ALIGNING LANGUAGE EDUCATION WITH THE CEFR: A HANDBOOK

April 2022

BRITISH COUNCIL
UKALTA UK Association for Language Testing and Assessment
EALTA www.ealta.eu.org
ALTE Association of Language Testers in Europe

Aligning Language Education with the CEFR: A Handbook

This resource has been prepared for those who are teaching, testing and developing materials in language education, as well as stakeholders concerned with education policy matters and decision-making.

Given the huge impact of the CEFR 2001 and the interest raised by the CEFR CV 2020, the editors expect that this first edition of the Handbook will soon be put to use in alignment projects by many different stakeholders in the field of language education. Those involved in such projects are invited to share not only their experience but also their views on the usefulness of the Handbook. There are plans to organize an event in early 2024 at which case studies in the use of the Handbook will be presented. The findings of such studies and other proposals from the field will be incorporated in a revised edition of the Handbook.

The Handbook is published jointly by the European Association for Language Testing and Assessment (EALTA), the UK Association for Language Testing and Assessment (UKALTA), the British Council, and the Association for Language Testers in Europe (ALTE).

ISBN: 978-1-7397544-1-9

CONTENTS

List of figures

List of tables

Preface

The *[Common European Framework of Reference for Languages](#)* (CEFR) was first published in 2001. Its common reference levels and illustrative scales were welcomed, especially by language testing agencies, but soon language education specialists and stakeholders began to ask questions like "How do we know your B1 is the same as our B1?" This prompted calls for the Council of Europe to set up a system of validation that would confirm the alignment of language tests with one or more levels of the CEFR. Such a system, however, lay beyond the remit of the Council of Europe, which instead published two guides, one for aligning language tests with the CEFR[1] and the other for developing tests with reference to the CEFR.[2]

A project to update the CEFR's illustrative scales was launched in 2013. A provisional version of the CEFR Companion Volume (CEFR CV) appeared in 2017 and the definitive version in 2020. An EALTA Symposium held in Dublin in January 2018 and an EALTA–UKALTA Symposium hosted by the British Council in London in February 2020 both identified the need for a new handbook to support alignment with the greatly expanded descriptive scheme of the CEFR CV. In undertaking to produce the handbook, the steering group (which emerged from the February 2020 event) recognized that alignment applies not only to language tests but to policy, curriculum guidelines, curricula, syllabuses, textbooks and other teaching/learning resources. The group also decided that the handbook should serve to inform policy makers, teacher educators, teachers and other language education stakeholders, as well as supporting the more or less technical processes on which alignment depends. The handbook also aims to help users to navigate the wide range of CEFR-related reference documents now available.

This handbook has been prepared with **two audiences** in mind, each of which comprises a number of specialized subgroups, as shown below. The bulleted lists are not intended to be exhaustive.

AUDIENCE 1	AUDIENCE 2
Those wishing to undertake **a practical or applied CEFR alignment exercise** in a particular context and for a particular purpose (e.g. to be able to make or evaluate a claim concerning CEFR alignment). This includes: ▪ in-service teachers ▪ materials developers ▪ syllabus designers ▪ pre-service teachers ▪ test producers ▪ textbook writers.	Other stakeholders in education or in society at large who are **primarily concerned with policy matters and decision-making** relating to language education more generally (e.g. to evaluate the claims made by textbook or test publishers regarding CEFR level). This includes: ▪ administrators ▪ curriculum developers ▪ education ministry personnel ▪ employers ▪ managers ▪ policymakers ▪ publishers ▪ teacher trainers/educators.

[1] *Relating Language Examinations to the 'Common European Framework of Reference for Languages: Learning, Teaching, Assessment' (CEFR). A Manual*, Strasbourg: Council of Europe, revised version, 2009.

[2] *Manual for Language Test Development and Examining*, Strasbourg: Council of Europe, 2011.

Chapters 1 and 2 of the handbook have been written to be broadly accessible and directly relevant to both audiences. They provide an essential introduction to the CEFR and the CEFR alignment process, including essential familiarization. Chapters 3 to 6 offer both general and detailed guidance on the successive steps in an alignment process together with information on available tools and approaches to reporting on the activities undertaken.

This handbook was developed as follows. First, the steering group agreed on its scope, drafted Chapters 1 and 2, and sent them to selected language education professionals for critical feedback. Next, specialists with extensive experience of working with the CEFR were invited to draft Chapters 3 to 6. After that, David Little edited the draft handbook, which was circulated to the specialists who had provided feedback on the first drafts of Chapters 1 and 2. Finally, the handbook was revised to take account of their comments and suggestions.

Steering group ▪ Neus Figueras, European Association for Language Testing and Assessment (EALTA) ▪ David Little, Trinity College Dublin ▪ Barry O’Sullivan, British Council ▪ Nick Saville, Association of Language Testers in Europe (ALTE) ▪ Lynda Taylor, UK Association of Language Testing and Assessment (UKALTA)	**Editors** ▪ Neus Figueras, European Association for Language Testing and Assessment (EALTA) ▪ David Little, Trinity College Dublin ▪ Barry O’Sullivan, British Council
Authoring team ▪ Glyn Jones, Freelance consultant, UK ▪ Elif Kantarcioğlu, ID Bilkent University, Turkey ▪ Noriko Nagai, Ibaraki University, Japan ▪ Graham Seed, Cambridge Assessment English, UK ▪ Jamie Dunlea, British Council, UK ▪ Neus Figueras, European Association for Language Testing and Assessment (EALTA) ▪ David Little, Trinity College Dublin. Ireland ▪ Barry O’Sullivan, British Council, UK ▪ Nick Saville, Association of Language Testers in Europe (ALTE) ▪ Lynda Taylor, UK Association of Language Testing and Assessment (UKALTA)	**Reviewers** ▪ Elaine Boyd, Freelance consultant, UK ▪ Bronagh Ćatibušić, Trinity College Dublin, Ireland ▪ Karen Dunn, British Council, UK ▪ Peter Lenz, Institute of Multilingualism, University of Fribourg, Switzerland ▪ Vincent Folny, France Éducation International, France ▪ Jane Lloyd, Cambridge University Press & Assessment, UK ▪ Waldemar Martyniuk, Uniwersytet Jagiellonski w Krakowie, Poland ▪ Josep Ma. Montcada, Generalitat de Catalunya, Spain ▪ Felicity O’Dell, LT123, UK ▪ Edgar Rene Pacheco, Centro de Enseñanza para Extranjeros (CEPE), UNAM, México ▪ Lorenzo Rocca, Società Dante Alighieri, Italy ▪ Cristina Rodriguez, Xunta de Galicia, Spain ▪ Wu Sha, National Education Examination Authority (NEEA), China ▪ Joe Sheils, formerly Head of the Language Policy Division, Council of Europe ▪ Richard Spiby, British Council, UK ▪ Dina Vilcu, Babeș-Bolyai University, Rumania ▪ Kateřina Vodičkova, Charles University Prague, Czekia ▪ Caroline Westbrook, British Council, UK ▪ Russell Whitehead, LT123, UK ▪ Jessica Wu, Language Training and Testing Center (LTTC), Taiwan

CHAPTER 1: THE CEFR

THE COUNCIL OF EUROPE, THE *COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES* (CEFR), AND ALIGNMENT WITH THE CEFR

Chapter 1 introduces the background to the *Common European Framework of Reference for Languages* (CEFR) and explains how it can be used for alignment purposes in second language education. This chapter addresses the following questions:

1.1 What is the Council of Europe?

1.2 What is the *Common European Framework of Reference for Languages* (CEFR)?

1.3 What is CEFR alignment and why is it useful?

1.4 Why is it important to align the different components of a learning system with one another?

1.5 What is involved in CEFR alignment?

1.6 What resources are available to support CEFR alignment?

1.7 What guidance does the handbook provide?

By the end of this chapter, you should have a clear idea of:

- **the Council of Europe and its educational goals**
- **the purpose and content of the CEFR**
- **why CEFR alignment might be relevant and useful in your context**
- **why/how you might wish to undertake a CEFR alignment exercise in your context.**

1.1 What is the Council of Europe?

The Council of Europe was founded in 1949 to promote and defend human rights, democracy and the rule of law; its headquarters are in Strasbourg, France. Today it comprises 46 member states stretching from Iceland in the west to Azerbaijan in the east. (The Council of Europe should not be confused with the European Council, which is the committee of heads of state and government of the European Union. All member states of the European Union are also members of the Council of Europe.)

