

TEAP4 技能(リーディング・リスニング・スピーキング・ライティング)スコアと CEFR レベル対応付け

調査報告書

2018 年 2 月

公益財団法人 日本英語検定協会

目次

調査概要 p3

調査背景 p3

TEAP 特色 p3

TEAP タスク開発時点での CRELLA との連携 p4

TEAP スコア概要 p4

これまでに行われた TEAP と他テストとの相関性検証 p4

スタンダードセッティング手法 p4

対応付けパネリスト p5

事前トレーニング p5

対応付け当日のアクティビティ p5

分析 p6

結果 p7

考察 p7

参考文献 p7

付録 1. 事前トレーニング資料 p8

付録 2. TEAP4 技能対応付け スケジュール p33

調査概要

CEFR(The Common European Framework of Reference)はリーディング・リスニング・ライティングとスピーキングの英語能力を6つのレベル(C2, C1, B2, B1, A2, A1)に分けて記述するものである。本研究においては、事前にCEFRやTEAPに関する概要やタスク設計に関してトレーニングを行った英語教育やテスト開発に関連する専門家(各技能13名)がTEAP4技能それぞれのスコアをCEFR各レベルに対応付けを行った。リーディングとリスニングはブックマーク法を用い、ライティングとスピーキングにはコントラスティング・グループ法を用いて、Below A2/A2、A2/B1、B1/B2、B2/Above B2の4つの境界に当たるスコアを各技能で定めた。

調査背景

試験をCEFRに対応付けるとは、複数名の専門家がCEFR及びターゲットとなる試験に対するトレーニングを経た上で、CEFR各レベルに対応するターゲットとなる試験のスコアに対して一つまたは複数のスコア切断点を設定する「スタンダードセッティング」というプロセスとExternal Validationと呼ばれる外部指標との比較検証等の結果を組みあわせて、その時点でのCEFR各レベルに対応する試験のスコアを判定することを前提としている。CEFRと試験を対応付けることは試験自体が変更になった場合、または比較検証を行った場合を含め継続的に行うことを前提としている。

TEAP(Test of English for Academic Purposes, 以下TEAP)はイギリスのテスト研究機関CRELLA(Centre for Research in English Language Learning and Assessment)との共同研究に基づき、アカデミックな環境で英語を用いて学ぶために必要な4つの技能(リーディング・リスニング・スピーキング・ライティング)をそれぞれ測定するために開発されたテストであり、CEFR A2-B2範囲を測定できるようタスク設計されている

(<http://www.eiken.or.jp/teap/group/report.html>)。

今回TEAPをCEFRに対応付けするにあたっては、欧州評議会が2009年に発行したRelating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (以下、Manual)に提唱される5つのプロセス(Familiarization, Specification, Standardization training/benchmarking, Standard setting, Validation)を経て2018年2月に以下の通り行われた。

TEAP 特色

TEAPは主に高校3年生を対象とした大学入試を想定して開発された。テスト構成は日本における「大学教育レベルにふさわしい英語力」を測定するよう設計され、テスト内容はすべて大学教育(留学も含む)で遭遇する場面を考慮して作成されている。難易度の目安としては、CEFR A2~B2 また英検準2級~準1級程度である。リーディング(試験時間70分)・リスニング(試験時間50分)はマークシートによる択一選択方式、ライティング(試験時間70分)は解答用紙への記入、スピーキング(試験時間10分)は一對一の面接方式を用いている。

TEAP タスク開発時点での GRELLA との連携

GRELLA とはこれまでにリーディング・リスニング (Taylor, 2014)、ライティング (Weir, 2014)、スピーキング (Nakatsuhawa, 2014) においてそれまでに行われたフィールドトライアルテストのテストセットや受験者解答データを元に、タスク設計及び評価観点の面から CEFR との関係性を検証してきた。また TEAP が大学入試として用いられる際の教育現場への波及効果 (Green, 2014; Nakamura, 2014) を高校・大学における英語指導関係者及び大学生・高校生へのアンケート調査を通して、検証を行った。

TEAP スコア概要

TEAP で測定する 4 技能それぞれのスコアは項目反応理論 (IRT : Item Response Theory、以下 IRT) に基づいて標準化された 100 点満点、4 技能合計で 400 点満点となるよう設計されている。また、異なる回次で受験して取得したスコアは比較可能である。

これまでに行われた TEAP と他テストとの相関性検証

これまでに TEAP は英検・センター試験・TOEFL ITP・TOEFL iBT 等の他の英語試験とその比較検証を行っており、全て同一受験者が TEAP とその他の試験を受験した得られたデータが得られるよう設計された。これは Manual にも定義されている External Validation に該当し、今回行われた対応付けと合わせて最終的な CEFR 各レベルに対応する TEAP スコアを決定した。

スタンダードセッティング手法

スタンダードセッティングを行う方法は Manual においても多くのものが提案されており、対応付けを行う試験の置かれた状況等に応じて、適した方法を用いるのが良いとされている。今回の対応付けにおいて、リーディングとリスニングはブックマーク法を用い、ライティングとスピーキングにはコントラスティング・グループ法を用いて、Below A2/A2、A2/B1、B1/B2、B2/Above B2 の 4 つの境界に当たるスコアを各技能で定めた。

ブックマーク法とは、選択式解答のような正誤がつく問題や多段階評点がつくような問題に適しているとされ、TEAP リーディング・リスニングが選択式解答を行う形式テストであるため、今回の対応付け方法として採用した。具体的には CEFR 対象レベル (A2, B1, B2, C1) に最低限該当する受験者 (minimally competent person、以下 MCP) の共通イメージを持った各パネリストが該当技能の問題が難易度順に掲載された問題冊子 (Ordered Item Booklet, 以下 OIB) を用いて 1 題ずつその難易度を確認し、指定された CEFR 対象レベル相当の MCP が予め定められた正答率 (response probability, 以下 RP) で正答できる問題を指定する方法を用いた。Manual においては、RP を正答率 3 分の 2 に該当する約 67% を用いることが提案されているので、今回はリーディング・リスニング共に RP を 67% (3 分の 2) と設定した。また今回 OIB はリーディング・リスニング共に実際の試験で出題する問題設計にあわせてそれぞれ 60 問・50 問を IRT を用いて推定された難易度の昇順にソートして作成した。

