home languages as a resource for all learners. Lastly, a strong emphasis was placed on literacy skills in English, Irish, French, and home languages, and on parental involvement in their children's literacy development.

2.1 Outcomes

Implementation of this policy has resulted in high levels of pupil achievement in English, Irish, French, and home languages (in the case of immigrant pupils). In standardised tests of Maths and English, the school regularly performs above the national average[5]. In 2014, following a whole-school evaluation by Department of Education inspectors, the school was judged to be in the highest category for the teaching and learning of Irish, a category in which only 12% of primary schools nationally are included

---

3. Afrikaans, Amharic, Arabic, Bangla, Benin, Bosnian, Cantonese, Dari, Cebuano, Estonian, Farsi Foula, French, German, Hebrew, Hindi, Hungarian, Igbo, Ilonggo, Indonesian, Itshekiri, Isoko, Italian, Kannada, Kinyarwanda, Konkani, Kurdish, Latvian, Lingala, Lithuanian, Malay, Malayalam, Mandarin, Marathi, Moldovan, Polish, Portuguese, Romanian, Russian, Shona, Slovakian, Spanish, Swahili, Tagalog, Tamil, Ukrainian, Urdu, Vietnamese, Visaya, Xhosa, Yoruba.

4. This was a very welcome initiative that helped to answer many of the questions above and provide assistance for Language Support teachers working with EAL pupils. At IILT seminars principals and teachers were introduced to the *Common European Framework of Reference for Languages* (CEFR; CoE 2001); all the resources developed by IILT subsequently published in *Up and Away* (2006); *Primary School Assessment Kit* (2007). Funding was withdrawn from IILT in 2008.

5. Annual standardised test results carried out in all Irish state primary schools are not published but kept on file by the Department of Education and Skills (DES). Because they are not published, there is no reference here.

(Department of Education and Skills 2018: 9). SBC has had no additional resources or support other than what is normally provided to schools with pupils who learn English as an Additional Language (EAL).

The approach has also brought unexpected benefits. The first concerns the Irish language. Fears that it might be swamped by the presence of such a multitude of languages turned out to be unfounded. In fact, the opposite was the case. The status of Irish was raised and its use increased within the school as pupils came to see that Irish, like any other language, could be used as a means of communication. Another welcome outcome was a developing culture of learner autonomy within the school (Little 1991; Little et al. 2017). Levels of motivation increased with many pupils working on their own initiative, devising ambitious language projects for themselves. Pupils' awareness of language, and how languages interact with each other, was enhanced. A further outcome was the cultivation of self-confidence, well-being, and social cohesion through pupils' developing awareness and understanding of each other's languages and cultures.

3 Curriculum delivery

In keeping with the idea of a whole-school, integrated approach to language learning, it is important that not only the language of schooling and curricular languages are seen and heard throughout the school, but home languages as well. If "the child's existing knowledge and experience form the basis for learning" (Government of Ireland 1999: 8), it is important to include the language that is "the default medium of [children's] self-concept, their self-awareness, their consciousness, their discursive thinking, and their agency [and] is thus the cognitive tool that they cannot help but apply to formal learning, which includes mastering the language of schooling" (Little et al. 2017: 202). Teachers encourage pupils to explore similarities and differences between home languages, being aware that "the cognate connections between the languages provide enormous possibilities for linguistic enrichment" (Cummins 2000: 21).

Home languages are used in three ways in SBC's classrooms:

- In reciprocal communication with other pupils who have the same or a closely related home language during play at the beginning of the school day or in the yard; pair and group work.
- For non-reciprocal purposes of display: *This is how we say it in my language* when learning to count, working with shapes and colours, and later when discussing more complex aspects of structure and vocabulary where home languages scaffold the learning of English, Irish and later, French.
- As a source of intuitive linguistic knowledge that individual pupils make available to the teacher and the rest of the class to enrich curriculum content and consolidate curriculum learning (for further discussion see Kirwan 2014).

Parents were positive in their reaction to the valuing and encouragement of their home languages. A Ukrainian parent told the principal that 'a weight was lifted off my shoulders when I heard that it was alright to speak my language at home'. An Indian parent was happy with 'the school's interest in our language. Before, my daughter was ashamed to hear us speaking Malayalam. Now she wants to read and write in it'. And an Irish parent who appreciated her daughter's developing communication skills commented that 'it makes them want to speak the Irish more at home'.

From Junior Infants onwards, children engage in dialogic interaction with their teachers. During curriculum delivery pupils are encouraged to contribute in their home language(s), in Irish and in English. Teachers ensure that Irish, and later French, is part of each pupil's daily communicative experience. In this way, ordinary activities become multilingual activities and *vice versa*. Language awareness is enriched, and pupils' implicit understanding of language is made explicit. From the formal curriculum to highlighting different languages on wall displays, multilingual greetings in the school's annual Christmas cards, concerts, art exhibitions, and religious ceremonies, a culture of language awareness is nurtured at all class levels.

Literacy is central to all language learning in SBC, with parallel texts in Irish/English, home language/English or Irish being a regular activity (Little and Kirwan 2019; Kirwan 2020). These texts develop from words and simple sentences in two languages in the early years, to more complex trilingual and multilingual texts as children progress through the school. As children's literacy in the language of schooling progresses, the skills they acquire are transferred to their home language. Parental involvement in fostering children's literacy in the home language is essential. Encouraging parents to take an active part in educational initiatives in the school where their language skills can be highlighted also contributes to the development of inclusivity and confidence among immigrant families (Kirwan 2015). Teachers, too, have expressed positive views as regards use of home languages, the integrated approach to language teaching that they have been instrumental in developing, and children's learning. For example:

> Children are responding very positively to the open language policy – even their body language, demeanour within class; the speed and accuracy with which they answer questions when their own language is involved; regardless of subject, their interest increases if it is something to do with home or their own language or their own experience; therefore when they respond it is with much more developed thought ... equally in writing (Little and Kirwan 2019: 50).

A further aspect of the approach taken concerns issues of identity and social cohesion. Pupils themselves have shared their views about the way in which they were being taught and the valuing of their home languages:

> [It helps pupils to get] personal into each other's cultures and languages [and] is very useful for friendship, for knowledge, so in many ways we're all expanding ... it makes you feel closer because you have a perspective on that person's point of view (speaker of Kurdish) (Kirwan 2019: 43).

> Sometimes it's, like, when we learn a language it's easier to learn other ones; sometimes it's not really about which language you're learning it's, like, how to learn a language (speaker of English) (Kirwan 2019: 45).

> It's like when two people speak the same language there's a kind of a bond between both of them (speaker A of Yoruba) (Little and Kirwan 2018: 335).

In contrast, when asked how they might feel in a situation where their home languages had been excluded from their education they said:

> It's so, so sad because it's like blocking a huge doorway ... it's like taking away an advantage of exploring (speaker of Kurdish) (Kirwan 2019: 42).

> Don't hide away from your own language because it's what makes you, you, and it's special and it's, you can't, it's like having an arm or a leg, you can't take it away from you (speaker of German) (Little and Kirwan 2019: 49).

> A child without a language is a child without a soul (speaker B of Yoruba) (Little and Kirwan 2019: 152, 153).

Growing proficiency in literacy provides the confidence for learners to embark on their own initiatives and many of them begin to produce work in all the languages at their disposal. This autonomous learning can operate at both an individual and cooperative level. It also paves the way for the introduction of self-assessment which happens in SBC in the penultimate year of schooling and focuses on the skills of listening, speaking, reading and writing (for further explanation see Little and Kirwan 2019: 137, 138). It can also be argued that had the teaching staff in SBC been in a position to teach pupils through

the home languages present in the school, children would have been empowered to drive their own learning to a much lesser degree. Because teachers lacked proficiency in these languages, they used their flexibility and creativity to engage in an approach to education that encouraged the development of autonomous skills that might not otherwise have been nurtured.

4 Conclusion

There are four assumptions underlying the approach to teaching and learning in SBC. The first is that the most effective way for plurilingual pupils to learn is to encourage them to use all the languages at their disposal autonomously – whenever and however they want to. Secondly, even very young children can be trusted to know how to use their home language autonomously as a tool of learning. The third assumption is that developing oral proficiency, literacy and language awareness is a complex process, in which reading and writing support listening and speaking and *vice versa.* The fourth assumption sees language awareness as a tool to support learning but also one of learning's most valuable outcomes, and it develops spontaneously when pupils make autonomous use of the languages at their disposal.

The key features of this approach are rooted in a view of primary education that is child-centred so that reflective and analytical dimensions of learning are firmly rooted in what pupils themselves contribute. Because classroom interaction takes account of their existing knowledge, skills and interests, pupils tend to be fully engaged. When pupils are activated to be agents of their own learning, their ability to direct and evaluate their learning becomes increasingly apparent as they move through the school. Finally, the development of literacy in English as the principal language of schooling feeds into, but also depends on, the development of pupils' literacy in their home language, Irish and (in Fifth and Sixth Class) French.

5 References

Beacco, Jean-Claude, Michael Byram, Marisa Cavalli, Daniel Coste, Mirjam Egli Cuenat, Francis Goullier and Johanna Panthier. 2015. *Guide for the Development and Implementation of Curricula for Plurilingual and Intercultural Education*, Strasbourg: Council of Europe. Available online: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ae621 (accessed 24 July 2020).

Beacco, Jean-Claude & Michael Byram. 2007. *From Linguistic Diversity to Plurilingual Education: Guide for the Development of Language Education Policies in Europe*, Strasbourg: Council of Europe. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016802fc1c4 (accessed 15 May 2020).

Central Statistics Office. 2017. *2016 Census Results, Part I*, Dublin: Central Statistics Office. http://www.cso.ie/en/media/csoie/newsevents/documents/census2016 summaryresultspart1/Census2016SummaryPart1.pdf (accessed 13 May 2020).

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cummins, Jim. 2000. *Language, Power, and Pedagogy: Bilingual Children in the Crossfire*. Toronto: Multilingual Matters.

Department of Education and Skills. 2018. Chief Inspector's Report January 2013-2016, 'Executive Summary', Dublin: DES. https://www.education.ie/en/Publications/Inspection-Reports-Publications/Evaluation-Reports-Guidelines/ (accessed 13 May 2020).

Garcia, Ofelia. 2017. 'Problematizing Linguistic Integration of Migrants: The Role of Translanguaging across the Bilingual Continuum', in David Little, Constant Leung and Piet Van Avermaet (eds), *Managing Diversity in Education: Languages, Policies, Pedagogies*, 243–56, Bristol: Multilingual Matters.

Government of Ireland. 1999. *Introduction to the Primary Curriculum/Curaclam na Bunscoile: Réamhrá*. Dublin: Dublin: Stationery Office. https://curriculumonline.ie/getmedia/93de2707-f25e-4bee-9035-92b00613492e/Introduction-to-primary-curriculum.pdf (accessed 24 July 2020).