The Council of Europe's **key instruments** are: *The European Convention on Human Rights* (1950), *The European Cultural Convention* (1954) and *The European Social Charter* (1961, revised version 1996).

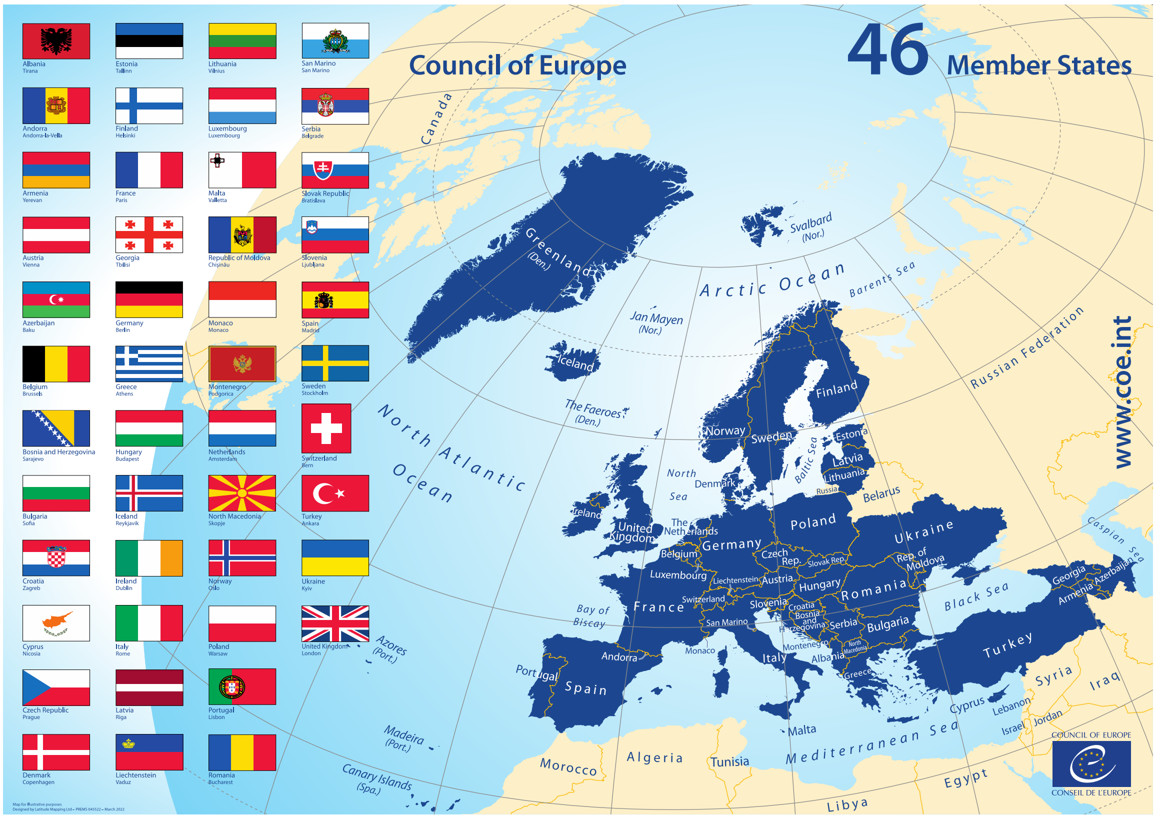

Figure 1.1: Council of Europe member states

Informed by its human rights and legal standards, the **Council of Europe's education programmes** aim to enhance the democratic competences of citizens, promote respect for human dignity and diversity, and combat discrimination.

Since the 1970s the Council of Europe has promoted **the learning and teaching of languages for communication** in order to:

1. preserve and encourage citizens to value Europe's linguistic and cultural diversity
2. facilitate international communication and exchange.

The Council of Europe's concept of **plurilingual and intercultural education** explicitly associates language teaching and learning with the Council's foundational values and political goals, especially social integration, respect for diversity, and intercultural dialogue.

1.2 What is the Common European Framework of Reference for Languages (CEFR)?

First published in 2001, the CEFR was developed to facilitate critical reflection on the various dimensions of language education – curricula, syllabuses, textbooks and examinations – and to promote common understanding across the education systems of Council of Europe member states.

The central component of the 2001 CEFR is a detailed taxonomic description of language proficiency in terms of language use, which it divides into four modes:

- Reception (listening and reading)
- Production (speaking and writing)
- Interaction (spoken and written)
- Mediation (i.e. facilitating communication between individuals or groups who for whatever reason cannot communicate directly).

The CEFR's **non-language-specific descriptive scheme** has two complementary dimensions: (1) the communicative tasks that the language user/learner may need to perform; and (2) the competences on which successful task performance depends. Both dimensions include illustrative scales that use "can do" statements to describe proficiency at six levels: A1 and A2 (basic user), B1 and B2 (independent user), C1 and C2 (proficient user).

In keeping with the Council of Europe's concern for the rights and responsibilities of the individual citizen, the CEFR views **the language user/learner as a social agent**: a member of society who has tasks (not exclusively language-related) to perform in a given set of circumstances, in a specific environment and within a particular field of action.

Published by the Council of Europe in 2020, the definitive version of the **CEFR Companion Volume** updates and extends the illustrative descriptors. It adds new descriptors to the 2001 scales; introduces a handful of new scales and a new pre-A1 level; adds scales for plurilingual and pluricultural competences; offers an extended definition of mediation for which it provides 24 illustrative scales; replaces the 2001 phonology scale by three new scales; and formulates descriptors so that they are gender-neutral and modality-inclusive (and thus refer to signed as well as spoken languages).

Throughout this handbook, general references to the CEFR should be understood as embracing the 2001 and 2020 documents.

1.3 What is CEFR alignment and why is it useful?

The CEFR offers a **comprehensive, coherent and transparent description of language proficiency** in terms of language use. With the help of the CEFR's common reference levels and illustrative scales, this description can be used to:

- develop curricula and syllabuses, textbooks and other teaching/learning materials, tests and other forms of assessment
- demonstrate alignment between two or more of these aspects of language education
- show that an existing curriculum, syllabus, textbook, test or other form of assessment is aligned with relevant parts of the CEFR's descriptive scheme and one or more of its proficiency levels
- compare curricula and syllabuses, textbooks and other teaching/learning materials, tests and other forms of assessment.

Since its publication in 2001, educational authorities, agencies and institutions have recognized certain advantages in claiming **alignment with the CEFR** for different aspects of their educational system. Perceived advantages include:

- achieving systemic coherence and transparency
- establishing a basis for principled comparison
- monitoring for purposes of quality assurance.

CEFR alignment promotes shared understanding and coherence among aspects of an education system, facilitates comparison between them, and supports quality improvement. For these reasons, it is likely to be of interest to:

- agencies charged with the development of national/regional curricula and syllabuses
- agencies charged with the development and administration of national/regional language exams and other forms of assessment
- autonomous educational institutions, especially universities and private language schools
- educators who have to develop their own teaching materials and/or assessments
- language teacher educators
- language teachers working in a context that is partly/wholly shaped by reference to the CEFR and CEFR CV
- national and international language testing agencies
- national and regional ministries of education
- professional associations concerned with quality assurance and quality improvement in language education
- textbook authors and publishers.

For those who are unfamiliar with the work of the Council of Europe and/or with the CEFR, it is important to point out that alignment with CEFR levels is **not** considered to be mandatory or obligatory. The CEFR should always be regarded as a resource for consultation rather than a package for implementation.