コントラスティング・グループ法とは、受験者解答に対して CEFR レベル判定を行う方法で、Manual においても提案されている手法の一つであり、TEAP ライティング・スピーキングの対応付け方法として採用した。具体的には過去の試験を実際に受験し、予め TEAP スコアが付与されている受験生の解答それぞれに対して、二つの隣接する CEFR レベルのいずれに該当するかを判定する方法を用いた。Manual においては、2つのカテゴリー分け判定を行う場合、または3つ以上のカテゴリー分け判定を行う場合等が提案されている。今回はスピーキング・ライティング両方とも、① Below A2/A2, ②A2/B1, ③B1/B2, ④B2/Above B2 の4つの境界に対してそれぞれ10名分の受験者解答を用意した。

対応付けパネリスト

4 技能それぞれの対応付けには各13名(全15名)が参加した。パネリストは大学・高校で英語指導を担当している者、または試験作成及び面接官/採点官としての経験があるものが参加した。全15名のうち、13名は英語母語話者、2名は日本語母語話者であった。

事前トレーニング(付録1参照)

パネリストは事前に以下の内容のトレーニング資料を精読し、与えられたタスクを終了することを義務付けられた。

- 1a: CEFR との対応付けを行う本プロジェクトの狙いを理解する。
- 1b: 他の参加者との交流を深めるため自己紹介を作成する。
- 2a: CEFR 概要や各レベルの記述文を理解する。
- 2b: CEFR に関する True/False クイズに取り組み理解を深める。
- 2c: ある CEFR レベルに該当する記述文に対してどのレベルに該当するかを当てはめるタスクに取り組み理解を深める。
- 2d: CEFR 各レベルの記述文を主な特徴を捉えながら、75語で要約し、理解を深める。
- 2e: CEFR 各レベルに該当する英語学習者が実生活において出来ていることを想像して記述することで、記述文と実際に出来ることとをリンクさせて理解する。
- 3a: CEFTTrain を活用して参加者自身の英語力を技能別に評価し、CEFR 記述文を用いた評価方法を理解する。
- 3b: CEFTTrain を活用して、CEFR 各レベル各技能の記述文を並び替えることで、より理解を深める。
- 3c/d: CEFTTrain を活用して、具体的な問題や英語学習者のパフォーマンスを評価する練習をする。
- 4a: TEAP の開発背景やターゲットとなる構成概念・英語力レベル・試験時間や各技能各タスクの構成概念、問題数、評価基準を理解する。

対応付け当日のアクティビティ(付録2参照)

・Familiarization

各技能対応付けを行う日程の最初に、事前トレーニングで予めパネリストが学習してきたことを復習した。具体的には CEFR の開発背景や各レベルの記述文や特徴に関して再度説明を行い、各レベルの記述文において特徴的な表現や内容を洗い出して、さらに理解を深めた。

・ CEFR training

事前トレーニングの理解を確認するため、再度各レベルの記述文を該当するレベルに当てはめることができるかの確認タスクを行った。また、各レベルの総合的な記述文に加え、各技能のさらに詳細な記述文をレベルごとに紹介し、その理解を深めた。

・ TEAP training

TEAP 各技能で使用するタスクを紹介し、過去に出題された問題に対して該当する CEFR レベルを判定するタスクを行い、グループや全体での議論を行い、共通のイメージを構築した。

・ Mini standard setting

最後に CEFR 各レベルに最低限相当する受験者 (Minimally Competent Person) のイメージを共有した上で、実際に過去の TEAP で出題された問題に対して、リーディング・リスニングに対して、Bookmark 法を用いて、どの CEFR レベルの英語学習者がおよそ 3 分の 2 の確率で正解できるかを判定し、グループや全体での議論を通して、共通のイメージを構築し、その手法を適用させる精度を高めた。またスピーキング・ライティングに対しては、Contrasting groups 法を用いて、実際の受験者のパフォーマンスが隣接する CEFR レベルのどこに該当するのかを判定し、グループや全体での議論を通して共通のイメージを構築し、その手法を適用させる精度を高めた。

・ Standard setting

リーディング・リスニングにおいては Bookmark 法を用いて、実際の試験で出題する問題設計にあわせてそれぞれ 60 問・50 問を IRT を用いて推定された難易度の昇順にソートして用意された問題冊子のどの問題で①Below A2/A2, ②A2/B1, ③B1/B2, ④B2/Above B2 の 4 つの境界を設定できるかを各参加者が判定した。

スピーキング・ライティングにおいては Contrasting groups 法を用いて、予め TEAP スコアが付与されている受験生の解答それぞれに対して、二つの隣接する CEFR レベルのいずれに該当するかを①Below A2/A2, ②A2/B1, ③B1/B2, ④B2/Above B2 の 4 つの境界に対してそれぞれ 10 名分の受験者解答を用意し、CEFR レベル判定を行った。

分析

最終的に行われた対応付けの結果はリーディング・リスニングに対しては、①Below A2/A2, ②A2/B1, ③B1/B2, ④B2/Above B2 の 4 つの境界にあたる問題とその問題に凡そ 3 分の 2 の確率で正答率のために必要な TEAP スコアを対応付けに参加した者全員分の集計を行い、平均点や標準偏差を算出した。スピーキング・ライティングに対しては、①Below A2/A2, ②A2/B1, ③B1/B2, ④B2/Above B2 の 4 つの境界において、それぞれのレベルと判定された受験者の TEAP スコアを対応付けに参加した者全員分の集計を行い、平均点を算出した。リーディング・リスニングに関しては、各パネリストの判定した上記 4 つの境界におけるアイテムに該当する TEAP スコアが 2 標準偏差を超えたも