Integrate Ireland Language and Training. 2006. *Up and away.* Dublin: IILT.

Kirwan, Déirdre. 2020. Converting Plurilingual Skills into Educational Capital. In Pauline Cogan (ed.). *Learn, The Journal of the Irish Learning Support Association,* 35-55. https://ilsa.ie/wp-content/uploads/2020/03/Learn-Journal-2020-13.3.20.pdf (accessed 13 May 2020).

Kirwan, Déirdre. 2015. *Cultivating Plurilingual Environments: Opportunities in Pre-school, Primary School and Beyond. 60-62.* Education and Training Boards Ireland. Spring 2015 edition. http://www.etbi.ie/wp-content/uploads/2015/02/ETBI_Spring15_final_low-res.pdf (accessed 15 May 2020).

Kirwan, Déirdre. 2014. From English Language Support to Plurilingual Awareness. In David Little, Constant Leung & Piet Van Avermaet. *Managing Diversity in Education*, 189-203. Multilingual Matters.

Kirwan, Déirdre. 2009. English language support for newcomer learners in Irish primary schools: A review and a case study. PhD thesis, University of Dublin, Trinity College.

Little, David (1991) *Learner autonomy: definitions, issues and problems.* Dublin: Authentik.

Little, David & Déirdre Kirwan. 2019. *Engaging with Linguistic Diversity: A Study of Educational Inclusion in an Irish Primary School.* London: Bloomsbury.

Little, David & Déirdre Kirwan. 2018. Translanguaging as a key to educational success: The experience of one Irish primary school. In Katrijn Maryns, Stef Slembrouck, Sven Sierens, Piet Van Avermaet & Koen Van Gorp (eds.). *The multilingual edge of education*, 313–339. Basingstoke, UK: Palgrave Macmillan.

Little, David, Leni Dam & Lienhart Legenhausen. 2017. *Language Learner Autonomy: Theory. Practice and Research*. Bristol: Multilingual Matters.

6 Biography

Déirdre Kirwan was principal of a linguistically diverse primary school in Ireland (1987-2015), Scoil Bhríde Cailíní, where she led the integrated approach to language teaching and learning that supported the use of pupils' home languages. Déirdre was awarded: *European Ambassador for Languages* (Léargas, 2008); PhD (Trinity College Dublin, 2009) for her research in language education. She has contributed to the new *Primary Language Curriculum* (Government of Ireland) and the *European Centre for Modern Languages*. Her most recent publication, co-authored with David Little, is entitled *Engaging with Linguistic Diversity: A Study of Educational Inclusion in an Irish Primary School* (2019) Bloomsbury Academic.

Developing an e-portfolio reflecting the concept of mediation for university students

Yukie Saito, Chuo University

Nearly 20 years have passed since the publication of CEFR (Council of Europe [CoE] 2001) and society has become more globalized with the development of information technology. At the same time, it has become more complex with many international issues needing to be solved. Reflecting on changes in society, the CEFR Companion Volume (CEFR/CV) was published in 2018, in which the concept of mediation is emphasized (CoE 2018). At the conference marking the launch of CEFR/CV, North (2018) explained mediation as a social and cultural process of creating conditions for communication and cooperation, which involves facing and hopefully defusing any delicate situations and tensions that may arise. In this increasingly globalized society, being able to play an active role as a mediator using English will be important. In Japan, the CEFR has been widely used in English education, including the new Courses of Study of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for junior high schools (MEXT 2017) and high schools (MEXT 2018). However, the concept of mediation emphasized in CEFR/CV has not been well recognized yet. Incorporating the concept of mediation is meaningful in English education in Japan for English learners to use English as a tool for global communication. In this paper, I will present the process of developing an e-portfolio based upon a learning management system (LMS) to promote students' use of English as a tool for global communication integrating mediation Can Do descriptors from the CEFR/CV.

Keywords: CEFR, CEFR/CV, mediation, e-portfolio, English as a tool for global communication

1 Introduction

Based on the CEFR (CoE 2001) and self-regulated learning (Zimmerman 2002), I developed a portfolio to help Japanese university students to use English as a tool for global communication (Saito 2017). In the process of the development, the European Language Portfolio (ELP) was also used as a reference. The ELP is defined as a document in which those who are learning or have learned one or more languages can record and reflect on their language learning and intercultural experiences (CoE 2001). The ELP has three components: a language passport summarizing linguistic achievements; a language biography that sets language learning targets, monitors progress, and reflects on language learning; a dossier that stores work in progress (Little et al. 2011). Saito (2017) included a language biography for goal-setting and self-assessment and a dossier to store learning documents. It was developed for first-year students of an English class at a private university in Japan, in 2015. Since then, the development of technology has advanced. In this digitized global society, acquiring Information and Communication Technology (ICT) literacy is crucial as one of the 21st Century skills (Griffin and Care, 2015). According to the summary of the results of the Programme for International Student Assessment (National Institute for Educational Policy Research 2019), at the age of 15, the time for Japanese students to use digital devices at schools is limited compared to students in other OECD countries. The department of a private university that I work for has introduced the policy of: Bring Your Own Device (BYOD); however, many students cannot use PCs well when they enter the department as freshmen. The previous portfolio (Saito 2017) was a paper-based portfolio. A new portfolio is being developed as an LMS-based e-portfolio so that more

opportunities for students to use PCs in English classrooms and at home may be created. Another important skill and learning outcome is that English learners are also expected to learn how to play the role of mediator using English. Because globalization has rapidly advanced, it has also brought about an unprecedented number of global issues that need to be solved. In this global society, the role of English as a tool for communication and as a tool to mediate to solve international problems is growing.

The outline of this paper is the following: first, the background to developing the e-portfolio reflecting mediation Can Do descriptors from the CEFR/CV for intermediate level university students is presented. Second, the process of adapting mediation Can Do descriptors to an English class with a designated ELT textbook is explained. Third, the process of developing an LMS-based e-portfolio and examples of the adapted Can Do descriptors on the LMS, Manaba, are presented. Forth, one example of a PowerPoint slide with the adapted Can Do descriptors for one lesson is presented. Also, classroom activities for students on how to apply these adapted descriptors are shared. Fifth, I shall discuss whether students understood the meanings of the adapted Can Do descriptors based on questionnaire survey data collected from students. Finally, I aim to make conclusions and draw implications based on drafting the LMS e-portfolio using the adapted mediation descriptors.

2 Background to developing the e-portfolio reflecting mediation of the CEFR/CV

The e-portfolio has been developed for the Integrated English Class for first-year students majoring in Global Informatics of a private university in Japan. Their English level is about 650 on the TOEIC Listening & Reading (L&R) Test. According to the Institute for International Business Communication (hereafter IIBC), a score of 650 on the TOEIC L & R Test has been benchmarked to level B1 of the CEFR (IIBC). The textbook used in the class is *Life 4* with topics from National Geographic published by Cengage, and it is aimed at B1+. It is also designed for cultivating students' critical thinking skills. Topics in the textbook, such as the issue of globalization, are educational and interesting and they are expected to promote these skills.

The textbook itself includes original Can Do descriptors after every unit; however, many of them are not related to critical thinking skills but to grammar. The topics in the textbook may be used for students to discuss actively and critically with fellow students. Thus, it is assumed that these mediation descriptors, such as summarizing a group discussion to others, can be integrated to promote deeper discussions. It is meaningful to incorporate the concept of mediation and develop an e-portfolio with mediation Can Do descriptors in order for university students to recognize and acquire skills for mediation, which can be helpful for them to work globally in the future.

3 Adapting mediation Can Do descriptors to an English class

Since the average student's English level is B1 and the textbook is aimed at B1+, I decided to use mediation Can Do descriptors for B1 and B2. However, those in the CEFR/CV (CoE 2018) are too long and complicated for students to understand. The following is one example of a B2 Can Do descriptor of overall mediation:

> Can establish a supportive environment for sharing ideas and facilitate discussion of delicate issues, showing appreciation of different perspectives, encouraging people to explore issues and adjusting sensitively the way he/she expresses things. (CoE 2018: 105)

As shown above, the descriptor is 32 words long. It can be understood by teachers who use it; however, it may be difficult for students to comprehend its meaning and use it and to understand their own goals and evaluate their learning. In the research by Pavlovskaya and Lankina (2019) to integrate the assessment of mediation competence into oral assessment in the context of CLIL, shortened mediation Can Do descriptors were used. Applying this to my context, they would need to be adapted for students to understand and set their own goals and evaluate their learning. Therefore, I decided to adapt mediation

Can Do descriptors for students in the Integrated English class to be able to more likely understand them. On the other hand, the adapted mediation Can Do descriptors also need to be aligned with the textbook contents. The following explains the process of how the mediation Can Do descriptors were aligned to the students' English level in the class and the textbook contents.

1. List B1 and B2 mediation Can Do descriptors from overall mediation (CoE 2018: 105), facilitating collaborative interaction with peers (CoE 2018: 209), collaborating to construct meaning (CoE 2018: 211), relaying specific information in speech (CoE 2018: 190, 191), and processing text in speech (CoE 2018: 197) from CEFR/CV (CoE 2018)

2. Adapt them to the students' English level so that they can understand and use them to identify their own goals and evaluate their learning

3. Analyze the textbook contents to see how the adapted mediation Can Do descriptors can be integrated into each unit.

Table 1 is an example of the process 1, 2, and 3 above, and they show Can Do descriptors from overall mediation for B1 and the adapted Can Do descriptors with units where these adapted descriptors are to be used. Table 2 is an example for B2.

Table 1. *Original mediation Can Do from B1 and adapted mediation Can Do*

Original mediation Can Do from overall mediation (B1) CEFR/CV (CoE 2018: 105)	Adapted mediation Can Do
• Can collaborate with people from other backgrounds, showing interest and empathy by asking and answering simple questions, formulating and responding to suggestions, asking whether people agree, and proposing alternative approaches.	• Can collaborate with classmates, showing interests by asking and answering simple questions (Unit 2a and 2b). • Can ask whether people agree or disagree and propose alternative approaches (Unit 2e and 2f).
• Can convey the main points made in long texts expressed in uncomplicated language on topics of personal interest, provided that he/she can check the meaning of certain expressions.	• Can convey the main points made in a longer text (Unit 2c and 2d).
• Can introduce people from different backgrounds, showing awareness that some questions may be perceived differently, and invite other people to contribute their expertise and experience, their views.	• Can introduce people from different backgrounds (Unit 2e).
• Can convey information given in clear, well-structured informational texts on subjects that are familiar or of personal or current interest, although his/her lexical limitations cause difficulty with formulation at times.	• Can convey information given in clear, well-structured informational texts (Unit1c).