1.4 Why is it important to align the different components of a learning system with one another?

O'Sullivan (2020) argues that the success of any learning system depends on the close alignment of elements that have traditionally been regarded as independent of one another: curriculum, teaching/learning materials, teaching approaches, teacher training, and assessment. Although these elements impact significantly on one another and on learning, curriculum developers, materials developers, teacher trainers and assessment specialists mostly work independently of one another. The need to consider these elements from a single unified perspective forms the basis of O'Sullivan's (2020) concept of the Comprehensive Learning System (CLS).

Figure 1.2 highlights the three core elements of the CLS:

1. **Curriculum** – informal as well as formal
2. **Delivery** – includes teacher selection, teacher training, accreditation, professional development and leadership; teaching and learning materials; the physical environment in which the delivery takes place
3. **Assessment** – includes developmental assessment (diagnostic, aspects of progress, formative, etc.) and judgmental assessment (placement, aspects of progress, achievement, proficiency, etc.)

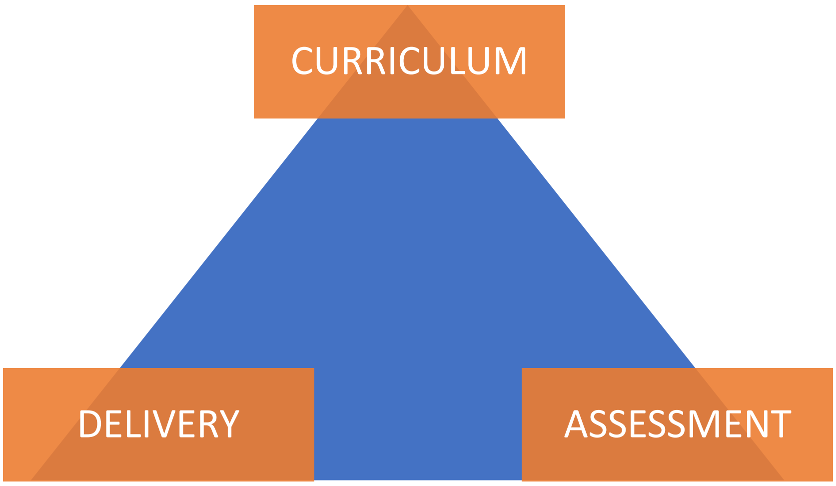

Figure 1.2: O'Sullivan's Comprehensive Learning System (CLS)

If one of these elements is in any way disconnected from the others, then the sytem is under threat. Here are some examples of how one or more elements may be disconnected:

Example 1: A new or revised curriculum is introduced but adequate training for teachers is not provided. It is thus very likely that teachers will not deliver the curriculum in the expected way, in which case the benefits intended by the change in curriculum will not be realized.

Example 2: The assessment does not fit with the requirements of the curriculum (e.g., a curriculum is focused on all four skills but the assessment is a multiple-choice grammar and reading test; or the curriculum is based on the needs of a particular educational context, whereas the assessment is not).

Example 3: The textbooks and materials used in the classroom do not reflect the needs of the curriculum (e.g., the curriculum calls for a communicative approach, but the materials follow a grammar-based approach; or the curriculum is designed for a particular educational context, whereas the materials are not).

Example 4: The curriculum, teacher training and assessment are all well aligned. However, the physical structure of the classroom makes effective delivery difficult or impossible (e.g., rigid furniture obstructs pair or group activities).

So, if a major change is to be introduced into a learning system, it should be clear from the above that all elements need to be considered. For example, if **a new or revised curriculum is introduced, teachers will be provided with adequate training so that they will deliver the curriculum in the expected way** while the tests will also be updated or re-written to meet the revised educational demands.

Before setting out to establish a link between a curriculum, an element of the delivery system or a test, it is important to remember that while it may be useful (and appropriate) to argue that what you are focused on may be linked to the CEFR in ways that support your claim, this does not mean that it is appropriate for use within a particular learning context. For example, a coursebook at B1 developed for use in a regional Spanish school system will not necessarily work for an equivalent system in Malaysia or India. It is not enough to argue that since both are focused on CEFR level B1, the coursebook will be appropriate in both contexts.

1.5 What is involved in CEFR alignment?

Undertaking a CEFR alignment exercise involves one of two processes:

PROCESS 1		PROCESS 2
Collecting evidence and developing an argument to show that an *existing* resource, e.g. a curriculum, a textbook or a language test, fulfils criteria derived from the CEFR.	**OR**	**Developing and documenting a *new* resource, e.g. a curriculum, a textbook or a language test, on the basis of the CEFR criteria.**

Process 1 and Process 2 both involve **a series of well-established and largely sequential steps**, or sets of procedures, as shown in Figure 1.3.

The first step in the alignment process is **FAMILIARIZATION** (explained more fully in Chapter 2). This is an essential stage at the outset of any alignment exercise. Experience from previous alignment studies has shown that in a reduced format it can also be a very useful, and sometimes necessary, preliminary activity for the other stages, e.g. standardization. For example, it can provide a useful opportunity for refreshing or for refocusing on specific elements relevant to the context for the alignment project.

FAMILIARIZATION
Ensuring that all participants in the alignment process have a sufficient knowledge of the CEFR, its levels and descriptors

STANDARD SETTING
Determining valid cut scores or decision judgments for assessment purposes

SPECIFICATION
Describing/profiling the content of a language syllabus/textbook/test in relation to the categories of the CEFR

VALIDATION
Collecting and presenting appropriate evidence in support of alignment claims

STANDARDIZATION
Ensuring, through training, a common understanding of the CEFR levels and the accurate benchmarking of local performance samples to relevant CEFR levels

Figure 1.3: Steps in the alignment process


The next steps are **SPECIFICATION** and **STANDARDIZATION** (Chapters 3 and 4) which are required for CEFR alignment projects in most, if not all, contexts of use.

For those involved in all aspects of establishing an empirically based link between a curriculum, a set of materials (e.g. textbook or online course), or an assessment or test, **STANDARD SETTING** procedures (Chapter 5) are a key requirement. For some of these contexts, the degree of standard setting activity required is likely to be less than we might expect for a test. The specific use of a test determines the choice and appropriateness of standard setting procedures.

VALIDATION (Chapter 6) is best understood as the continuous process of quality monitoring in order to gather the validity evidence to support any claims of CEFR alignment. Like Familiarization, Validation is to some degree relevant to all the other steps in the alignment process – by demonstrating that all stages have been followed in an appropriate way, we establish evidence of the validity of subsequent claims of a link to the CEFR.

1.6 What resources are available to support CEFR alignment?

The CEFR 2001 and the Companion Volume (CV) 2020 constitute a *single* resource to refer to when undertaking the CEFR alignment process. Here, for easy reference, is a brief outline of what each document contains:

CEFR 2001

- The Council of Europe's language education policy (Chapter 1)
- The rationale for the action-oriented approach (Chapter 2)
- The Common Reference Levels (Chapter 3)
- Communicative language activities but ignore the illustrative scales (Chapter 4)
- The user/learner's competences but ignore the illustrative scales (Chapter 5)
- Language learning and teaching (Chapter 6)
- Tasks and their role in language learning (Chapter 7)
- Linguistic diversification and the curriculum (Chapter 8)
- Assessment (Chapter 9)

+

CV 2020

- Key aspects of the CEFR for teaching and learning
- Updated CEFR illustrative descriptive scales including:
  - additions to 2001 scales
  - new scales for reception, production and interaction
  - new scales for phonology
  - scales for mediation
  - scales for signing competences
  - scales for plurilingual/pluricultural competences
- Communicative language activities and strategies (replaces CEFR 2001, section 4.4 in Chapter 4, adding substantial new material)
- Communicative language competences (replaces CEFR 2001, section 5.2 in Chapter 5)

The searchable database of CEFR CV descriptors may be useful throughout the alignment process. For those involved in creating or producing language tests, examinations and other forms of assessment aligned to the CEFR, other useful resources include the *Manual for Relating Language Examinations to the CEFR* (Council of Europe, 2009) and the *Manual for Language Test Development and Examining* (Council of Europe, 2011). There are also publications which describe case studies exemplifying and reflecting on the test linking process and experience, e.g. Figueras and Noijons (eds), 2008; Martyniuk (ed), 2010.