のは外れ値として扱い、削除して最終的な閾値を算出した。具体的にはリスニング

- ① Below A2/A2 判定において 2 名分、④B2/Above B2 判定において 1 名分を削除し、リーディング Below A2/A2 判定において 1 名分を削除した。スピーキング・ライティングはコントラスト法を用い、パネリストの判定は各受験者のパフォーマンスを隣接する CEFR レベルのどちらか 2 択で判定するため、外れ値の検出は行わなかった。

結果

上記スクリーニング後、パネリスト全員の技能別判定を集計した結果、以下の通りとなった。

CEFR	A2	B1	B2	C1
Total	135	225	309	375

考察

先述の対応付け方法を用いて算出された TEAP 各技能の閾値は上記の通りとなった。これまでに行われた他の英語試験との比較検証結果も踏まえ、4 技能毎の閾値を足し合わせて 4 技能合算スコアにおける CEFR 閾値を上記の通り設定した。

参考文献

- 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.
- Martyniuk, W. (ed.). 2010. *Aligning Tests with the CEFR*. Cambridge University Press.
- Taylor, L. 2014. *A Report on the Review of Test Specifications for the Reading and Listening Papers of the Test of English for Academic Purposes (TEAP) for Japanese University Entrants*.
- Green, A. 2014. *The Test of English for Academic Purposes (TEAP) Impact Study: Report 1- Preliminary Questionnaires to Japanese High School Students and Teachers*.
- Nakamura, K. 2014. *Examination of Possible Consequences of new Test within the Context of University Entrance Exam Reform in Japan*.
- Weir, C.J. 2014. *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Writing Test for Japanese University Entrants*.
- Nakatsuhara, F. 2014. *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants-Study 1&2*.
- Nakatsuhara, F. 2014. *A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Speaking Test for Japanese University Entrants-Study 3&4*.

付録 1. 事前トレーニング資料

Relating the Test of English for Academic Purposes (TEAP) to the Common European Framework of Reference for Languages (CEFR)

Preparation Booklet

Introduction

Welcome to the Eiken Foundation of Japan's project for relating the TEAP to the CEFR.

This booklet contains a series of training tasks that you must complete before attending on-site workshops. **The instructions for each task will be written in blue.**

We will ask you to email this completed booklet to us in two stages: once after you have completed tasks 1a through 2e, and once again at the end of the training period (see the schedule below for details).

Please type your answers and comments directly in this booklet as you go through each of the tasks. The exception to this is task 1b (self-introduction), which should be emailed to Eiken separately. Please email us this self-introduction as soon as possible.

Deadlines for the Home Training Tasks

It is very important that you are well prepared in advance for the actual Standard Setting workshops. We have provided a schedule for the Home Training. Please pace yourself and complete the tasks based on the schedule below so that your knowledge will be accumulated and you will have enough time to reflect on the material and email the completed tasks on time.

Please adhere to the following schedule. It has been set so that you avoid cramming all of the tasks into a short period.

Task Completion Schedule

1st period: Jan 12-19	Tasks (1a – 2e)
Jan 19	Email booklet to Eiken with above tasks completed
2nd period: Jan 20-28	Tasks (3a – 4c)
Jan 28	Email completed booklet to Eiken

Checklist for Training Tasks (Table of Contents)

Below is a table containing a list of all the tasks you will be required to complete for this preparatory training. Please check them off as you complete each one.

Check✓	Your Tasks	Page
	<u>Task 0a: Read a description of the CEFR project</u>	
	<u>Task 1b: Self-introduction</u>	
	<u>Task 2a: Read an overview of the CEFR, the CEFR Global Scale, and an excerpt from the CEFR official document</u>	
	<u>Task 2b: True/false comprehension check</u>	
	<u>Task 2c: Descriptor and level check</u>	
	<u>Task 2d: Write CEFR level summaries</u>	
	<u>Task 2e: Real-world tasks</u>	
	<u>Task 3a: Assess your own foreign language proficiency (CEFTrain website)</u>	
	<u>Task 3b: Reconstruct CEFR scales (CEFTrain website)</u>	
	<u>Task 3c: Practice relating items to CEFR (Part 1) (CEFTrain website)</u>	
	<u>Task 3d: Practice relating items to CEFR (Part 2) (CEFTrain website)</u>	
	<u>Task 4a: Read about TEAP</u>	
	<u>Task 4b: Complete a quiz about TEAP</u>	
	<u>Task 4c: Review TEAP materials (sample items, rating criteria, etc.)</u>	

1. About the CEFR Project

[Task 1a: Read a Description of the CEFR Project](#)

About CEFR

The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (Council of Europe, 2001) is a framework of reference that describes foreign language proficiency at six levels, ranging from beginner to advanced. It was developed by the Council of Europe in order to provide an objective, transparent, and cohesive planning tool in language education. The CEFR is widely used by language teaching practitioners in Europe and around the world as a reference for language curriculum design, material development, and assessment.

Project Goal

The goal of this project is to investigate the relationship between TEAP and the CEFR.

Procedure

- ① Home Training: Panelists complete preparation assignments before the on-site workshops in order to become familiar with the CEFR bands and descriptors and with TEAP.
- ② Standardization Training: Panelists attend the on-site workshops and do more training activities, using additional descriptor scales and assessing sample test items for practice.
- ③ Standard Setting: Panelists review TEAP test items and examinee performances and place them into the appropriate CEFR level.

The above procedure is based on recommendations in *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*. (Council of Europe, 2009).

[Task 1b: Email a self-introduction](#)

Please send us an email in which you introduce yourself in 100 words or less. As we will all be working closely together during the workshops, let's get to know each other. Please include such information as where you are from, where you work, your professional background, and anything else you would like to share. Your self-introduction will be shared with the other panelists via email.