Table 2. *Original mediation Can Do from B2 and adapted mediation Can Do*

Original mediation Can Do from overall mediation (B2) CEFR/CV (CoE 2018: 105)	Adapted mediation Can Do
• Can establish a supportive environment for sharing ideas and facilitate discussion of delicate issues, showing appreciation of different perspectives, encouraging people to explore issues and adjusting sensitively the way he/she expresses things.	• Can establish a supportive environment for sharing ideas (Unit 2a and 2b). • Can facilitate discussion of various issues (Unit 3a and 3b). • Can show appreciation of different perspectives and encourage people to explore issues (Unit 3c and 3d).
• Can build upon other's ideas, making suggestions for ways forward. Can convey the main content of well-structured but long and propositionally complex texts on subjects within his/her fields of professional, academic and personal interest, clarifying the opinions and purposes of speakers.	• Can convey the main content of well-structured texts and clarify the opinions and purposes of speakers (Unit 3c and 3d).
• Can work collaboratively with people from different backgrounds, creating a positive atmosphere by giving support, asking questions to identify common goals, comparing options for how to achieve them and explaining suggestions for what to do next.	• Can work collaboratively with other students creating a positive atmosphere by giving support and asking questions (Unit 4c and 4d).
• Can convey detailed information and arguments reliably, e.g., the significant point(s) contained in complex but well-structured texts within his/her fields of professional, academic and personal interest.	• Can convey significant point(s) contained in well-structured texts (Unit 4a and 4b).

For example, as shown in Table 1 and Table 2, in general, mediation Can Do descriptors are long, and one Can Do descriptor includes several goals. Therefore, I divided one descriptor into several Can Do descriptors or made one descriptor shorter for students to better understand them. In parentheses in the right column in Table 1, the unit in which each adapted Can Do descriptor is to be used is listed. For example, the descriptor of *Can establish a supportive environment for sharing ideas* is to be integrated into Unit 2, in which students are to learn about performances. In 2a of the unit, they learn how the development of technology has changed the way we listen to music, and in 2b of the unit, they are to ask and answer questions related to performances around the world. To create a supportive environment for sharing ideas, I introduce backchanneling expressions and follow up questions to the students. More details about the contents from Unit 1 to Unit 4 for the spring semester, the original Can Do from the textbook, the adapted mediation Can Do descriptors, and activities for the adapted mediation Can Do descriptors are shown in Appendix. In the next section, I present how I have been developing an LMS-based e-portfolio with the adapted mediation Can Do descriptors.

4 Developing an LMS-based e-portfolio and its use in the class

The portfolio with the adapted Can Do descriptors is for students to use on the LMS called Manaba. Manaba is a convenient and useful LMS because teachers can make a mini-test and a questionnaire survey, assign a report and a project, and give a grade. For example, teachers set assignments such as making a report on Manaba and students can submit their assignments on it. Manaba is a widely used LMS in universities in Japan. For this class, I use the LMS to upload PowerPoint slides before each lesson and assign students mini-tests, individual reports, and individual and group projects. My students use it to prepare for each lesson, work on assignments, and review their studies and to work cooperatively with other students in preparing for a presentation. Thus, it is assumed that integrating an e-portfolio on the commonly used LMS is more effective and useful than a paper-based portfolio for them to identify their learning goals and to observe and evaluate their learning as well as to get accustomed to using PCs.

Yet, the LMS-based e-portfolio has been designed using the shape of a questionnaire as this is an available tool. Figure 1 shows an example of the e-portfolio for lessons 8 and 9. The unit number, the topic of the unit, and the topic title are shown on the top. The Can Do descriptors are differentiated as Can Do descriptors from the textbook (unbolded) and the adapted mediation Can Do descriptors (bolded). Students can evaluate their learning by checking the descriptors. There are three choices for answering: *I Can Do it*, *I Can Do it to some degree*, and *I cannot do it now*. At the bottom is some space where students write self-reflection comments after each lesson. At the end of the semester, they are to check the descriptors and write reflections for all classes.

Figure 1. A part of the e-portfolio with the adapted Mediation Can Do descriptors on the LMS (Manaba).

5 Adapted Can Do descriptors and their use in the classroom

For this project, as mentioned before, for every lesson the Can Do descriptors were provided on PowerPoint slides and shared at the beginning of the lesson so that students can identify their learning goals for that lesson. Figure 2 shows the first PowerPoint slide for lesson 8, where students study Unit 3. As shown, the Can Do descriptors in Figure 2 are the same as the Can Do descriptors in the example of the e-portfolio in Figure 1.

In this lesson, students read a passage related to the wreck of the Titanic followed by a prompt "Do you think the remains of the Titanic should be left on the seabed, or should they be put in a museum?". One of the adapted mediation Can Do descriptors is *I can facilitate discussion of various issues*. In order to facilitate students' discussions, useful expressions are introduced such as "What's your opinion about the question?", "Do you agree or disagree with the statement?", "Why do you agree with it?", "Why do you disagree with it?", "How about you, (student' name)?" and then let them practice the expressions. In each group, they decide a facilitator, who helps other students to express their opinions. In addition to the prompt, I add three more discussion questions related to the passage so that every student in a group of three or four can be a facilitator in turn. Another adapted mediation Can Do descriptor is *I can give a summary of the group's view(s) in a new group*. To help students report a summary of the discussion, useful expressions are given such as "I would like to share a summary of our group's views.", "In our group, one student agreed with the statement and the other three students disagreed with it", and "They agreed with it because …". After practicing the students are assigned to new groups, and each student in the new group would summarize the previous group's views.
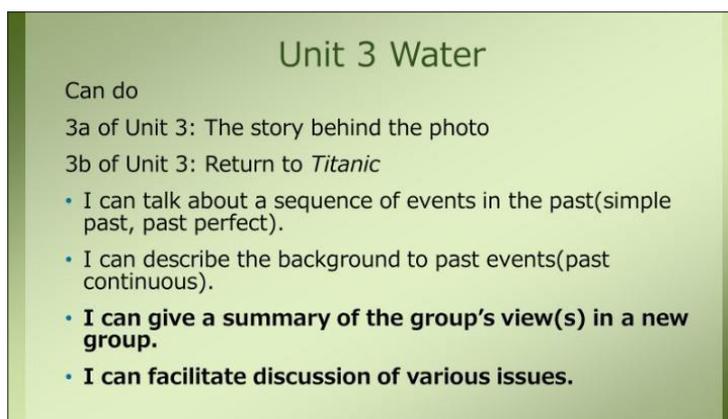
Figure 2. One example of PowerPoint slides with the adapted Can Do descriptors .

This is one example of how activities are planned to support students acquire skills described in the adapted mediation Can Do descriptors.

6 Student reactions to the adapted mediation Can Do descriptors

The original plan was to have students use the LMS e-portfolio and the provided Can do descriptors for unit 1 to unit 4 in the spring semester in 2020. However, some contents of the lessons could not be covered by the sudden change to online lessons due to the spread of Covid-19. Thus, instead of having them use the LMS e-portfolio after every lesson, I had students evaluate their learning with a shortened list of Can Do descriptors covering Unit 1 to Unit 3 at the end of the semester as mentioned above (see Section 4). One of the objectives was to evaluate whether they were able to do what were described in the Can Do descriptors. The other objective was to evaluate whether there were mediation Can Do descriptors which were difficult to understand. At the end of the questionnaire, the following questions in Japanese were added: (i) *Are there any of the Can Do descriptors that are difficult to understand?* and (ii) *If there were difficult ones, please write the number of the Can Do descriptor and give reasons why they were difficult*.

All of the 20 students enrolled in the class answered the questionnaire and one student answered question (i) with *Yes*. The first Can Do descriptor of mediation was *Can ask group members to give the reason(s) for their view.* To this first descriptor, the student wrote the comment: "As the first question in the questionnaire, it (the descriptor) was too abrupt, so I was a little confused. I think it's better to have it (the descriptor) later."

Although I introduced the Can Do descriptor on a PowerPoint slide at the beginning of the lesson, seeing it after a while at the end of the semester might have made the student think the descriptor was too abruptly introduced. The student's comment suggests that students need to be familiarized with Can Do descriptors so that they can understand their learning goals. Therefore, using and showing the adapted Can Do descriptors several times for different lessons to the students is important to have students make sure of their learning goals. Mediation competency can be cultivated over time through repeated opportunities to engage in such activities.

There was no comment about other adapted Can Do descriptors in terms of difficulty in understanding their meaning. Also, all of them evaluated their learning with the adapted Can Do descriptors. As for the detailed results of the questionnaire, they will be shared in a future publication. The process of developing the LMS-based e-portfolio indicates that by adapting original mediation Can Do descriptors from CEFR/CV (CoE 2018) to a current teaching context, they can be used for students to evaluate their learning.

7 Conclusions and implications

English has been and will be used as a tool for communication and mediation in this global society. It is assumed that university students need to prepare for using English to play a role as mediators in their future careers. Introducing and incorporating the concept of mediation can be done by adapting mediation Can Do descriptors in an EFL context with a designated ELT textbook. As shown here, this had been combined with an LMS-based e-portfolio. The portfolio can be helpful for students to understand and set their goals and evaluate their learning. For the continuing process, to reflect the cyclical phase of self-regulated learning, setting the phase of the students' monitoring their own learning will be important; thus, a part of the dossier where students can create a record of their learning process is expected to be added to the e-portfolio. The project of drafting the LMS-based e-portfolio introduced in this paper is only a single case for one English class at a university in Japan; however, it could be widely applied to different teaching contexts using a different LMS and adapting mediation Can Do descriptors to various contexts.

Incorporating the concept of mediation emphasized in the CEFR/CV will be important in this increasingly globalized society where many issues need to be solved internationally. However, it has not been reflected in language policy in Japan yet. Though it is possible to reflect on the concept of mediation at an individual level, as reported in this paper, it may be essential to consider how the concept of mediation can be incorporated in English education in Japan.

8 References

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press. . http://www.coe.int/en/web/commoneuropean-framework-reference-languages.

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg, France: Council of Europe. https://www.coe.int/en/web/common-european-framework-reference-languages/documents (accessed 31 July 2020).

Little, David, Francis Goullier & Gareth Hughes. 2011. The European Language Portfolio: The story so far (1991–2011). Strasburg, France: Council of Europe. https://rm.coe.int/16804595a7(accessed 31 August 2020).