1.7 What guidance does this handbook provide?

The remainder of this handbook provides **practical and accessible guidance** to assist with the process of CEFR alignment in a variety of specific contexts. Each of the following chapters focuses on one of the five procedures outlined above:

- Chapter 2 – Familiarization
- Chapter 3 – Specification
- Chapter 4 – Standardization
- Chapter 5 – Standard Setting
- Chapter 6 – Validation

Each chapter begins with **an explanation of the procedure and its importance**, followed by **general advice and practical activities that are relevant to all users** regardless of language education context (e.g. curriculum, teaching, assessment).

The remainder of each chapter offers **more targeted guidance and practical activities specific to different contexts of use and their associated subgroups**, e.g. language teachers, curriculum designers, textbook writers, test producers. You can select from and focus on these according to need.

We understand that you may be undertaking your alignment exercise using a group approach that involves one or more coordinator(s) and participants. For this reason, we provide some **practical advice and suggestions relating to these differing roles**, including tasks to be completed with likely timings.

In the Appendix, you will find **photocopiable summary forms** to use and complete. The summary forms will also be available online in editable format. This additional practical tool will assist you with your ongoing monitoring and validation throughout the alignment process. While we suggest that you use the forms as they are presented here, there will be situations where you feel they need to be adapted to fit the needs of a particular alignment approach or resource. Any changes to the approach proposed in this handbook should be carefully considered and rationalized in your final report. For some additional information, please see the **Notes for Your Own Implementation** section in Chapter 6.

Notes for your own implementation …

Remember that YOU will be the best judge of what is (and is not) possible or realistic within your specific context of operation.

For this reason, we strongly encourage you to **tailor what you do to your specific context**, taking account of your resources and limitations. **Avoid being tempted to do too much** – or more than is actually required for your purposes. Equally, **avoid being tempted to underestimate what you need to do!**

Key questions for you to keep in mind at all times during an alignment process are:

1. **Why are we doing this?**
2. **Which steps are essential for us in our context and for our purposes?**
3. **Which steps do we prioritize?**
4. **Which steps may not be necessary?**
5. **How long will the process take?**
6. **How much will it cost?**
7. **What expertise will we need, or have access to?**
8. **What resources (i.e. time, funding, expertise) do we need to plan for?**

CHAPTER 2: FAMILIARIZATION

Chapter 2 focuses on familiarization as an important first step in any alignment project. The chapter is organized as follows:

- **2.1 What is familiarization?**
- **2.2. Generic familiarization activities**
- **2.3. Specific familiarization activities**
  - **2.3.1. Coordinator roles and decisions**
  - **2.3.2. Preparing materials**
- **2.4. Guidelines for reporting**

Sections 2.1, 2.2. and 2.4 are addressed to all those involved in an alignment project, whatever their role or responsibility, whereas Section 2.3. is addressed to those responsible for the organization and delivery of specific familiarization activities.

By the end of this chapter, you should have a clear idea of:

- **the importance of familiarization and what it entails**
- **the activities you need to complete to achieve a good understanding of the CEFR approach and descriptor scales**
- **how to document the completion of the activities for reporting purposes.**

2.1 What is familiarization?

Familiarization is designed to ensure that those involved in an alignment project have an appropriate knowledge of the CEFR and share a common understanding of the purposes of the project. This handbook envisages two types of familiarization: generic and specific (see Figure 2.1.) Generic familiarization can take place in a single session and can be completed individually, whereas specific familiarization normally takes place in a group, in more than one session, and with the guidance of a coordinator.

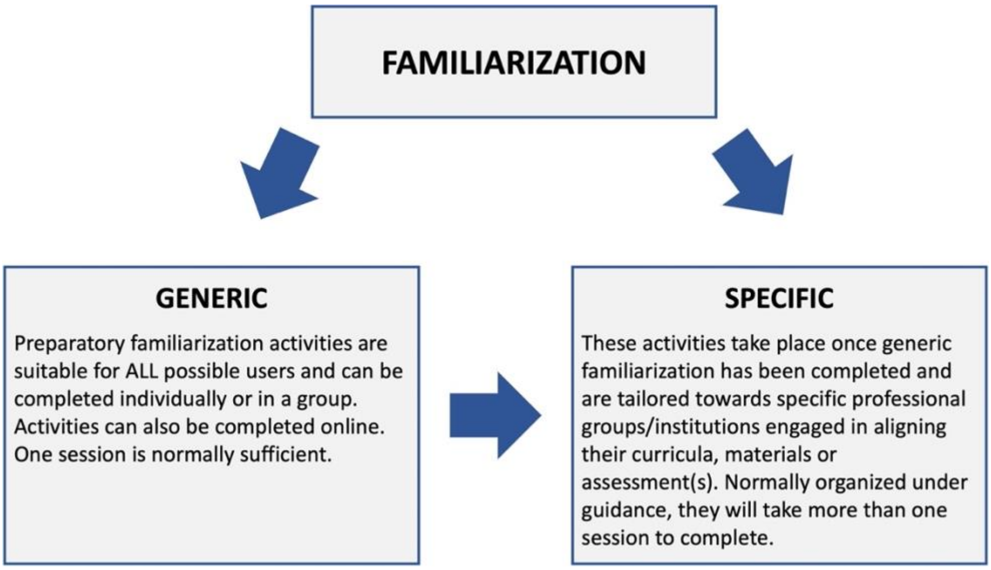

Figure 2.1: Generic and specific familiarization

The time needed for the activities in each familiarization session depends on participants' existing level of familiarity with the CEFR (2001) and the CEFR CV (2020) and also on the aim and scope of each specific alignment project, but we recommend that you plan a full morning or afternoon for each session.

Some individual users may want to follow the links to additional documentation provided in the CEFR CV for reasons of personal interest or because of specific needs in their context. Curriculum designers, for example, may want to access Beacco and Byram's *From Linguistic Diversity to Plurilingual Education: Guide for the Development of Language Education Policies in Europe* (2007). Language testers may want to access North et al.'s *Relating Examinations to the CEFR. A Manual* (2009) or ALTE's *Manual for Language Test Development and Examining* (2011).

2.2 Generic familiarization activities

Using the electronic or printed version of the CEFR CV 2020, complete the five activities shown in Figure 2.2 in the order proposed, working individually or in a group, face-to-face or online. Note down your answers to each question as they will help you to complete summary form 2.1 in Activity 5.

Figure 2.2: Five generic familiarization activities

1. **Key Aspects:** Read Chapters 1 and 2 of the CEFR CV 2020 to learn about the additions and changes made to the CEFR 2001. Ask yourself:
   - What does alignment to the CEFR mean?
   - What are the implications for me and my context?

2. **Descriptive Scheme:** Focus on Section 2.4 of the CEFR CV, *The CEFR descriptive scheme*, and ask yourself:
   - Which activities, competences and strategies are most relevant to my context?
   - Are they all equally relevant? If not, how can I prioritize them?

3. **Salient Features:** Read Appendix 1 of the CEFR CV, *Salient features of the CEFR levels*, which outlines the progression from level to level. Ask yourself:
   - What level(s) will I mainly be working with?
   - What are the main features of that level and the levels on either side of it?

   Go to Appendix 2, *Self-assessment grid*, and identify the main aspects of each mode of communication (reception, production, interaction, mediation). Ask yourself:
   - Which aspects are most relevant in my context?

4. **Relevant Scales:** Look at the diagrams that show the different illustrative descriptor scales and identify the most relevant scales for your context:
   - Communicative language activities and strategies:
     - Reception (Figure 11, page 47)
     - Production (Figure 12, page 61)
     - Interaction (Figure 13, page 71)
     - Mediation (Figure 14, page 90)
   - Plurilingual and pluricultural competence (Figure 15, page 123)
   - Communicative language competences (Figure 16, page 129)
   - Signing competences (Figure 17, page 144).