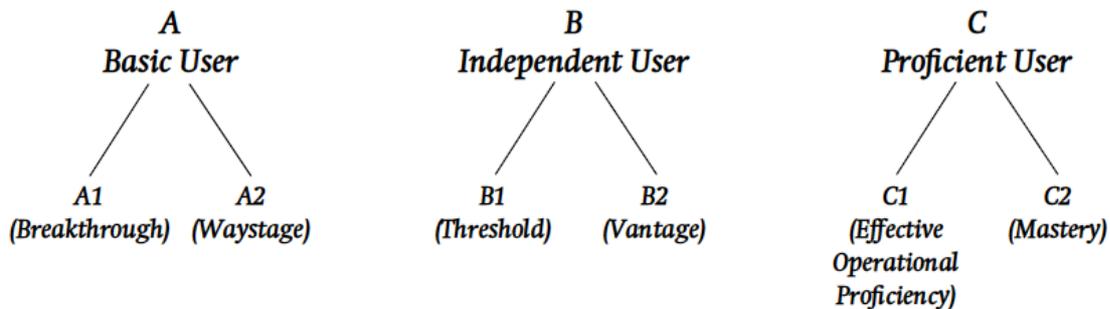
Please email your self-introduction to standardsetting@eiken.or.jp before proceeding on to the next section.

2. Overview of the CEFR and its Levels

[Task 2a: Read an overview of the CEFR, the CEFR Global Scale, and an excerpt from the official CEFR document.](#)

You are encouraged to take notes and highlight key points; this will be useful when completing subsequent tasks.

The CEFR describes what language learners are capable of doing in terms of the four skills—reading, writing, listening, and speaking—at different stages of progress. The six broad stages (or “Common Reference Levels”) of the CEFR are represented in the following diagram (Council of Europe, 2001, p.23):



In the on-site workshops, we will be working with the CEFR Global Scale as well as a number of the Illustrative Scales. The Global Scale describes holistically what language learners can do at each level. The Illustrative Scales present a comprehensive breakdown of the broad scales into subscales representing descriptive categories for each of the four skills. The Illustrative Scales also include “plus levels” (A2+, B1+, B2+), which generally represent “stronger” performance levels within the main bands. You will see the descriptors for these plus levels to some extent in the online training and in the on-site workshops, but please note that the focus of the standard setting will be on the six main CEFR levels.

Table 1: CEFR Global Scale (Council of Europe, 2001, p.24)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce himself/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

The following description of the CEFR uses descriptors taken from the Illustrative Scales, which you will work with in the online training materials further down in this booklet, as well as in the on-site workshop.

Excerpt from the official CEFR document (Council of Europe, 2001, pp.33-36)

- Level A1 (Breakthrough) – is considered the lowest level of generative language use - the point at which the learner can *interact in a simple way, ask and answer simple questions about themselves, where they live, people they know, and things they have, initiate and respond to simple statements in areas of immediate need or on very familiar topics*, rather than relying purely on a very finite rehearsed, lexically organized repertoire of situation-specific phrases.

- Level A2 does appear to reflect the level referred to by the Waystage specification. It is at this level that the majority of descriptors stating social functions are to be found, like *use simple everyday polite forms of greeting and address; greet people, ask how they are and react to news; handle very short social exchanges; ask and answer questions about what they do at work and in free time; make and respond to invitations; discuss what to do, where to go and make arrangements to meet; make and accept offers*. Here too are to be found descriptors on getting out and about: the simplified cut-down version of the full set of transactional specifications in 'The Threshold Level' for adults living abroad, like: *make simple transactions in shops, post offices or banks; get simple information about travel; use public transport: buses, trains, and taxis, ask for basic information, ask and give directions, and buy tickets; ask for and provide everyday goods and services*.

- The next band represents a strong Waystage (A2+) performance. What is noticeable here is more active participation in conversation given some assistance and certain limitations, for example: *initiate, maintain and close simple, restricted face-to-face conversation; understand enough to manage simple, routine exchanges without undue effort; make him/herself understood and exchange ideas and information on familiar topics in predictable everyday situations, provided the other person helps if necessary; communicate successfully on basic themes if he/she can ask for help to express what he wants to; deal with everyday situations with predictable content, though he/she will generally have to compromise the message and search for words; interact with reasonable ease in structured situations, given some help, but participation in open discussion is fairly restricted*; plus significantly more ability to sustain monologues, for example: *express how he/she feels in simple terms; give an extended description of*

everyday aspects of his/her environment e.g. people, places, a job or study experience; describe past activities and personal experiences; describe habits and routines; describe plans and arrangements; explain what he/she likes or dislikes about something; give short, basic descriptions of events and activities; describe pets and possessions; use simple descriptive language to make brief statements about and compare objects and possessions.

- Level B1 reflects the Threshold Level specification for a visitor to a foreign country and is perhaps most categorized by two features. The first feature is the ability to maintain interaction and get across what you want to, in a range of contexts, for example: *generally follow the main points of extended discussion around him/her, provided speech is clearly articulated in standard dialect; give or seek personal views and opinions in an informal discussion with friends; express the main point he/she wants to make comprehensibly; exploit a wide range of simple language flexibly to express much of what he or she wants to; maintain a conversation or discussion but may sometimes be difficult to follow when trying to say exactly what he/she would like to; keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.* The second feature is the ability to cope flexibly with problems in everyday life, for example *cope with less routine situations on public transport; deal with most situations likely to arise when making travel arrangements through an agent or when actually travelling; enter unprepared into conversations on familiar topics; make a complaint; take some initiatives in an interview/consultation (e.g. to bring up a new subject) but is very dependent on interviewer in the interaction; ask someone to clarify or elaborate what they have just said.*