MEXT. 2017. Chugakkogakushushidoyoryo [The Course of Study for junior high schools guidelines]. http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2018/05/07/1384661_5_4.pdf (accessed 8 August 2020)

MEXT. 2018. Kotogakkogakusyushidoyoryo [The Course of Study for senior high schools guidelines]. http://www.mext.go.jp/a_menu/shotou/new-cs/1384661.htm (accessed 8 August 2020)

National Institute for Educational Policy Research. 2019. Key Features of OECD Programme for International Students Assessment 2018 (PISA2018). https://www.nier.go.jp/kokusai/pisa/pdf/2018/01_point-eng.pdf (accessed 8 August 2020)

North, Brian. 2018. Completing the CEFR descriptive scheme: the CEFR companion volume. Paper presented at CEFR Companion Volume Launching Conference, Council of Europe, 16-17 May. https://rm.coe.int/completing-the-cefr-descriptive-scheme-north/16808ae70c

Griffin, Patrick. & Esther Care. 2015. The ATC21S Method. In Patrick Griffin & Esther Care (eds.). Assessment and Teaching of 21st Century Skills: Methods and Approach, 3-33. Cham: Springer International Publishing.

Pavlovskaya, Irina. Y. & Olga. Y. Lankina. 2019. How new CEFR mediation descriptors can help to assess the discussion skills of management students-Global and analytical scales. *CEFR Journal-Research and Practice. 1*, 33-40. https://cefrjapan.net/journal (accessed 31 July 2020)

Saito, Yukie. 2017. Developing a portfolio for English as a tool for global communication. In Fergus O'Dwyer, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.), Critical, constructive assessment of CEFR informed language teaching in Japan and beyond (English Profile Studies 6), 49-64. Cambridge: Cambridge University Press.

The Institute for International Business Communication. TOEIC® Program Kakutesuto to CEFR no Taishohyo [Comparison of TOEIC® Program Test Scores with the CEFR]. https://www.iibc-global.org/toeic/official_data/toeic_cefr.html (accessed 31 July 2020)

Zimmerman, Barry. J. 2002. Becoming a self-regulated learner: an overview. *Theory into Practice. 41*(2), 63–70.

9 Biography

Yukie Saito is an associate professor at the Faculty of Global Informatics of Chuo University. She obtained an MA from the Teachers College of Columbia University. Her research areas are the CEFR and its application in English education in Japan, and teacher cognition and classroom practice.

Appendix

Lesson plans in the spring semester with adapted mediation Can Do descriptors

Name of the Course: Integrated English I
Semester: Spring semester
Numbers of classes: 14 classes including the introduction of the course in the first class and a final exam in the last class.
Required textbook: Life 4 (Cengage Publication)
Levels of students: the average TOEIC score is about 650
Number of students: 20

14 lessons in Spring Semester	Content of the lesson	Original Can Do from the textbook, Life 4	Added Can Do from adapted mediation descriptors

Lesson1	Introduction of the course	Explanation of the course, the textbook, a portfolio with Can Do, online assignments, evaluation, a group presentation, and a final exam	
Lesson2	Unit1: Culture and Identity 1a of Unit1: How we see other cultures 1b of Unit1: Culture and color	I can ask and answer questions about things that are always and generally true, and routines (simple present). I can ask and answer questions about things happening now (present continuous). I can talk about professions and states: thoughts and mental process, etc. (stative verbs). I can use different questions forms: direct and indirect question.	I can ask a group member to give the reason(s) for their views. I can collaborate with classmates, showing interests by asking and answering simple questions. Activities: The topic of 1 a is related to stereotype images. I would like to make them discuss why people have stereotype images and how we Can Do not to have stereotype images which may lead to biases. • Introduce how to express their opinions and supporting reasons. • Introduce how to ask and answer questions showing interests in other students' opinions.
Lesson3	Unit1: Culture and Identity 1c of Unit1: A world together 1d of Unit 1: First impressions	I can introduce myself in formal and informal situations. I can open and close a conversation. I can ask for and give personal information.	I can consider two different sides of an issue, giving arguments for and against, and propose a solution. I can present my ideas in a group and ask questions for other students' opinions. I can convey information given in clear, well-structured informational texts. Activities: The topic of a passage of 1 c is related to globalization. I would like to make them discuss what globalization is, whether they are for or against globalization, and how we can maximize advantages of globalization and minimize disadvantages of globalization. • Introduce how they can agree or disagree with other people's opinions.
Lesson4	Unit1: Culture and Identity 1e of Unit1: About us 1f of Unit 1: Faces of India		I can summarize the main points made in clear, well-structured spoken and written texts on subjects that are familiar or of personal interest. I can summarize the main points made during a conversation on a subject of personal or current interest. Activities: 1 e is related to reading and writing a business profile. I will make them read the profile and summarize the main points of the profile. In 1f, students will watch a video related to a photographer of National Geographic. He is talking about his ambition to become a photographer and his visit in India as a photographer. I will have them summarize the contents of the video. • Introduce how they can scan the profile and make the summary. • Introduce how they can take a note to while watching a video in order to summarize the contents of the video.

Lesson5	Unit2: Performing 2a of Unit2: Music today 2b of Unit2: Learning to dance	I can describe different types of music. I can talk about things that happened in a time period up to or including the present (present perfect). I can use the correct tense when talking about things that have happened in the past tense (present perfect and simple tense).	I can establish a supportive environment for sharing ideas. I can collaborate with classmates, showing interests by asking and answering simple questions. Activities: 2 a is a passage related to how the development of technology has changed the way we listen to music. I will make them discuss what have been affected by the development of technology positively and negatively. 2b is dancing. I will make them ask and answer questions related to performances including dancing and use follow up questions to be active listeners. • Introduce backchannel expressions to be active listeners. • Introduce follow up questions to be active listeners.
Lesson6	Unit2: Performing 2c of Unit2: Living statues 2d of Unit2: What's playing?	I can talk about performers and performances. I can give my opinion about art events. I can ask for and give information about arts events.	I can convey the main points made in longer texts. I can help organize the discussion in a group by reporting what others have said and summarizing different points of view. Activities: 2c is a passage with 5 paragraphs. Have them do jigsaw reading and make a summary of one paragraph and put them summaries of five paragraphs together and make a summary of the long text. Have them discuss what performance, expedition, event, or concert that tourists in Japan shouldn't miss experiencing in a group and report the results of the discussion in a new group. • Introduce how they can scan the longer text and make a summary. • Introduce how to report and summarize the discussion results.
Lesson7	Unit2: Performing 2e of Unit2: A portrait of an artist 2f of Unit2: Taiko master		I can summarize a short narrative or article, a talk, discussion, interview or documentary and answer further questions about details. I can ask whether people agree or disagree and propose alternative approaches. I can introduce people from different backgrounds. Activities: In 2e, students are to read a portrait of an artist. After reading it, they will make a summary of the portrait and write a review of an artist whose work they like. After writing it, they will have another student read it and ask some questions about the review. 2f is a video about a taiko master who moved to the U.S and has been passing on taiko performing to Americans there. I will have them discuss whether traditional performing should be kept or not. • Introduce how they can scan the portrait and make a summary.

Lesson8	Unit3: Water 3a of Unit 3: The story behind the photo 3b of Unit 3: Return to Titanic	I can talk about a sequence of events in the past (simple past, past perfect). I can describe the background to past events (past continuous).	I can give a summary of the group's view(s) in a new group. I can facilitate discussion of various issues. Activities: 3 b is a passage related to Titanic. I will make them discuss the meaning of the passage's conclusion "The story of Titanic is not about the ship- it's about the people" and Titanic's future about whether it should be kept on the ocean floor or not. • Introduce the expressions to be used for giving a summary of the discussion. • Introduce the expressions to be used for facilitating discussion.
Lesson9	Unit3: Water 3c of Unit 3: Love and death in the sea 3d of Unit 3: No way! 3e of Unit3: What a weekend!	I can talk about water sports and activities. I can use adverbs to describe experiences.	I can convey the main content of well-structured texts and clarify the opinions and purposes of speakers. I can show appreciation of different perspectives and encourage people to explore issues. Activities: 3 c is a relatively long passage about a person who was almost killed in the ocean. I will have them discuss the main content of the passage and the person's opinion based on the experience. After reading the passage, they will talk about their experience they had where they learned a lesson. While they listen to other students' stories, they show appreciation and interests by expressing comments on the stories. • Introduce how they can skim for finding a writer's conclusion in a passage. • Introduce how to express comments on other students' stories.
Lesson10	4 a of Unit4: Will a robot take your job 4 b of Unit4: What's next?	I can show different degrees of certainty about predictions (may, might, could). I can make predictions about future events (predictions with will). I can ask and answer questions about future plans and arrangements (going to, present continuous). I can talk about stages in education and job training.	I can convey significant point(s) contained in well-structured texts. Activities: 4a is a passage about the future in which robots may take people's jobs in the future. I will have them find, convey, and discuss significant points including possible future of driverless cars. • Introduce how a topic sentence and a concluding sentence can be helpful to understand significant points in a passage.

Lesson11	4 c of Unit4: A better life? 4 d of Unit4: Would you mind...?	I can describe different jobs, job requirements, and conditions. I can make and respond to requests.	I can work collaboratively with other students creating a positive atmosphere by giving support, asking questions, and compare options. I can work towards a common goal in a group by asking and answering straightforward questions. Activities: They will read a passage about one person's experience of the fast- changing life in China. I will have them discuss and make a suggestion about how university students in Japan can have a better life in terms of the place they go to school, the place they live, the careers they choose, and their family.
Lesson12	4 e of Unit4: I'm enclosing my resume 4 f of Unit4: Everest tourism changed Sherpa lives		I can relay the main points contained in formal correspondence and / or reports on general subjects and on subjects I'm interested in. I can further develop other people's ideas, pose questions that invite reactions from different perspectives and propose a solution. Activities: 4 f is a video about Sherpas' life which has been changed by tourism. After watching the video, they will read Part 1 or Part 2 of the video script of 2 parts, find the main points, and pass them with students who worked on a different part. After the activity, they will discuss how modern tourism has affected people's life positively and negatively and propose solutions to minimize the negative sides.
Lesson13	Review 1~4 Group presentation		I can collaborate on a shared task by making suggestions and proposals of different approaches. I can act as a rapporteur in a group discussion, noting ideas and decisions, discussing these with the group. Activities: They will make a group presentation about effects of globalization in many perspectives such as culture, arts, technology, education, work, and the environment. To prepare for the group presentation, they are expected to work cooperatively with other members and act as a rapporteur. • Introduce how to make a presentation effectively.
Lesson14	Final Exam	Final exam and self-evaluation of the course with the portfolio with Can Do. Ask students to do a questionnaire survey about the course, their self-evaluation, and the portfolio.	