   You may find the introductory text that precedes each diagram helpful. Note that you may need to read some of the scales in detail in order to decide on their relevance for your purposes. Write the scale labels down for later reference. Given the number of scales available, make sure you select *only* those most relevant to your work. Ask yourself:
   - Which are the most relevant scales in my context? Why?
   - What is the rationale for my decisions? (Your answer to this question is key because it will support the claims you make.)

5. **Summary:** Complete Summary Form 2.1 in the Appendix.

2.3 Specific familiarization activities

Specific familiarization activities are tailored to the particular needs of groups of professionals who have already completed the generic familiarization activities. Wherever possible, specific familiarization should be a group activity because discussion helps to build a shared understanding of how to interpret the different descriptors especially where these are not fully transparent. For example, in the *Analysis and criticism of creative texts (including literature)* descriptor for B1, we find one can-do statement which reads: "Can describe the key themes and characters in short narratives involving familiar situations that contain only high frequency everyday language". Group discussion around specific words or phrases like 'short', 'familiar' or 'high frequency everyday language' can help participants come to an agreed (and, we hope, accurate) understanding of how to interpret this statement at the B1 level.

Participants should be ready to share their completed Summary Form 2.1 at the request of the coordinator.

For a project involving all modes of communication (reception, production, interaction, mediation), a minimum of two sessions will be needed to complete the activities. Precise timings are likely to vary depending on how familiar the participants are with the CEFR and which scales the generic familiarization activities have identified as relevant.

Subsections 2.3.1 and 2.3.2 are addressed to the professional who will take on the role of coordinator. Before preparing the activities, it may be useful to read Chapter 3 of the *Manual for Relating Examinations to the CEFR*, especially if the participants are to align a high-stakes test to the CEFR.

Coordinators may also find it useful to consult the searchable database of all CEFR CV descriptors.

2.3.1 Coordinator roles and decisions

Coordinators play a key role in the organization of specific familiarization activities. They should have extensive knowledge of the CEFR and the implications of an alignment process, and should be familiar with the context and the requirements of the organizing group/institution. Only on the basis of such knowledge can they:

- decide on the logistics and content of the session(s)
- prepare the materials for the successive activities
- write a report on the familiarization process.

The time that coordinators need to invest in the preparation, delivery and documentation of each specific familiarization session will depend on a number of variables: their previous experience of running CEFR 2001 familiarization sessions; the responsibilities they are given (type of report required, amount of information to be compiled, number and type of sessions, etc.); the amount of administrative help available (to prepare materials, organize premises and catering, manage the online platform, etc.). The number and background of the participants will also have an impact on the amount of time coordinators will need to devote to planning. As a general rule, preparing specific familiarization will take at least twice as long as the participant sessions.

Before engaging in the preparation of specific familiarization activities, you should:

1. Decide whether activities will be carried out in person, online or a combination of the two. Make sure that the number of participants accepted allows for adequate discussion and the successful completion of activities. Limit attendance in face-to-face sessions to a maximum of 18–20 participants. For online sessions, check the resources available to facilitate discussion and the completion of the activities (software, IT support, etc.).

2. Read and assess the responses of participants to the generic familiarization questions in Summary Form 2.1, which may have been completed individually, in order to decide:
   a) which communication modes/skills, scales, levels and descriptors to focus on (including rationale)
   b) how many sessions will be needed
   c) how much revision will be needed before you start work on communication modes – reception, interaction, production, mediation (this will depend partly on how recently generic familiarization took place)
   d) the focus, structure and content(s) of each session
   e) the number, type and duration of activities, discussion time(s), break(s), etc.
   f) which examples to use of materials relevant to the context in question (an exam task, a classroom assessment activity, a section of a curriculum, a lesson plan, etc.) and how to provide them to the participants.

   If the project covers all four modes of communication, we recommend that you focus on reception in Session 1, interaction and production in Session 2, and mediation – with which participants may be less familiar – in Session 3. Having one whole session for mediation will allow you to work on mediation strategies as well as mediation activities, and perhaps include familiarization activities with the plurilingual and pluricultural competence scales.

   You will need to decide when to address the scales for communicative language competences (we suggest Session 2), and whether or not participants need familiarization with the signing competence scales.

3. On the basis of the stakes attached to the project and the brief you have received, decide what type of report is needed and what evidence(s) you should collect.

2.3.2 Preparing materials

The preparation of materials and the management of the sessions should be adapted to the type of delivery envisaged (in person or online) and its timing (synchronous or asynchronous).

Specific familiarization – Session 1

Session 1 is divided into two parts. Part 1 is a review of the activities completed in generic familiarization and should be carried out as presented. Part 2 may address reception, production and interaction, or mediation, though we recommend that you start with reception. The activities in both Part 1 and Part 2 are graphically presented in Figure 2.3.

In each activity, and prior to engaging in discussion, encourage participants to ask themselves some questions. Their answers will be useful when they complete Summary Form 2.2 at the end of the session.

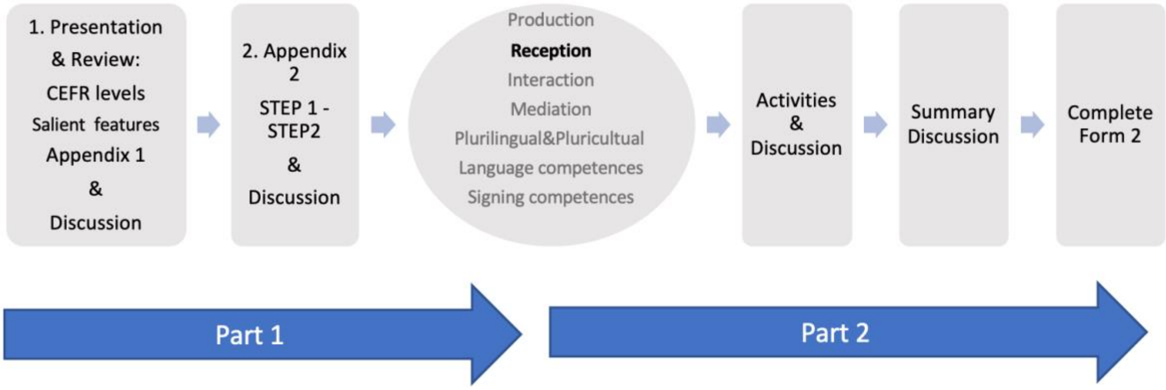

Figure 2.3: Activities to be carried out during the first familiarization session

1. Presentation and review

Prepare materials (PowerPoint presentation, quiz, etc.) to introduce the objectives of specific familiarization and the structure and content of the session(s).

Following the presentation, refer participants to Appendix 1 of the CEFR CV (if necessary, provide photocopies) and ask them to read and individually to underline/highlight the salient features of each CEFR level. Ask them to check their completed Summary Form 1 and to note their answer(s) to the following question:

- What features in the CEFR levels are most relevant in your context?

Participants discuss their answers working in pairs or small groups.

2. Appendix 2 of the CEFR CV: Self-assessment grid

Prepare the self-assessment grid for a two-step activity.

Step 1. As the focus in this first step will be on reception, interaction and production, photocopy pages 177–179 of the CEFR CV and

a) blank some of the cells in the grid (those considered most relevant to the project)
b) prepare cut-up versions of the descriptors removed from the blanked cells, which participants will use to complete them
c) photocopy the completed grid for reception, interaction and production for correction and discussion.

Ask participants to write their individual responses to the following questions:

- What key words (consider actions, type of information, conditions and limitations) best identify:
  - oral comprehension?
  - reading comprehension?
  - oral production and oral interaction?
  - written production and written/online interaction?

Ask participants to work in pairs or small groups to fill the blanked cells with the cut-up versions of the descriptors prepared in Step 1 and to identify the main features of reception, interaction and production across the levels.