- The subsequent band seems to be a Strong Threshold (B1+). The same two main features continue to be present, with the addition of a number of descriptors which focus on the exchange of quantities of information, for example: *take messages communicating enquiries, explaining problems; provide concrete information required in an interview/consultation (e.g. describe symptoms to a doctor) but does so with limited precision; explain why something is a problem; summarize and give his or her opinion about a short story, article, talk, discussion, interview, or documentary and answer further questions of detail; carry out a prepared interview, checking and confirming information, though he/she may occasionally have to ask for repetition if the other person's response is rapid or extended; describe how to do something, giving detailed instructions; exchange accumulated factual information on familiar routine and non-routine matters within his/her field with some confidence.*

- Level B2 represents a new level as far above B1 (Threshold) as A2 (Waystage) is below it. It is intended to reflect the Vantage Level specification. The metaphor is that, having been progressing slowly but steadily across the intermediate plateau, the learner finds he has arrived somewhere, things look different, he/she acquires a new perspective, can look around him/her in a new way. This concept does seem to be borne out to a considerable extent by the descriptors calibrated at this level. They represent quite a break with the content so far. For example at the lower end of the band there is a focus on effective argument: *account for and sustain his opinions in discussion by providing relevant explanations, arguments and comments; explain a viewpoint on a topical issue giving the advantages and disadvantages of various options; construct a chain of reasoned argument; develop an argument giving reasons in support of or against a particular point of view; explain a problem and make it clear that his/her counterpart in a negotiation must make a concession; speculate about causes, consequences, hypothetical situations; take an active part in informal discussion in familiar contexts, commenting, putting point of view clearly, evaluating alternative proposals and making and responding to hypotheses.* Secondly, running right through the level there are two new focuses. The first is being able to more than hold your own in social discourse: e.g. *converse naturally, fluently and effectively; understand in detail what is said to him/her in the standard spoken language even in a noisy environment; initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly; use stock phrases (e.g. 'That's a difficult question to answer') to gain time and keep the turn whilst formulating what to say; interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without imposing strain on either party; adjust to the changes of direction, style and emphasis normally found in conversation; sustain relationships with native speakers without unintentionally amusing or irritating them or requiring them to behave other than they would with a native speaker.* The second new focus is a new degree of language awareness: *correct mistakes if they have led to misunderstandings; make a note of 'favorite mistakes' and consciously monitor speech for it/them; generally correct slips and errors if he/she becomes conscious of them; plan what is to be said and the means to say it, considering the effect on the recipient/s.* In all, this does seem to be a new threshold for a language learner to cross.

- At the next band – representing a Strong Vantage (B2+) performance – the focus

on argument, effective social discourse and on language awareness which appears at B2 (Vantage) continues. However, the focus on argument and social discourse can also be interpreted as a new focus on discourse skills. This new degree of discourse competence shows itself in conversational management (co-operating strategies): *give feedback on and follow up statements and inferences by other speakers and so help the development of the discussion; relate own contribution skillfully to those of other speakers*. It is also apparent in relation to coherence/cohesion: *use a limited number of cohesive devices to link sentences together smoothly into clear, connected discourse; use a variety of linking words efficiently to mark clearly the relationships between ideas; develop an argument systematically with appropriate highlighting of significant points, and relevant supporting detail*. Finally, it is at this band that there is a concentration of items on negotiating: *outline a case for compensation, using persuasive language and simple arguments to demand satisfaction; state clearly the limits to a concession*.

- Level C1, the next band, was labelled Effective Operational Proficiency. What seems to characterize this level is good access to a broad range of language, which allows fluent, spontaneous communication, as illustrated by the following examples: *Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language*. The discourse skills characterizing the previous band continue to be evident at Level C1, with an emphasis on more fluency, for example: *select a suitable phrase from a fluent repertoire of discourse functions to preface his remarks in order to get the floor, or to gain time and keep it whilst thinking; produce clear, smoothly flowing, well-structured speech, showing controlled use of organizational patterns, connectors and cohesive devices*.

- Level C2, whilst it has been termed 'Mastery', is not intended to imply native-speaker or near native-speaker competence. What is intended is to characterize the degree of precision, appropriateness and ease with the language which typifies the speech of those who have been highly successful learners. Descriptors calibrated here include: *convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices; has a good command of idiomatic expressions and colloquialisms with awareness of connotative level of meaning; backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it*.

Task 2b: True/False Comprehension Check

Please complete the true/false quiz below. For each question, choose true or false using the drop-down list. You can check your answers in the appendix on the last page of this booklet.

1. Users at the B1 level can write a well-developed argumentative essay on a complex topic.

True / False

2. Users at the A1 level can understand the main points in a company meeting about changes in workplace policy.

True / False

3. Users at the B2 level can explain to a salesperson why they deserve a refund for an electronic appliance.

True / False

4. Users at the C1 level can give an effective sales presentation to potential clients and respond smoothly to subsequent questions.

True / False

5. Users at the A2 level can explain to a friend why they recommend a certain film.

True / False

6. Users at the B1 level can arrange new airline flights with a travel agent after a cancellation.

True / False

7. Users at the A1 level can conduct a job interview.

True / False

8. Users at the C2 level can compete in a parliamentary debate competition.

True / False

Task 2c: Descriptor and Level Check

Please complete the quiz below. For each question, read the descriptor and choose which

CEFR level it matches using the drop-down list. You can check your answers in the appendix.

1. Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. **Choose CEFR level**
2. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices. **Choose CEFR level**
3. Can introduce himself/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. **Choose CEFR level**
4. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. **Choose CEFR level**
5. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. **Choose CEFR level**
6. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. **Choose CEFR level**
7. Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. **Choose CEFR level**
8. Can use language flexibly and effectively for social, academic and professional purposes. **Choose CEFR level**
9. Can produce simple connected text on topics which are familiar or of personal interest. **Choose CEFR level**
10. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need. **Choose CEFR level**
11. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. **Choose CEFR level**
12. Can understand with ease virtually everything heard or read. **Choose CEFR level**

Task 2d: Write CEFR Level Summaries

Briefly summarize each CEFR level in 75 words or less. Consider these questions:

- What are the key features that set each level apart from adjacent levels?
- Can you identify any key words?