Classroom-based assessment of group discussion: Challenges and opportunities

Olga Y. Lankina, St. Petersburg State University

Yulia V. Petc, St. Petersburg State University

Collaborating in a group and leading group work are often used in the modern language classroom. However, it still seems relevant to find an effective way to assess group discussions and to identify problems that students experience when they have to mediate texts and concepts or collaborate on shared tasks. This paper analyzes the engagement of students with a task to mediate and discuss information in a group and how students approach the tasks at the B1 and B2 Council of Europe (CoE) or CEFR levels. The objective of this paper is to show that group discussions can be used for the formative and summative assessment of mediation skills as they are described in the CEFR Companion Volume (CEFR/CV; CoE 2018). We will describe the process of how learners can receive global achievement marks for mediation on par with analytical marks for interaction, discourse management, range, accuracy, and phonological control. We also provide an example of how assessment of a group discussion can be done by giving students a global achievement mark for mediation and analytical marks. The outcome of the research is a test that can be used in the classroom and to provide criteria for assessing mediation when it is part of a group discussion.

Keywords: mediation, classroom-based assessment, descriptors for mediation, global and analytical marks, CEFR/CV, group discussion

1 Introduction

Conveying information and discussing it is a prerequisite for effective cooperation. In order to understand the nature of communication in a group, we looked at the competencies that underlie group discussions. Interactional competence, which is a basis for group discussions, includes discourse competence, non-verbal communicative strategies, and strategic competence (Celce-Murcia et al. 1995). This is demonstrated when people "interact as speakers and listeners to construct meaning" (Fulcher and Davidson 2007: 49). These are the most social speech acts because they are co-constructed (McNamara 2006: 64). Researchers point out the complexity that this social element adds to co-constructed speech (Luoma 2004; Lazaraton 2014). Discussants construct the event together and share the right to influence the outcomes. Besides, in situations of information-related talk aimed at conveying information on a particular issue, speakers have to deploy several strategies to establish common ground, give information in bite-sized chunks, ensure logical progression, and maintain a comprehension check. (Luoma 2004: 20). These features of co-constructed speech pose certain challenges for students who learn how to mediate and discuss information in an academic environment.

CEFR frames mediation as a language activity in its written or oral form, which makes communication possible between people who are not able to communicate with each other directly (CoE 2001: 14). The co-construction of meaning while collaborating in a group is described in the (CoE 2018), which emphasizes the mediator as a social agent of communication. It is very important that when we use a language we combine several activities and switch between the modes of communication: mediation,

reception, production, and interaction (CoE 2018: 33).

Since 2001, linguistic, cultural, and social aspects of mediation have been thoroughly discussed and mediation has been linked to the context of school education (Zarate et al. 2004; Coste and Cavalli 2015). Mediating concepts in a collaborative work context implies, on the one hand, the learner's contribution to the effective work of the group towards achieving a shared objective, and on the other hand, his/her stimulating and developing ideas as a member of the group. The latter is characterized by the learner's ability to further develop other people's ideas and opinions, co-develop ideas, and evaluate problems and suggestions. The definition of mediation was developed by Coste and Cavalli and their proposition about the distinction between *cognitive mediation* (constructing or conveying meaning) and *relational mediation* (facilitating relationships) (Coste and Cavalli 2015: 28) became the key principle for designing the Global Achievement Scale used in this research. We also used mediation activities as listed by North and Piccardo (North and Piccardo 2016: 21) and described in the Illustrative Descriptor Scales (CoE 2018: 116) to write the contents of the Global Achievement Scale.

On the assumption that young adults should learn a foreign language and develop their social competencies together (Canale and Swain 1980; Long and Porter 1985; Sharan 1990; Slavin 1990; Pavlovskaya 2017), group work is a fundamental condition of language learning. It also plays an important role in the occupational and academic domains and therefore, has to be presented in an adequate format in language tests.

2 Research objective and participants

Having faced the task of teaching students to mediate texts and concepts, we noticed that students find it difficult to engage in mediation in group discussions and develop each other's ideas.

The objective of the research was to investigate the extent to which there is a deficiency of certain skills, i.e. developing other people's ideas among CEFR B1 and B2 level students, which does not allow students to hold discussions successfully.

The research participants were undergraduate students in the Management Department of St. Petersburg State University (Russia; 91 people) and two trained examiners. The students' levels of English language proficiency varied between CEFR B1 and B2. In compliance with the standards for foreign language education applied at the university, students at the CEFR B1 and lower levels are taught General English. Students who have achieved the level of language proficiency equal to CEFR B2 or higher go on to study Academic English and English for Specific Purposes, particularly Business English in the Management department. Most of the students can be described as motivated and willing to learn the language for academic purposes (e.g., a semester abroad program, lectures in English delivered by non-native and native speakers), professional development (e.g., scientific research), and career opportunities.

Students were grouped into three cohorts according to the results of the placement test: B1 low, B1 average, and B2. The research tests were carried out in the middle of the academic year. By that time the students' proficiency had not changed considerably. The students worked in three-person groups. In cases, when it was not possible to organize a three-person group, groups of four were formed. Because the number of four-person groups was minimal, the impact of group size on the results was not taken into consideration. The details of the groups are presented in the table below.

3 Methodology – Test description and test marking

Three Speaking tasks were used for measuring oral performance in group discussions at the CEFR B1 and B2 levels. The assessment tasks were designed to conform to the foreign language curriculum content. Though the research participants were offered different tasks in terms of content, each task was tailored to comply with the students' language proficiency level and the course requirements. The tasks were designed to require the students to discuss a situation by mediating the background information, expressing opinions, and negotiating an agreement. For example, in General English lessons, students

worked in groups of three or four people. Each student had to present their charity event proposal and the group decided which event would be the most successful. More advanced students engaged in Business English study also worked in groups of three or four people, and had to discuss different leadership styles then choose those which they thought would best suit their educational institution. To familiarize the participants with this kind of task, they completed similar tasks with different course materials under the guidance of a teacher prior to the assessment stage.

Table 1. *Student profile*

Characteristics of groups	Cohorts		
	B1 low	B1 average	B2
3 students	5 groups	4 groups	16 groups
4 students	2 groups	2 groups	-
Total number of students	23	20	48
Type of English class	General English	General English	Academic Communication skills in English for Managers
Course details	Two semesters, 90-min face-to-face class twice a week	Two semesters, 90-min face-to-face class twice a week	Two semesters, 90-min face-to-face class twice a week.

Table 2. *Test details*

Test level	B1	B1 +	B2
Test topic	Charity (B1 test)	Education (B1 test)	Leadership (B2 test)
Input	Written (80-word role cards)	Written (150-word role cards)	Video (8-minute videos)
Preparation time	2 min	3 min	5 min
Output (speaking time)	10 min	12 min	15 min
Test length	12 min	15 min	28 min
Procedure	Each student is given a written stimulus, a unique option, and a task, then asked to present their options and discuss an issue.		The procedure is the same as the B1 level, except for input, which was a video.
Task focus	Mediating factual and general information, e.g., options for charity or education.		Mediating factual, general and abstract information, e.g., leadership styles.
	Using functional language to discuss options, invite to speak, ask straightforward questions, give reasons for views, repeat back to confirm mutual understanding, define the task, collaborate on a shared task, and maintain the focus of a discussion.		Using functional language to further develop other people's ideas and opinions, present ideas and invite reactions, and consider different sides of an issue; organize a discussion, refocus it, highlight the main issue, and collaborate in decision-making.

Each student in a group was provided with a short input text on a common topic and was given 2-5 minutes to familiarize themselves with the content. After that, the texts were retrieved, and the students were required to convey the main idea of their input text ensuring all of the group members' comprehension. The task rubric was formulated in a manner that required students to facilitate discussion by inviting others to speak, express their agreement, and contribute towards fulfilling the goal. To create the conditions for uninterrupted communication, the teacher would refrain from guiding students through the task. The teacher evaluated the students, awarding Analytical marks and a Global Achievement mark. The tasks were recorded to provide the possibility for marking multiple times.

While discussing the topic, students had to relay information by processing, reformulating, summarizing, or streamlining it (cognitive mediation). At the same time, they aimed to build rapport with the other students in the discussion group, creating conditions that were instrumental for facilitating and moving the discussion towards accomplishing a shared communicative goal (relational mediation). Thus, mediation of a text, concepts, and communication became the global objective of the task and was reflected in the Global Achievement mark, which was awarded on the basis of such descriptors as relaying specific information in speech, facilitating collaborative interaction with peers, and collaborating to construct meaning.

The Analytical criteria were interaction, discourse management, range, accuracy, and phonological control. The CEFR/CV descriptors for overall mediation, mediating concepts via collaborating in a group, leading group work, and mediating communication (CoE 2018: 101, 116-117, 119, 120-123) were adapted by the authors to write the Global Achievement Scale, while the Analytical Scale was drawn up without amending from Table 3: Common Reference Levels: qualitative aspects of spoken language use (CoE 2001: 28-29). The authors of the paper who were the students' teachers and examiners agreed on the 3 and 5-band performances which were used as standardized performances to refer to while assessing.

Both the Analytical and the Global Achievement scales had 5 points, where 1, 3, and 5 bands were described. Students' performance was reported in terms of 1-5 marks where '1' is the lowest and '5' is the highest mark.

4 Results and Analysis

The tests' results were analyzed using the Rasch Model (FACETS) and Classical Test Theory (ITEMAN).

Table 3. *Summary statistics of ITEMAN and FACETS analysis*

ITEMAN								FACETS				
Level	Topic	Number of students	SD (standard deviation)	Min Score	Max Score	Item mean (aver. of the criterion) marks	Alpha (KR 20)	SEM (standard error of measur.)	Mean Rater correlate. (PtMea)	Item mean	Mean Resd	Mean StRes	Mean S.D.
B1	Charity	23	4,734	13	30	3.485	0.951	1.044	0.81	3.48	0.00	0.02	1.02
B1+	Education	20	5,089	11	30	3.833	0.952	1.113	0.75	3.83	0.00	0.00	1.00
B2	Leadership	48	3,627	18	30	3.917	0.925	0.992	0.84	3.92	0.00	-0.01	0.99