Step 2. As the focus in this step is on mediation, print/photocopy pages 180–181 of the CEFR CV and

a) cut up the descriptors for the three types of mediation – mediating a text, mediating concepts, mediating communication (all levels)
b) put them in an envelope, jumbled
c) prepare three grids, one for each type of mediation, so that participants can select which scale and level each of the cut-up descriptors belongs to
d) when the participants have completed the task, photocopy their completed grids for mediation for correction and discussion.

Ask participants to work in pairs or in small groups to assign the descriptors to type of mediation and level. Distribute the complete grid for correction and discussion. Ask participants to work individually to identify the main features of reception, interaction and production across the levels. Ask them to write their answers to the following questions before engaging in discussion:

- What key words (consider action(s), type of information, conditions and limitations) best identify the following:
  - mediating a text?
  - mediating concepts?
  - mediating communication?

3. Activities with descriptor scales and discussion

Prepare activities appropriate to the context and level in question, decided on the basis of Summary Form 2.1. Only one mode of communication can be addressed in the remaining time in Session 1. If the project focuses on all modes of communication, we advise you to focus on reception in this first session.

Activities may include qualitative analysis of relevant level descriptors, reordering jumbled cut-up selected descriptors, ranking the descriptors selected and placing them on a grid, assigning different descriptors to the appropriate scale and level, etc. (for further details on the preparation of these activities see Chapter 3 of the *Manual for Relating Examinations to the CEFR*).

When participants have completed each activity and before they engage in group discussion, ask them to respond to the following questions:

- Why is this scale relevant in my context?
- What key words best identify my purpose(s)?

At this point each participant should complete Summary Form 2.2.

4. Prepare a summary session

Prepare **a summary session** so that participants have an overall view of the work completed and what has been achieved.

If you plan to include agreement indices[3] in the final report, make sure activities are presented in such a way that you tally and analyse participants' ratings/rankings.

5. Collect Summary Form

Collect Summary Form 2.2 for Session 1.

6. Write a report summarizing responses

Write a report summarizing the responses to Summary Form 2.2.

[3] These are indications of the degree of agreement among participants and can be quantitative (described in the Reference Supplement to the Manual referred to above) or qualitative (comparison of comments/observations). The latter will be most relevant to alignment of curricula or materials.

Materials for further sessions

Based on the decisions taken in Session 1 regarding the modes of communication to focus on, and depending on participants' existing familiarity with the CEFR, you will need to prepare one or two further sessions, following Steps 3, 4, 5 and 6 described above.

Like Session 1, further sessions will include (see Figure 2.4):

- an introductory activity that recapitulates the work done in previous sessions, in order to feed back and feed forward
- activities and discussion on the scales selected, normally separated by a break
- a concluding session and a summary discussion, guided by the coordinator
- the completion of the corresponding summary form by the participants.

Figure 2.4: Activities to be carried out in further specific familiarization sessions

2.4 Guidelines for reporting on the familiarization stage

To bring together all of the evidence gathered during the familiarization stage, the coordinator should summarise the activities and responses of the participants using Form 2.3. This will then create the basis of later reports – both of this session and also the final validation report, where the final claims are made and supported.

Reporting on each alignment step has two purposes. First, it is necessary to maintain an internal record, a kind of institutional memory; secondly, the evidence collected in the report will form part of the final alignment report (Chapter 6). Reports can be written for internal use only or with an external audience in mind.

The main author(s) of the report will be the appointed coordinator(s), who will be responsible for confirming that all activities are completed and documenting the outcomes. The coordinator(s) will submit the report to the commissioning institution, which will be responsible for its dissemination.

2.4.1 Reporting for internal use

For internal reporting, administrative procedures specific to the institution should be documented for possible future reference:

- protocols followed and approvals needed (choice of coordinator(s), selection of participants, etc.)
- resources and preparation necessary (budgets, technology, materials, clerical help, etc.)
- logistics (premises, catering, etc.)
- lessons learnt and recommendations.

The report should also include copies of the completed familiarization forms and a brief summary document, signed by the coordinator(s) and including the following:

- the number of participants
- the activities/contents covered in the sections, as well as the timings and calendar
- the degree of satisfaction of the participants
- an overall evaluation by the coordinator(s).

2.4.2 Reporting for external use

The nature of a report for external use will vary depending on the specific project and the target audience, which may be broader (e.g. schools, publishers, test users, test-takers, etc.) or more specialized (e.g. teaching or assessment professionals).

When a report is aimed at a broad audience, the information provided in the report for internal use (excluding non-relevant institution-specific information) should suffice.

In the case of a report for a specialized audience, additional detail will be required, such as information on participants' biodata, their performance(s) in the different activities (e.g. agreement indices when ranking descriptors or accuracy ratings when assigning descriptors to scales or levels).

Notes for your own implementation …

If you are working on your own:

- Completing familiarization activities is hard work. Regular breaks contribute to success.
- Avoid being too ambitious, aiming at too much. Careful is better than fast.
- Make sure there is sufficient time to complete each activity.
- Document decisions as you go along. That will make completing summary forms easier.

If you are a coordinator, in addition to the above:

- Allow for sufficient discussion time.
- Do not rush participants.
- Vary the format of activities within and across sessions.
- Make sure you collect all relevant information you will need for the final report, including the Summary Forms at the end of each session.

CHAPTER 3: SPECIFICATION

Specification analyses the content of any resource, existing or new, in terms of approach and coverage in relation to the categories presented in the CEFR. As well as serving to report alignment to the CEFR, the process of specification has an awareness-raising function that may help to further improve the resource. Specification seeks to answer the question: "How far does my resource cover the content(s) of the CEFR?"

The level of detail to which specification is carried out depends on the alignment claims you wish to make.

Engaging in specification procedures requires familiarization with the CEFR (as described in Chapter 2) and a thorough understanding of the target proficiency level, the context and the nature of the resource (e.g. a syllabus or textbook) that you wish to align to the CEFR. The person(s) responsible for carrying out specification will be able to select the most relevant CEFR descriptor scales and descriptors and relate them to the aims and content of the resource in order to claim alignment to the CEFR.

This chapter is organized as follows.

3.1. What is specification?

3.2. Generic specification activities

3.3. Specific specification activities

3.4. Guidelines for reporting

By the end of this chapter you should have a clear idea of:

- **what specification entails**
- **what approach is most relevant/useful in your context**
- **how to analyse and describe your resource**
- **how to profile your resource in relation to the CEFR**
- **how to document the degree to which your resource is aligned to the CEFR.**

3.1 What is specification?

Specification is the process of analysing your resource and mapping it to the CEFR with a level of detail appropriate to your needs, so for example, with a national curriculum the level of detail would be expected to be high but for an institutional-level equivalent a broader high-level mapping might suffice. It is the first step in the collection and presentation of evidence to support your alignment claim.

In order to undertake specification, you need to have completed at least the generic familiarization activities in Chapter 2, which we recommend that you revisit and explicitly relate to the resource you plan to align to the CEFR. Specification will require that you engage in the careful, detailed reading of CEFR descriptors.

This handbook proposes two types of specification procedure: **generic and specific**. The generic procedure involves a broad, overall linking of the structure and content of your resource – curriculum, materials or assessments – to the content and level descriptors of the CEFR. A series of generic specification activities are provided in Section 3.2. The specific procedure applies the same principles but at a much more detailed and fine-grained level, as described in Section 3.3.

Each procedure moves in one of two directions, top-down or bottom-up. The top-down procedure starts from the CEFR as the basis for creating a totally new, CEFR-informed curriculum, set of materials or assessment, whereas the bottom-up procedure starts from the analysis of existing resources and practices which are to be aligned to the CEFR. In both cases, the purpose behind the resource needs to be made explicit, clearly rooted in its context and based on learners' needs. Also in both cases, an in-depth understanding of and familiarity with the CEFR and the CEFR CV is presumed (see Section 2.1 above).