Short summaries are better (bullet points are acceptable).

Level	Summary
-------	---------

C2	
C1	
B2	
B1	
A2	
A1	

Task 2e: Real-world tasks

- For each CEFR Level, please give two real world tasks that a language learner should be able to accomplish at that level.
 - For example: A1-Introduce themselves to a new acquaintance
 - For example: C2-Write a letter to an elected official detailing why a new policy might be ineffective.

***When you have completed this task, please email this booklet to**

standardsetting@eiken.or.jp before proceeding on to Part 3.

Level	Real-world task
C2:	1.
	2.
C1:	1.
	2.
B2:	1.
	2.
B1:	1.
	2.
A2:	1.
	2.
A1:	1.
	2.

3. Training on the CEFRTrain Website

The Council of Europe has created a website to help CEFR users become more familiar with the CEFR descriptor scales. In this part of the training, you will complete three activities on this website. This should help you understand which texts and activities match each CEFR level.

Please note that the purpose of this section is for you to gain a deeper understanding of various CEFR scale descriptors. The tasks in this section are somewhat similar to, but NOT exactly the same as, the ones we are going to do in the on-site workshops. Moreover, not all the CEFR scales you will use in the online tasks will be used in the workshops, but they will help to improve your understanding of the CEFR.

Task 3a: Assess Your Own Foreign Language Proficiency

Access the following page on the CEFRTrain website and complete the self-assessment activity.

<http://www.helsinki.fi/project/ceftrain/index.php.190.html>

Input your results in the table below using the pulldown menus.

Skills:	(Example)	Listening	Reading	Spoken	Spoken	Writing
---------	-----------	-----------	---------	--------	--------	---------

	Listening			Interaction	Production	
Your Level:	B1	Choose your level				

Then, write some of your reflections on the self-assessment task in the space below. Please limit your comments to around 200 words.

Self-Assessment Task Reflections:

Some questions to consider:

- How was assessing yourself different from assessing someone else?
- Do you think self-assessment is useful for helping one to understand the CEFR better? Please give reasons why you think so.
- Which CEFR descriptors do you feel are ambiguous, and why? If you could make those descriptors clearer, how would you define them?
- What difficulties did you encounter when using the CEFR descriptor scales?

Your Comments:

Task 3b: Reconstructing CEFR scales

Access the following page on the CEFTrain website and complete the restructuring activities

<http://www.helsinki.fi/project/ceftrain/index.php.191.html>

Task 3c: Practice Relating Items to the CEFR (Part 1)

Access the following pages on the CEFTrain website and practice relating items/performances to

CEFR levels:

<http://www.helsinki.fi/project/ceftrain/index.php.192.html>

(The tasks may take a few seconds to load, so please be patient).

Relating Task Part 2 Reflections

Now, in the space below, in 200 words or less, write some of your reflections on the relating task. Can you identify what might have led you to assess inaccurately?

- Which CEFR descriptors do you feel are ambiguous, and why? If you could make those descriptors more specific, how would you define any ambiguous terms?

Your Comments:

Task 3d: Practice Relating Items to the CEFR (Part 2)

Access the following pages on the website and practice relating items/performances to CEFR levels. This task is more complicated as each item/performance needs to be related to several different CEFR scales.

Please visit the links below to access the items.

Listening:

<http://www.helsinki.fi/project/ceftrain/files/listening3.swf>

Speaking:

http://www.helsinki.fi/project/ceftrain/files/speaking_4.swf

http://www.helsinki.fi/project/ceftrain/files/speaking_8.swf

Reading:

http://www.helsinki.fi/project/ceftrain/files/reading_7.swf

Writing:

http://www.helsinki.fi/project/ceftrain/files/writing_2.swf

http://www.helsinki.fi/project/ceftrain/files/writing_9.swf

Relating Task Part 2 Reflections

- What factors in the items were most influential on your decisions?
- What do you think was the most challenging part of the task?
- Can you identify what might have led you to assess inaccurately?

Your comments:

4. Introduction to TEAP

In order to effectively relate TEAP items to the CEFR, you will also need to be familiar with the TEAP test.

[Task 4a: Read about TEAP](#)

General Introduction

The Test of English for Academic Purposes (TEAP) is a four-skill (reading, listening, writing, and speaking) test that began as a collaborative project between the Eiken Foundation of Japan and Sophia University. Later, the Centre for Research in English Language Learning and Assessment (CRELLA), one of the world's leading language-testing research centers, was also enlisted to help with test development. From the outset, TEAP was intended to have the potential for wider application beyond one institution. Another long-term aim of the project was to have a positive impact on English education in Japan by revising and improving the widely varying approaches to English tests used in university admissions and by serving as a benchmark standard of the English skills needed by Japanese students to study at the university level in the English as a foreign language (EFL) context of Japan. As such, the tasks included in the test are designed to reflect tasks that Japanese students are likely to encounter when participating in academic activities in English-medium universities in Japan. TEAP is now accepted for admissions purposes by over 90 universities across Japan.

The levels measured by TEAP

TEAP used the CEFR as a reference point for defining relevant levels of proficiency and as a springboard from which TEAP-specific rating descriptors appropriate for the academic context of Japan could be developed. It was felt that bringing the CEFR into the test design would facilitate stakeholders' understanding of the test scores and task requirements. It was also considered important to report scores not only as scale scores but in terms of bands which indicate to test takers their approximate level in terms of some external criterion, and the CEFR offered a practical way to do this.