As we can see in the table, the most difficult test, or the less able students, were the B1 cohort who took the Charity test (Item mean 3.48), while the least difficult test or the best-prepared students were those who took the B2 Leadership test. According to the FACETS User Manual, when the parameters of the test are successfully estimated, the mean Resd is 0.0. In our tests, it was 0.00. When the data fit the Rasch model, the mean of the Standardized Residuals (StRes) is expected to be near 0.0 (in our tests it varied from -0.01 to 0.02), and the Sample Standard Deviation (S.D.) is expected to be near 1.0 (1.02; 1.00; 0.99). The raters' correlation was quite strong and exceeded 0.75. If we look at the raters' agreement closely in Figure 1 below, we can see that mediation and interaction have a larger discrepancy than the other criteria, which can be explained by the fact that even though the raters had prepared for the assessment session, assessing mediation was still a new experience for them.
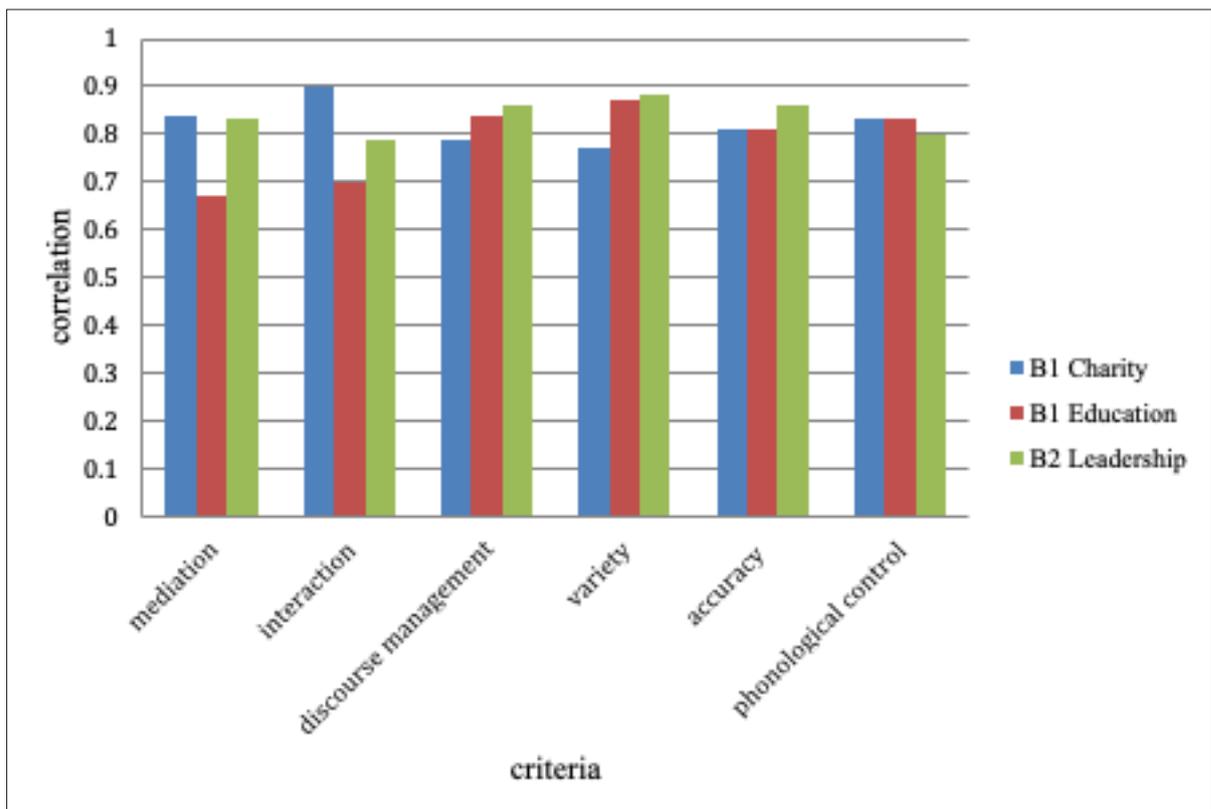
Figure 1. Raters' correlation.

Our aim was to analyze the factors that impede students' ability to co-construct the new meaning. In order to do that, we compared the amount of time students spent on presenting their own views (or producing their long turns) and collaborating on the task. The quantitative analysis of students' oral performances revealed that students at a low B1 level tended to retell the content of their input cards to other group participants almost without debating the issue. At the B1 level the following pattern of group performance prevailed: a series of long turns followed by a short and quite rudimentary collaboration phase (B1 low 0.8:0.2; B1 average 0.66:0.34; the numbers here represent the long turn and collaboration phases as they relate to the length of the test). This changes at the B2 level where mediation became naturally integrated into the discussion, and long turns made up only one-fifth of the total length of the task (0.2:0.8, the long turn and collaboration phases respectively), as shown in Figure 2. This integration of mediation manifests itself in the way discussants (1) took turns: they became more confident in balancing contributions from other group members with their own; (2) switched easily between modes of communication, so that mediation, interaction, and production in their speech became intertwined. For example, while discussing leadership styles, a student briefly outlined the contents of the video she

watched and then pointed out one feature she liked most about "silent leadership", that is "connectedness with the community". The other student built upon that by specifying the circumstances when "silent leadership" would suit the society best, i.e. "a period of prosperity", according to how she understood it. The third student provided details from the video and his own experience and argued that "paternalistic leadership" seemed to be the most effective for a big company. As we can see, students blended their references to the input videos and their own opinions, thus mixing mediation with interaction. It was interesting to note this feature of a discussion at the B2 level, and to match it with how "Facilitating collaborative interaction with peers" is described in the CEFR/CV (CoE 2018: 115).
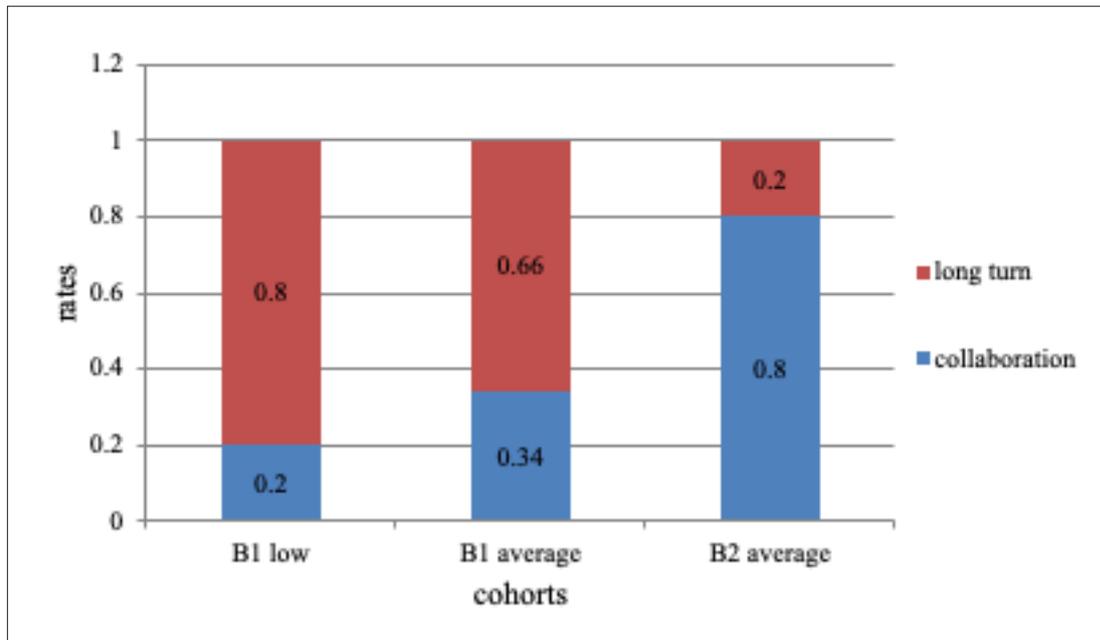
Figure 2. The rates of long turn and collaboration phases as they relate to the length of the whole test.

We also looked at how descriptors for mediation (collaborating in a group) work in discussions in the hope of better understanding the development of the students' ability to build on what other people say. For this purpose, we made a list of 44 descriptors that were most suitable for our tasks and levels: descriptors for overall mediation, managing interaction, encouraging conceptual talk, and facilitating pluricultural space in CEFR/CV (CoE 2018: 101, 116, 119, 120-121). Then we shortlisted those descriptors to 17 that students were using in their speech and gave them 'short names', e.g., 'collaborate on a shared task' stands for "Can collaborate on a shared task, for example formulating and responding to suggestions, asking whether people agree, and proposing alternative approaches." (B1. Facilitating Collaborative Interaction). 'Consider different sides' is used for the descriptor "Can consider two different sides of an issue, giving arguments for and against, and propose a solution or compromise." (B2. Collaborating to Construct Meaning). (CoE 2018: 116). The occurrences for each descriptor were counted in each cohort and descriptors were sorted in order of frequency, as shown in Figures 2-4.

At the B1 low level, when faced with the problem of shared decision-making, students resorted to straightforward strategies such as stating the goal of the discussion (e.g., *We have to choose the best idea*), turn-taking, inviting partners to contribute to the discussion (e.g., *What do you think? Lena, what's your idea?*), expressing agreement/disagreement (e.g., Yes, I agree; No, I don't like it.) They also repeated back (A*: Do you mean that children from a hospital can cook and take part in the competition?* B: *They can, I think. But I say that all people who want can cook something and people choose a winner*.) and maintained the focus of the discussion (A: *Where will we get products for the competition?* B: *I'm sorry but we should talk not about where we take food or the place where we keep equipment … we should think about what we can do to get some money for the children's hospital.*) As we can see in Figure 3, the descriptors of B1

level dominate here; however, there are some instances of using descriptors from higher levels (B2; considering the pros and cons of an issue). The numbers in the figure show the occurrences for this particular cohort.
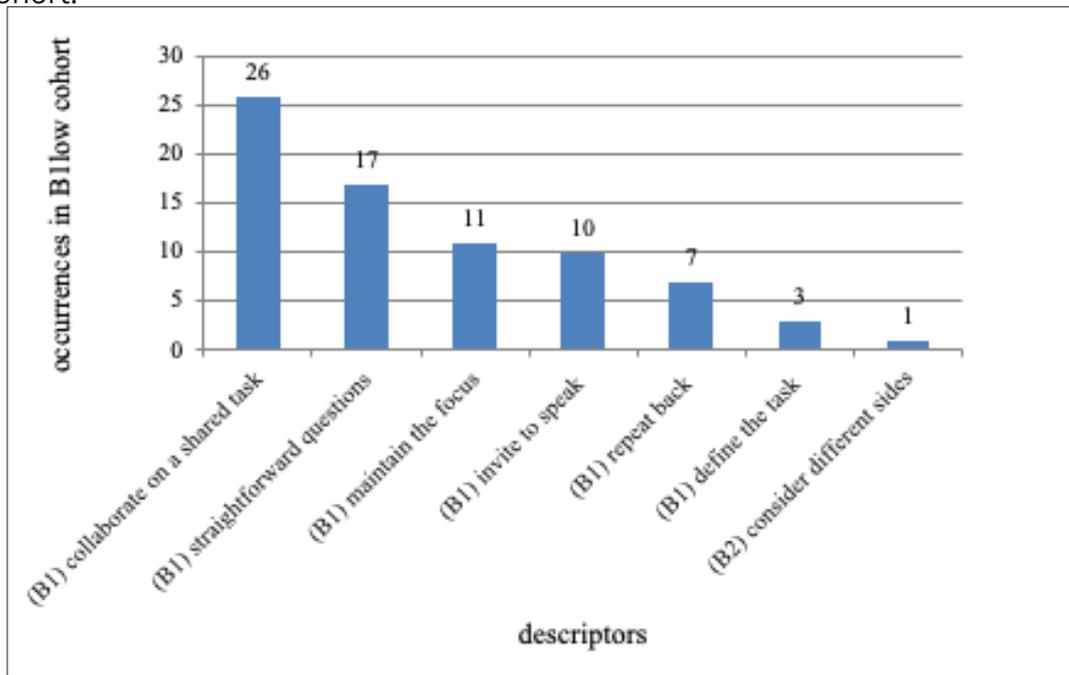

Figure 3. B1 Low cohort: descriptors distribution.