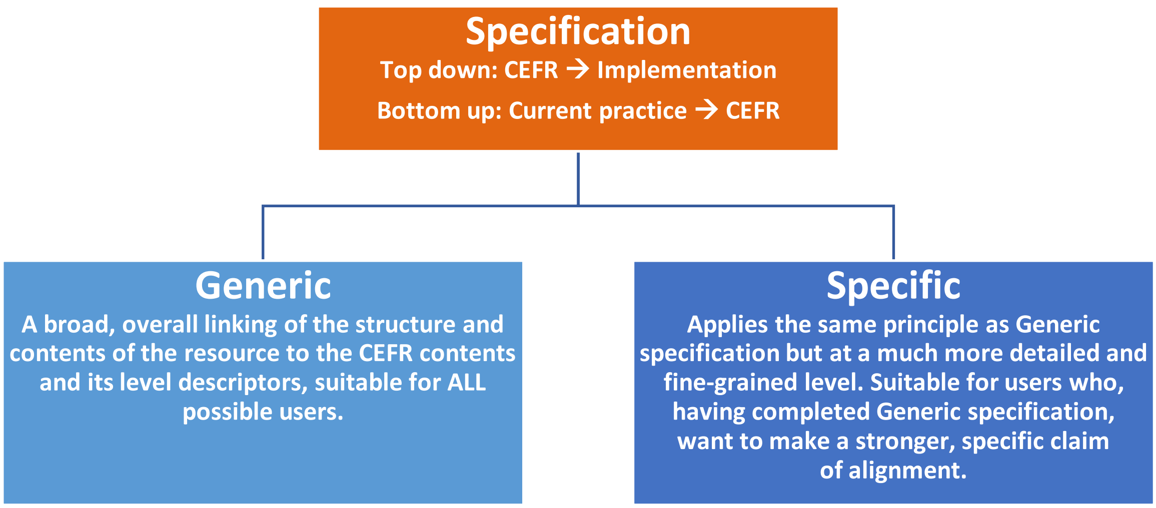

Figure 3.1: The specification process

In the generic procedure, the salient features of each proficiency level (Appendix 1 in the CEFR CV) and the self-assessment grid (Appendix 2 in the CEFR CV) are used for the contextualization and alignment by all users, just as they were in the generic familiarization discussed in Section 2.1. In the specific procedure, on the other hand, the illustrative descriptor scales for the particular communicative language activities, strategies, and competences that are relevant to one's context and purpose should be identified and selected. Detailed steps for each process are proposed in the following sections.

3.2 Generic specification activities

With the activities presented in this section you can begin to establish a general claim that your resource is aligned to the CEFR. When you have completed them, you will have a general description of your resource in terms of the CEFR's categories and proficiency levels.

The five generic specification activities (Figure 3.2) can be carried out by one person or by a group working collaboratively. As you proceed, make a note of the answers to the questions attached to each activity as they will help you complete Summary Forms 3.1 to 3.3 in activity 5.

Figure 3.2: Steps in generic specification procedure

Activity 1: Relevant proficiency level and salient features

Go to Appendix 1 in the CEFR CV (p. 173), which you already consulted to complete activity 3 in the familiarization process in Chapter 2.

a) Take a close look at the table summarizing the Common Reference Levels (p. 175). Read the descriptors for each of the six proficiency levels. Underline the verbs related to the different communicative activities: "understand" and "summarize" refer to receptive ability (oral and written); "express" and "produce" refer to productive ability (spoken and written); "interact" and "exchange" refer to interactive ability.

Ask yourself:

1) What proficiency level(s) is/are most relevant/pertinent for the context and purpose of my resource?
2) Do different activities in my resource cover, or aim to cover, more than one level?

b) Read carefully the text on salient features for each proficiency level and identify words and phrases which define each level and differentiate one level from another.

Ask yourself:

1) Has reading the salient features text confirmed my responses to the questions in 1.a above?
2) Can I explain critical differences between the level I identified as most relevant for my resource and its adjacent levels?

Activity 2: Language activities and strategies

The CEFR organizes its descriptive scheme according to four modes of communication. These are shown in Figure 3.3 together with the language activities and strategies associated with them. As there are no overall scales for strategies, the specific strategy scales pertaining to each mode of communication have been included.

Summary Form 3.4 should be completed by the developer of the resource prior to the data gathering stage. This presents an overview of the intended outcomes of the stage and can later be compared with the actual findings in order to validate the claims made following the activities.

Ask yourself:

1) Which mode(s) of communication, language activities and language strategies are most relevant to my resource?

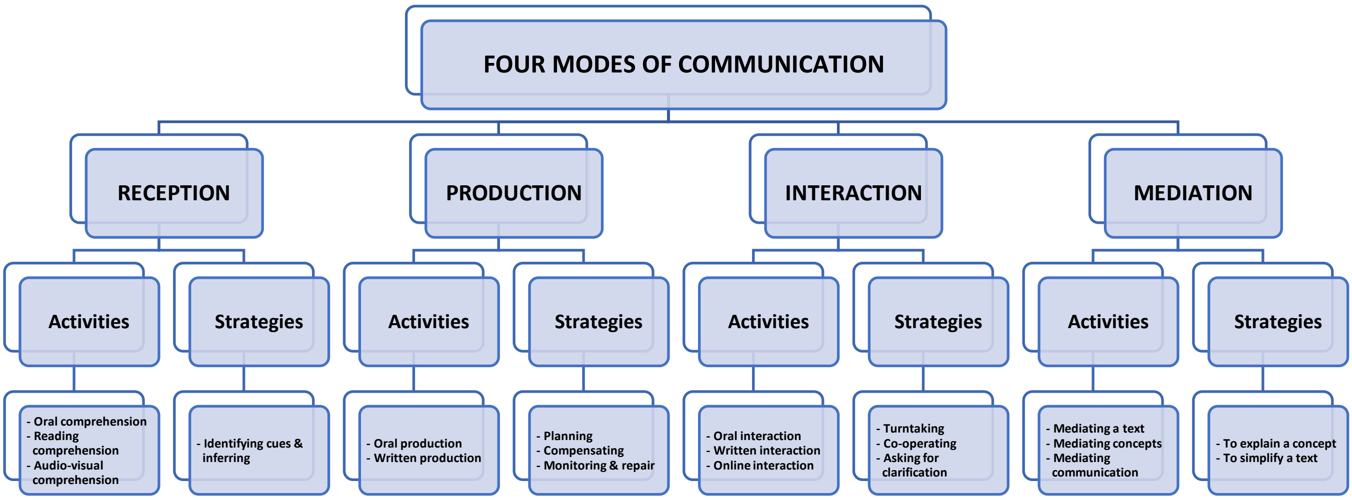

Figure 3.3: Four modes of communication

Activity 3: Competences

Look at the following figures in the CEFR CV:

- Figure 15, *Plurilingual and pluricultural competence* (p. 123)
- Figure 16, *Communicative language competences* (p. 129)
- Figure 17, *Signing competences* (p. 144).

Ask yourself for each figure:

1) Which competence(s) if any does my resource target?
(In most cases it is unlikely that it will target competences across all three figures.)

Activity 4: Graphic profiling of alignment

Go to the self-assessment grid in Appendix 2 of the CEFR CV (p. 177).

Ask yourself:

1) Which descriptors are most relevant to my resource in terms of proficiency level and communicative language activities?
2) How can I make a graphical profile which depicts the relationship between my resource and the CEFR level descriptors in the self-assessment grid?

Activity 5: Summary

On the basis of your responses to the questions in activities 1–4, complete Summary Forms 3.1, 3.2 and 3.3 in the Appendix.

3.3 Specific specification activities

A specific specification relates the content of your resource, section by section, to particular CEFR descriptors. As a result of carrying out the procedure, you should be able to show which of the CEFR's illustrative scales relate to your resource, and for each one:

- which level is targeted (which particular descriptor(s) applies)
- where it is addressed in the resource.

The procedure starts from an analysis of the resource. How you arrive at your analysis will depend on two factors:

- whether you are designing a new resource (top-down) or aligning an existing resource (bottom-up)
- the nature of the resource: whether it is a curriculum, a course plan, a coursebook, a set of materials, or an assessment.