TEAP also took into consideration the guidelines published by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). MEXT has publicly recommended two levels of English proficiency as goals for high school graduates, and it has stated these goals in terms of commonly recognized English proficiency benchmarks, including the EIKEN grade levels. For high school students, MEXT has suggested the EIKEN Grade 2 and Pre-2 levels as appropriate benchmarks. Based on research into the comparability of the EIKEN grades with the CEFR, these levels of proficiency can be considered relevant to the B1 and A2 levels, respectively (See Eiken website for further information).

The key focus of TEAP is therefore a level of proficiency relevant to the B1 level of ability as defined in the CEFR. In addition, TEAP is designed to provide useful feedback to students at the A2 level. At the same time, in order to look forward to the more demanding academic context of English-medium universities in Japan, TEAP also contains tasks capable of discriminating between students at a B1 level and the more advanced B2 level. This allows for the provision of useful feedback to students at this more advanced level of ability. TEAP test tasks are thus designed to broadly span the CEFR A2, B1 and B2 levels.

Levels measured by TEAP:

CEFR	EIKEN	MEXT ¹ benchmarks	TOEFL® iBT ²	TEAP
C2				
C1	1		100	
B2	Pre-1	English teachers	80	
B1	2	High school graduates	45	
A2	Pre-2	graduates	32	
A1	3, 4, 5			

1 Ministry of Education, Culture, Sports, Science and Technology

2 Based on comparability studies undertaken by Eiken at the University of Hawaii in conjunction with Professor J.D. Brown & Dr. Yao Hill. TOEFL® is a registered trademark of Educational Testing Service (ETS), Inc. This table has not been endorsed or approved by ETS.

Test Structure

TEAP has three modules: Reading and Listening (RL); Reading, Listening, and Writing (RL+W); and Reading, Listening, Writing, and Speaking (RL+W+S).

Module	Time
RL	90 minutes
RL+W	70 minutes
RL+W+S	Approx. 10 mins

※Test takers can select one of the three modules. W and S cannot be taken without taking RL.

Details of each section of the test are described in the following tables:

Reading

Part	Intended CEFR level	Skill focus	Task description
R1	A2—B2	Vocabulary and word usage	Read a short text from which a word or phrasal verb has been deleted and choose the best word or phrase to fill the gap. (Multiple choice with 4 answer options.)
R2A	A2—B1	Reading graphs and charts	Look at information displayed in a graph or chart and choose the best response to answer a question about the graph or chart. (Multiple choice with 4 answer options.)
R2B	A2—B1	Reading advertisements and notices	Look at the information displayed in a notice, announcement, or e-mail, etc. and choose the best response to answer a question about it. (Multiple choice with 4 answer options.)
R2C	A2—B2	Reading short texts	Read a short expository text and then choose the best response to answer a question about the text (one question per text). (Multiple choice with 4 answer options.)
R3A	B1—B2	Reading extended texts	Read longer texts from which several words and phrases have been deleted. Choose the best response to fill each gap. Gaps target discourse-level understanding and require reading across sentences and paragraphs. (Multiple choice with 4 answer options.)
R3B	B1—B2	Reading extended texts (including graphs and charts)	Read an extended argumentative or expository text and choose the best response to answer questions about it. (Multiple choice with 4 answer options.)

Listening

Part	Intended CEFR level	Skill focus	Task description
L1A	A2—B1	Listening to short dialogues	Listen to a short dialogue and choose the best response to answer a question about it. Dialogue and question are heard once. (Multiple choice with 4 answer options.)
L1B	A2—B1	Listening to short monologues	Listen to a short monologue and choose the best response to answer a question about it. Questions ask test takers to choose from 4 graphs or charts. Monologue and question are heard once. (Multiple choice with 4 answer options.)
L1C	A2—B1	Listening to short monologues describing charts or graphs	Listen to a short monologue and choose the best response to answer a question about it. Questions ask test takers to choose from 4 graphs or charts. Monologue and question are heard once. (Multiple choice with 4 answer options.)
L2A	B1—B2	Listening to long dialogues	Listen to a long dialogue and choose the best response to answer questions about it. Dialogue and question are heard once. (Multiple choice with 4 answer options.)
L2B	B1—B2	Listening to long monologues (including graphs and charts)	Listen to long monologues and answer 4 questions about each monologue. Monologue and question are heard once. (Multiple choice with 4 answer options.)

Writing

Task	Intended CEFR level	Skill focus	Task description
Task A	Input text: A2+ Task: B1	Conveying the main ideas of a text.	Summary writing. Extended response (about 70 words, structured as a paragraph).
Task B	Input texts: B1 Task: B2	Synthesizing and evaluating information and arguments from multiple verbal and nonverbal sources.	Essay writing. Extended response (about 200 words, structured as an essay).

TEAP Writing scoring system and rating scales:

TEAP Writing responses are double rated. Scores are assigned by two different raters. In the case of

significant discrepancy between the raters, responses are reviewed by a senior rater and/or a rating supervisor.

Task A: Scores are assigned based on four rating categories. There are four score bands for each category:

1. Main Ideas
2. Coherence and Cohesion
3. Grammatical Range and Accuracy
4. Lexical Range and Accuracy

Score bands
B2
B1
A2
Below A2

Task B: Scores are assigned based on five rating categories. There are four score bands for each category:

1. Main Ideas
2. Coherence
3. Cohesion
4. Grammatical Range and Accuracy
5. Lexical Range and Accuracy

Score bands
B2
B1
A2
Below A2

Speaking (face-to-face interview)

Part	Intended CEFR level	Task	Targeted language functions
1	A2	Interview	Providing specific personal information.
2	B1	Role play (test taker interviews examiner)	Initiating interaction; Asking for information/opinions; Commenting.
3	B1—B2	Monologue	Agreeing/disagreeing; Justifying opinions; Elaborating.
4	B2	Extended interview	Expressing opinions; Justifying opinions; Comparing; Speculating; Elaborating.

TEAP Speaking scoring system and rating scales:

TEAP Speaking performances are double-rated. Examiners assign holistic scores on the test day, and raters later listen to recordings of each test and assign analytic scores.