Figure 4 is related to the average B1 level, where one can find even more attempts to use sophisticated language functions, such as developing other people's ideas and considering two different sides of an issue (e.g., A: *In my class, there were strong and weak students.* B: *I see what you are saying. It's good to mix students, but it can be hard for those who are ambitious*); or invite other people's reactions (e.g., A: *What do you think we can borrow from the educational systems of other countries?... B: I think it would be nice if school in Russia started at 9 o'clock or later like in the Netherlands. What do you think?*) In our tests, *developing other people's ideas* as a language function was first used at this level (position number 8 in Figure 4).
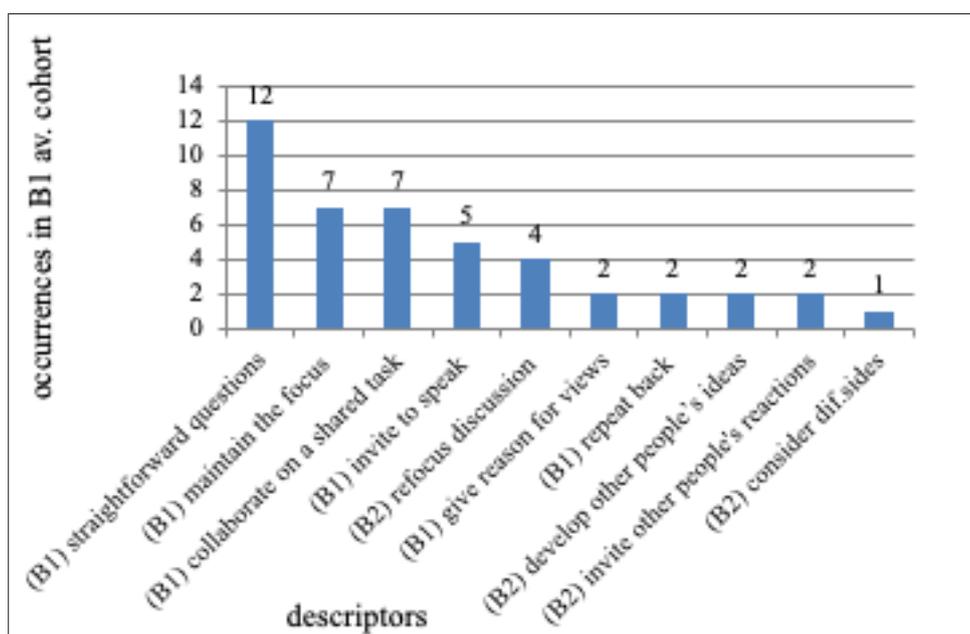
Figure 4. B1 Average cohort: descriptors distribution.

At higher levels, discussions became lengthy and detailed, and students demonstrated a wider repertoire of exploited strategies. They showed their ability to organize and manage collaborative work (e.g., *Today we're going to talk about different types of leaders.... And first of all, we should understand what type of leadership everyone watched a video about. Let's talk about different types of leaders and then we'll discuss what leadership style can be used at our school*). Students co-developed ideas, offered suggestions, compared different points of view, summarized, etc. (A: *The type of leadership depends on a person's characteristics.* B: *Apart from a leader's personality, social milieu plays an important role in making this or that type of leadership popular.* C: *Right, so we have personal and social factors here. Which would be more important?*) Even though developing other people's ideas as a language function was used more often by the B2 students, it still found itself at the tail of the descriptors distribution (position number 8 in Figure 5).
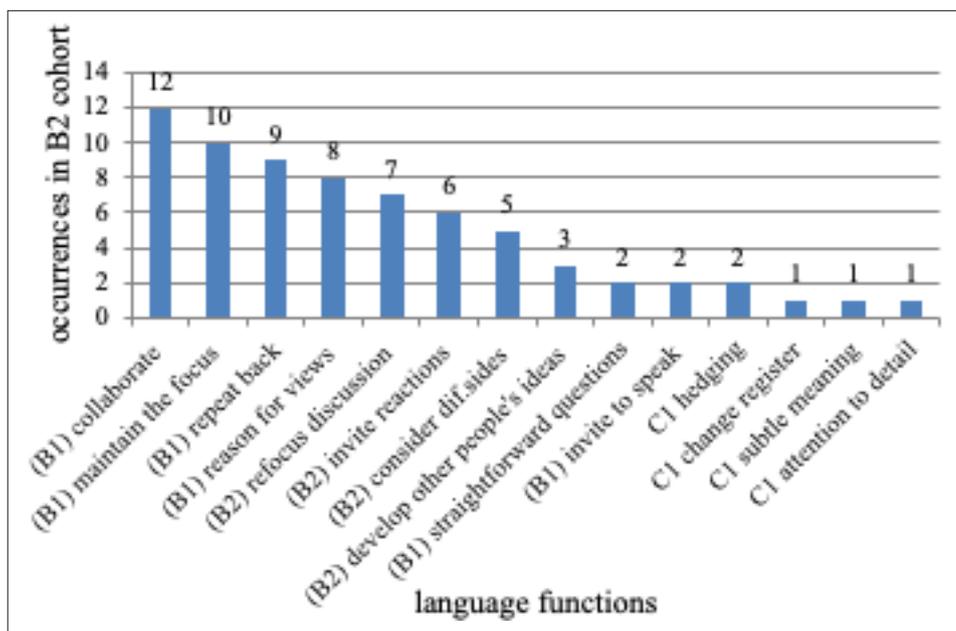
Figure 5. B2 cohort: descriptors distribution.

5 Conclusions

The tests that have been described in the paper were designed with the aim of assessing students and identifying gaps in their language skills including mediation. This type of test task can be used for the formative and summative assessment of mediation in group discussions, and language proficiency in general.

The descriptor analysis showed which descriptors were used by students often and successfully, and which descriptors students did not pick. Judging by the marks that students received, we can see which language skills had been developed and to which extent as well as which skills were lacking. In this way, the use of CEFR/CV descriptors can help identify problem spots in students' skill development, and backward design can be used to cover the lacunae.

We also observed changes in language behavior with the progression of the language level: (1) if students are given a task to mediate and discuss some information, at lower levels they tended to complete the task in two distinct phases, i.e. relaying information first and then interacting to solve the task; while at higher levels relaying information and interacting were blended and continued throughout the task; (2) the variety of descriptors for mediation which can be attributed to a speech act increases from lower to higher levels. These features of students' language behavior can give an examiner or a teacher a good sign when determining the level of a test-taker.

6 Limitations and areas of further research

The descriptors that were not observed in the test were either not required for the task, or students lacked the skill to use them. Backward design in planning the curriculum can be used to cover the outstanding descriptors. In backward design, educators first identify learning outcomes, then write tasks for assessment, and finally create learning activities to achieve the desired results (Wiggins and McTighe 2008: 17-18).

The group format (3 or 4 people) and the lack of interference of the Interlocutor into group work creates an authentic context, which can be viewed as a positive feature of this type of exam. At the same time, one can argue that the absence of the Interlocutor can disadvantage shy or lower level candidates; however, we believe that the situation when candidates have an opportunity to communicate freely provides plenty of room for mediating communication and its assessment. It may also have a positive effect on teaching and sharpening those skills in the classroom setting.

A larger sample of oral performances and the involvement of more assessors would increase the reliability of the measurements. Also, a more refined focus on the nature of collaboration in a group can give more information on how to teach and assess group discussions. For example, it might be worthwhile researching whether the cooperative and competitive modes of communication affect successful task completion.

We have noticed that the skills for mediating concepts are quite often underdeveloped as a social skill among the target population of students who lack the ability to articulate their stance clearly. This can be viewed as a problem area which requires the attention of educators and test developers. These tests, as well as other similar tests, can be offered as a tool of measurement of progress on the way to working out these problems.

7 References

Canale, Michael & Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1). 1–47.

Celce-Murcia, Marianne, Zoltan Dörnyei & Sarah Thurrell. 1995. Communicative competence: a pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6(2). 5–35.

Coste, Daniel & Marisa Cavalli. 2015. *Education, mobility, otherness. The mediation functions of schools*. Strasbourg: Council of Europe, Language Policy Unit. DGII.

Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment (CEFR).* Strasbourg: Council of Europe. https://rm.coe.int/16802fc1bf (accessed 20 August 2020).

Council of Europe. 2018. *The CEFR Companion Volume with new descriptors.* Strasbourg: Council of Europe. https://rm.coe.int/cefr- companion-volume-with-new-descriptors-2018/1680787989 (accessed 03 August 2020).

Fulcher, Glenn & Fred Davidson. 2007. *Language Testing and Assessment. An advanced resource book.* London and New York: Routledge.

Lazaraton, Anne. 2014. Second language speaking. In Celce-Murcia M., D. M. Brinton & M. A. Snow (eds.), *Teaching English as a second or foreign language*, 4th ed., 106-120. Boston: Cengage Learning.

Long, Michael H. & Patricia A. Porter. 1985. Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly* 19(2). 207-228.

Luoma, Sari. 2004. *Assessing speaking*. Cambridge: Cambridge University press.

McNamara,Tim & Carsten Roever. 2006. *Language Testing: The social dimension.* Malden, MA: Blackwell.

North, Brian & Enrica Piccardo. 2016. *Developing illustrative descriptors of aspects of mediation for the Common European Framework of Reference (CEFR)* https://www.rm.coe.int/developing-illustrative-descriptors-of-aspects-of-mediation-for-the-co/1680713e2c (accessed 03 August 2020).

Pavlovskaya, Irina (ed.) 2017. *Актуальные вопросы языкового тестирования. [Modern issues of language testing].* 2. St. Petersburg, 684 p. URL http://hdl.handle.net/11701/7188 (accessed 03 August 2020).

Sharan, Shlomo. 1990. *Cooperative learning: theory and research.* New York: Praeger Publishers.

Slavin, Robert E. 1990. *Cooperative learning: Theory, research and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Wiggins, Grant & Jay McTighe. 2008. *Understanding by Design.* Alexandria, Va.: Association for Supervision and Curriculum Development.

Zarate, Genevieve, Aline Gohard-Radenkovic, Denise Lussier & Hermine Penz. 2004. *Cultural mediation in language learning and teaching.* Strasbourg: Council of Europe Publishing.