Scores are assigned based on five rating categories. There are four score bands for each category:

1. Pronunciation
2. Grammatical Range and Accuracy
3. Lexical Range and Accuracy
4. Fluency
5. Interactional Effectiveness

Score bands
B2
B1
A2
Below A2

[Task 4b: Complete a Quiz about TEAP](#)

After reading about TEAP, read each question and choose either true or false using the drop down menu.

- | | |
|---|--------------|
| (1) TEAP contain test tasks that cover all levels of the CEFR. | True / False |
| (2) TEAP is designed for university entrance purposes in Japan. | True / False |
| (3) TEAP measures all four skills (reading, listening, writing, speaking). | True / False |
| (4) All items in the TEAP Reading and Listening sections are multiple-choice questions. | True / False |
| (5) Part R1A is designed to test vocabulary and word usage. | True / False |
| (6) Part R2B includes graphs. | True / False |
| (7) Parts L1A and L2A involve listening to dialogues. | True / False |
| (8) Part L1B requires test takers to identify the correct graph or chart from four options. | True / False |
| (9) Task A in the Writing test is an essay-writing task. | True / False |
| (10) The TEAP Speaking test lasts for approximately 10 minutes. | True / False |
| (11) In Part 4 of the Speaking Test, test takers must give and justify opinions. | True / False |

[Task 4c: View Sample Items from the TEAP-CBT Test](#)

[Please access the links below to view TEAP sample items for each section of the test and the public rating scales for Writing and Speaking. "Task 6.3: View Sample Items from the TEAP](#)

Test.”

Reading and Listening

Sample test: http://media.eiken.or.jp/teap/reading/teap_sample_reading.pdf

Sample test listening script: http://media.eiken.or.jp/teap/listening/sampletest_script.pdf

Writing

Sample test: <http://media.eiken.or.jp/teap/writing/sampletest.pdf>

Sample response: http://www.eiken.or.jp/teap/construct/model_ans.html

Public rating criteria: http://www.eiken.or.jp/teap/construct/rating_crit.html

Speaking

Sample test: <http://www.eiken.or.jp/teap/construct/#rltest> (scroll down to “Speaking test”)

Sample video: <http://www.eiken.or.jp/teap/construct/#rltest> (scroll down to “見本問題 応答例” and click on “見本問題 応答例”)

Public rating criteria: http://www.eiken.or.jp/teap/construct/sp_rating_crit.html

For more information about TEAP, please also see the following links.

Research reports: <http://www.eiken.or.jp/teap/group/report.html>

TEAP Website (Japanese): <http://www.eiken.or.jp/teap/>

5. References

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.

Appendix

Answers to CEFR Quizzes

Task 2b quiz:

1. F
2. F
3. T
4. T
5. F
6. T
7. F
8. T

Task 2c quiz

1. A2
2. C1
3. A1
4. B1
5. C2
6. B2
7. A1
8. C1
9. B1
10. A2
11. B2
12. C2

付録 2. TEAP4 技能対応付け スケジュール

Timeline	Day 1: Reading
9:00-9:10	Introduction to the workshop <ul style="list-style-type: none"> About the CEFR What is standard setting?
9:10-10:40	Familiarization <ul style="list-style-type: none"> CEFR scales
10:40-10:50	Coffee break
10:50-12:00	CEFR Task SS Training <ul style="list-style-type: none"> CEFR benchmarks
12:00-12:50	Lunch
12:50-15:00	TEAP SS Training <ul style="list-style-type: none"> About TEAP TEAP benchmarks Explanation of the bookmark method Mini standard setting
15:00-15:10	Coffee break
15:10-16:20	Standard setting <ul style="list-style-type: none"> The bookmark method
16:20-16:30	Closing of Day 1
Timeline	Day 2: Listening
9:00-9:10	Introduction to the workshop <ul style="list-style-type: none"> About the CEFR What is standard setting?
9:10-10:40	Familiarization <ul style="list-style-type: none"> CEFR scales
10:40-10:50	Coffee break
10:50-12:00	CEFR Task SS Training <ul style="list-style-type: none"> CEFR benchmarks
12:00-12:50	Lunch
12:50-15:00	TEAP SS Training <ul style="list-style-type: none"> About TEAP TEAP benchmarks Explanation of the bookmark method Mini standard setting
15:00-15:10	Coffee break
15:10-16:20	Standard setting <ul style="list-style-type: none"> The bookmark method
16:20-16:30	Closing of Day 2

Timeline	Day 3: Writing
9:00-9:10	Introduction to the workshop <ul style="list-style-type: none"> About the CEFR What is standard setting?
9:10-10:15	Familiarization <ul style="list-style-type: none"> CEFR scales
10:15-10:25	Coffee break
10:25-11:30	CEFR Task SS Training <ul style="list-style-type: none"> CEFR benchmarks
11:30-12:00	TEAP SS Training Part 1: About TEAP
12:00-12:45	Lunch
12:45-14:10	TEAP SS Training Part 2 <ul style="list-style-type: none"> TEAP benchmarks Explanation of the Contrasting Groups Method
14:10-14:20	Coffee break
14:20-16:20	Standard setting
16:20-16:30	Closing of Day 3

Timeline	Day 3: Writing
9:00-9:10	Introduction to the workshop <ul style="list-style-type: none"> About the CEFR What is standard setting?
9:10-10:15	Familiarization <ul style="list-style-type: none"> CEFR scales
10:15-10:25	Coffee break
10:25-11:30	CEFR Task SS Training <ul style="list-style-type: none"> CEFR benchmarks
11:30-12:00	TEAP SS Training Part 1: About TEAP
12:00-12:45	Lunch
12:45-14:10	TEAP SS Training Part 2 <ul style="list-style-type: none"> TEAP benchmarks Explanation of the Contrasting Groups Method
14:10-14:20	Coffee break
14:20-16:20	Standard setting
16:20-16:30	Closing of Day 3