Software

- ITEMAN http://www.assess.com/iteman/
- FACETS https://www.winsteps.com/minifac.htm
- FACETS User Manual, https://www.winsteps.com/facetman/index.htm

8 Biography

Olga Y. Lankina is a researcher. Her field of research includes language testing and mediation. She holds a specialist degree in applied linguistics and works as an ESP teacher and teacher of English for academic purposes at St. Petersburg State University (Russia). Olga is a Speaking Examiner and examiner trainer. In 2013-2014 she took part in linking the University test to the CoE. She is a member of EALTA and UKALTA.

Yulia V. Petc holds a Master's degree in Pedagogy and Linguistics. She works as a language test developer at the Language testing department of St. Petersburg State University (Russia) and an ESP teacher and teacher of English for academic purposes at St. Petersburg Mining University (Russia). Yulia is a Speaking Examiner. In 2013-2014 she took part in linking the University test to the CoE. She is a member of EALTA.

News
JALT CEFR & LP SIG

Maria Gabriela Schmidt, SIG Coordinator

There are three announcements concerning the JALT CEFR & LP SIG that I would like to share:

(1) CEFR & CLIL Symposium & Conference on October 23-25 2020 (online)
(2) New research project – Call for Collaboration
(3) Outcome of the research project now published

(1) CEFR & CLIL Symposium & Workshop on October 23-25 2020 (online)

With increased interest in the integration of CEFR and CLIL, the JALT CEFR & LP SIG wants to promote research and practice linking the CEFR and CEFR/CV with CLIL. This international symposium & workshop serve to further the educational discussion within CEFR in all levels of education, with the additional understanding of CLIL (Content and Language Integrated Instruction). Our goal is to create a forum for discussions promoting the understanding of concepts related to, applications of, and future directions of CEFR and CLIL in Japan and beyond. The theme is

The praxis of teaching, learning, and assessment with CEFR and CLIL

Invited speakers are Masashi Negishi (Tokyo University of Foreign Studies) and Yuen Li Lo (University of Hong Kong). The conference will be held online and will be free or charge. Yet, attendance is limited to 100 people on each day on a first-come basis.

For signing up, please use: https://tinyurl.com/y22g9ev6

For more details see the special conference homepage:
https://cefrjapan.net/events/22-events/83-cefr-and-clil

Timeline:
- Friday 23 October 2020: 7:00-9:00 PM Tokyo, Japan Standard Time (CEST 12-2 pm)
- Saturday 24 October 2020: 9:00 AM-6:15 PM JST (CEST 2-11:15 am)
- Sunday 25 October 2020: 9:00 AM-1:40 PM JST (CEST 2-6:40 am)

This is a joint event of JALT CEFR & LP SIG, JALT Akita Chapter, and Akita International University.

(2) New research project – Call for Collaboration

In April 2020, we were granted the third JSPS Grant-in-Aid (KAKEN) research project (no. 20K00759) directly involving the SIG, and the fifth involving SIG members, on

Foreign Language Education Reform through Action Research
Putting CEFR educational principles into practice

The JALT CEFR & LP SIG is looking for researchers to participate in our latest KAKEN JSPS research project. The aim of this project is to support numerous small-scale action research (AR) projects related to foreign language teaching in Japan and beyond, in which practitioners reflect on and find ways to improve their teaching practice using the CEFR as a reference tool. Each AR team will be provided with support and guidance to ensure that their research is conducted in a systematic way in relation to the AR literature and reflective of CEFR principles.

The teams will be grouped thematically around five central themes:
- Designing a curriculum and/or course
- Materials Development
- Assessment
- Learner autonomy and the European Language Portfolio
- Classroom Implementation.

These categories roughly follow the chapters of the CEFR & LP SIG's latest publication, **CEFR-informed Learning, Teaching and Assessment: A Practical Guide** (Nagai et al. 2020). The book authors and SIG officers will offer support and guidance in their respective areas of expertise while undertaking their own research projects. Each team will also be provided with a succinct review of the AR literature, and an action research framework to guide their research efforts. Ultimately, we plan to publish these case studies in either an edited volume or our in-house journal: *CEFR Journal – Research and Practice.*

Schedule:
- **2020: Work with CEFR & LP SIG members to identify research foci and to plan AR projects.**
- **2021: Carry out action research/Collect and analyze data.**
- **2022: Reflect on AR projects and write research up for publication.**

This project is related to the JSPS KAKEN project no. 20K00759「アクションリサーチの手法を用いた言語教育改善: CEFRの教育理念を参考にして」. Taking part in it, presenting and/ or publishing will be acknowledged by including your name in the yearly reports and in the final report of the KAKEN as research collaborators (研究協力者) during the research period (2020-2022). The names of research collaborators will appear in the final report of the ongoing KAKEN project, as shown in the previous KAKEN project report (see KAKEN database https://kaken.nii.ac.jp/en/grant/KAKENHI-PROJECT-16K02835/).

JALT and the CEFR & LP SIG are included explicitly in the research proposal. Being a member is favorable but not mandatory for taking part in the project. As this is a JSPS KAKEN project, the outcome aims to contribute substantially to foreign language education (English or other languages) in Japan. Yet, the language for presentations and publications should be either in English or in Japanese. If you are interested in participating, please contact the research team.

For more details, see the special project homepage: https://cefrjapan.net/kaken-5

Contact: jalt.cefrlp.sig@gmail.com

(3) Outcome of the research project now published.

One direct outcome of the second JSPS Grant-in-Aid research project (no. 16K02835) is the following book published in July 2020:
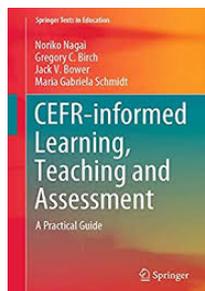
Nagai, Noriko, Gregory C. Birch, Jack V. Bower & Maria Gabriela Schmidt. 2020. *CEFR-informed Teaching, Learning, Assessment – A Practical Guide.* Singapore: Springer.

For more details, see the special book homepage: https://cefrjapan.net/publications/books/85-books-practical-guide

If you are interested in reviewing this book, please contact the editors of this volume.

By the way, another outcome of the first JSPS Grant-in-Aid research project (no. 26370624) is the following book with case studies:
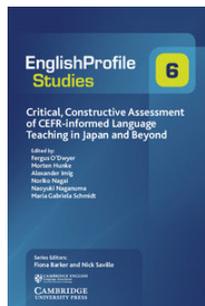
O'Dwyer, Fergus, Morten Hunke, Alexander Imig, Noriko Nagai, Naoyuki Naganuma & Maria Gabriela Schmidt (eds.). 2017. *Critical, Constructive Assessment of CEFR-informed Language Teaching in Japan and Beyond*. English Profile Studies, volume 6. Cambridge (UK): Cambridge University Press.

If you are interested in reviewing this book, please contact the editors of this volume.

If you want to share something which is related to CEFR or CEFR/CV research or practice, please submit it to the news section of CEFR & LP SIG.

We constantly try to reach out to other peers and other groups interested in and using the CEFR. If we do not help each other and work together, who else will support us?

Links

- CERF & LP SIG: https://cefrjapan.net
- CEFR Journal: https://cefrjapan.net/journal
- CEFR events: https://cefrjapan.net/events
- JALT: https://jalt.org/
- Language Portfolio for Japanese University, bilingual (English/Japanese): https://sites.google.com/site/flpsig/flp-sig-home/language-portfolio-for-japanese-university

CEFR JOURNAL—RESEARCH AND PRACTICE
VOLUME 3

Submission (Call for Abstracts)

This journal attempts to fall somewhere in between an inaccessible academic journal (long waiting times, fairly strict guidelines/criteria) and a newsletter (practical in nature but lacking in theoretical support/foundation), linking research of a practical nature with relevant research related to foreign language education, the CEFR, other language frameworks, and the European Language Portfolio. While the CEFR was introduced by the Council of Europe and intended for use, first and foremost, within Europe, the influence of the CEFR now has to be attested in many places beyond European borders. It has become a global framework, impacting a variety of aspects of language learning, teaching, and assessment across countries and continents beyond the context for which it was originally created. As such, there is a pressing need to create a quality forum for sharing research, experiences, and lessons learned from applying the CEFR in different contexts. This journal provides such a forum where people involved or interested in processes of applying the CEFR can share and learn from one another.

We are continuously seeking contributions related to foreign language education, the CEFR, other language frameworks, and the European Language Portfolio. We are particularly interested in specific contextual adaptations.

Currently, we have a new Call for Abstracts out. Due to current necessities and demand, we are looking to give your experiences with **online, remote, and e-learning in conjunction with the CEFR, the CEFR/CV, or portfolio work** the spotlight it deserves. In these months many practitioners are accruing valuable best and potentially also worst practice experience. We would like to offer a forum to share such valuable insights in future volumes. Until 30 November 2020 we are looking for abstracts at:

journal@cefrjapan.net

Guidelines

Submission:	30 November 2020
Contributions:	Articles (research), reports (best practice), news (work in progress), research notes, book reviews
Language(s):	English (British, American, international) preferred, but not mandatory. Other languages by request, with an extended abstract in English.
Review type:	Peer review, double blind

Peer review guidelines:

We ask all peer reviewers to make every reasonable effort to adhere to the following ethical guidelines for the **CEFR Journal – Research and Practice** submissions that they have agreed to review:

1. Reviewers must give unbiased consideration to each manuscript submitted for consideration for publication, and should judge each on its merits. Since, we employ a double-blind review, the text you have been provided with ought to have no reference to race, religion, nationality, sex, gender, seniority, or institutional affiliation of the author(s). Please, notify us immediately were any such information still detectable in the anonymised text you received.

2. Reviewers should declare any potential conflict of interest prior to agreeing to review a manuscript, including any relationship with the author that may potentially bias their review.

3. Reviewers are strongly advised to keep the peer review process confidential; information or correspondence about a manuscript should not be shared with anyone outside the peer review process.

4. Reviewers should provide a constructive, comprehensive, evidenced, and appropriately substantial peer review report. For your convenience, we are providing you with a 'reviewing matrix' you may choose to use at your own discretion. We would also like to kindly ask you to provide us in the journal editorial team with a final overall assessment of the text's publication potential – please, see bottom of this document.

5. Reviewers must avoid making statements in their report, which might be construed as impugning any person's reputation.

6. Reviewers should make all reasonable effort to submit their report and recommendation in a timely manner, informing the editor if this is not possible.

7. Reviewers should call to the journal editor's attention any significant similarity between the manuscript under consideration and any published paper or submitted manuscripts of which they are aware.

Author instructions:

- Adapted version of deGruyter Mouton guidelines for Language Learning in Higher Education (CercleS) and style sheet.