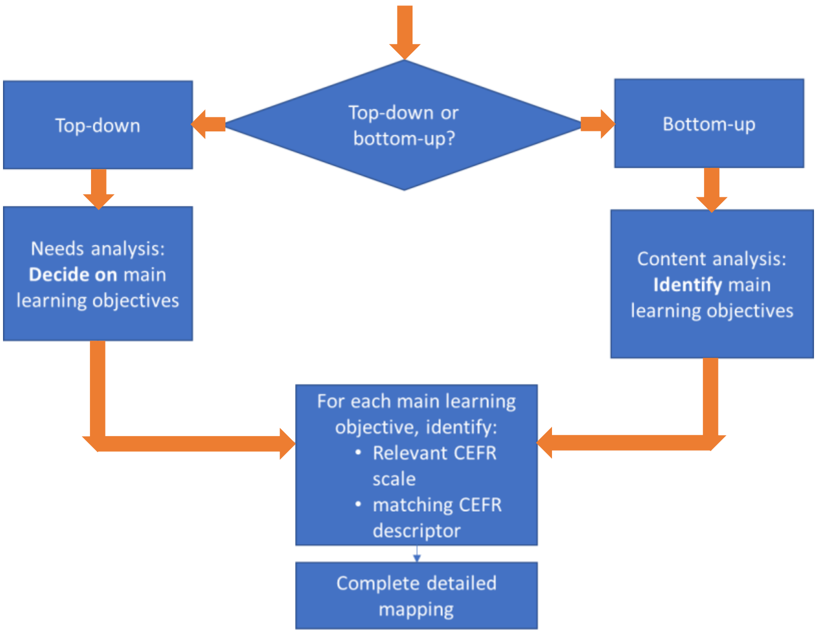

Figure 3.4: Overview of the specific specification procedure

Activity 1: Analyse your resource

If you are designing a new resource (top-down), your starting point should be a needs analysis. What should a successful learner (or test-taker) be able to do having completed the programme (or having passed the assessment)? You will already have a general answer to this question if you have completed a generic specification (Section 3.2). At this stage you go into greater detail, aiming to compile a list of discrete learning objectives.

If you are aligning an existing resource, then your starting point will be a content analysis. This may have been done in the process of developing the resource. There may be an overview, a table of contents or a test specification that shows what the separate learning objectives (or test constructs) are. This may make analysis easier, but in any case, you should look beyond the stated objectives to find, as far as possible, what learners will actually be expected to do. In a curriculum or course outline, look for descriptions or examples of activities; in learning materials look at the learning activities; in an assessment programme or an exam, look at the activities planned and/or at sample test items and tasks. Do they match the stated objectives?

Whether you are designing a new resource or aligning an existing one, you have to decide what level of granularity your analysis should have. For example, a coursebook may consist of a number of units, each comprising a number of lessons, and each lesson may contain a sequence of different learning activities. Should you attempt to relate each unit to a CEFR descriptor, or each lesson, or (to go the finest level of detail) each separate activity? There is no simple answer to this question, but a general principle is that you should look for clearly definable learning objectives that are (or can be) stated in terms of what a successful learner *can do*, because you want to relate these to the "can do" descriptors in the CEFR. To pursue the example of a coursebook, a particular lesson may contain a ten-minute class activity designed to teach ten vocabulary items. This is not a learning objective that can easily be related to a CEFR descriptor, but it may function as *scaffolding* for the main activity in the lesson, which consists in reading an authentic magazine article. This is an objective that can be linked to a CEFR scale (such as "Reading for information and argument"). In this case, then, it is at the level of the lesson as a whole that you can best relate the material to the CEFR.

The other issue is the adaptation or localization of CEFR descriptors. This happens where existing descriptors do not quite fit into the context in which the resource is being prepared. While a broad discussion is beyond the scope of this handbook, users of the manual should be aware that the CEFR descriptors are not written in stone. The can-do statements can be:

- adopted as they are currently worded
- deleted if it is felt they are not relevant to the context
- adapted to suit a particular context or population
- re-written altogether (again to suit a particular context or population).

In the case of the latter two, you will need to consider conducting a study to validate the alternative wording or focus. This would rely on expert judgment much the same as the methodology used to build the CEFR itself.

Activity 2: Identify corresponding CEFR scales

Now look at the list of learning objectives resulting from Activity 1. For each one, ask yourself:

- Which illustrative scale is most applicable?

Depending on the nature of the objective, there may be more than one answer to this question. For example, if the objective is to assess whether a test-taker can plan and deliver a short, informative presentation, you may identify "Addressing audiences" (under *Production Activities*) as the most relevant scale. However, you may also consider that "Thematic development" (under *Pragmatic Competences*) is relevant. When you look at the actual descriptors in these scales you may find that one matches the learning objective more closely than the other. On the other hand, you may find that both are equally applicable, in which case you could include them both in the next step.

Activity 3: Identify the CEFR level

Within the scale(s) that you have identified, look for the descriptor which most closely matches the objective. Start with the descriptor for the targeted or expected level (the level that you have identified for this category of language activity, strategy or competence in the course of drawing up your generic specification).
If the descriptor comprises several "can do" statements, as is often the case, which one is most relevant? Then look at the descriptors for the adjacent levels, to check if the current objective is actually above or below the targeted or expected level. Identify the phrases in each descriptor that differentiate it from the adjacent levels. For example, if the learning objective is to be able to write short, connected texts about oneself and the target level is A2, look at the descriptors for A1, A2 and A2+ in the scale for "Creative writing". Identify the most relevant "can do" statements in those descriptors, with special attention to the phrases which specify the topics that the learner should be able to write about (shown in blue in Figure 3.5) and the type of text that they should be able to produce (shown in red). In this example, it would seem that the closest matching descriptor (as the learning objective asks for *connected* text) is actually that for A2+.

A2+	Can describe everyday aspects of their environment e.g. people, places, a job or study experience in linked sentences.
A2	Can produce a series of simple phrases and sentences about their family, living conditions, educational background, or present or most recent job.
A1	Can produce simple phrases and sentences about themselves and imaginary people, where they live and what they do.

Figure 3.5: Selected "can do" statements from creative writing descriptors for levels A1, A2 and A2+

Activity 4: Complete the detailed mapping

For each matching descriptor that you identify, record the corresponding learning objective, together with its reference (an indication of where it occurs in the structure of your resource, such as a page, unit or section number), the CEFR scale, the CEFR level and the “can do” statement itself. Summary Form 3.5 in the Appendix is designed for this purpose.

As well as documenting a detailed mapping of your resource’s alignment to the CEFR, the completed form will enable you to see or confirm certain general aspects at a glance. Looking down the column for “CEFR level”, for example, you will be able to judge how far the detailed mapping confirms the level profile that you completed in Summary Form 3.3, or whether the different components of an assessment are balanced with respect to target level. Note that Summary Form 3.5 also has space for you to record the domains, themes and text types associated with each learning objective. Once completed, these columns will also help you to judge how far the detailed mapping conforms to the overview that you produced in Summary Form 3.4.

3.4 Guidelines for reporting on the specification stage

The guidelines for reporting on the familiarization process (Chapter 2) also apply to specification, provided that additional detail is included on:

- the expertise of the analyst(s) in relation to the resource being analysed
- whether the outcomes of the activities reported in the forms were cross-checked, by whom and how.

The forms can also help you to record the activities completed and document their outcomes. Completed summary forms may be used in various ways:

- to stimulate discussion and reflection on the possible needs and challenges in the different phases of the alignment project
- to illustrate which specification activities have been completed, how and by whom (this constitutes a source of validity evidence in the alignment process)
- as part of the validity evidence to support any claims of CEFR alignment.

Notes for your own implementation …

If you are working on your own:

Completing specification activities is very time-consuming and you need to decide whether your alignment project requires only a generic approach (and will prioritize overall coverage) or also aims at completing specific activities (and will focus on overall coverage and detailed alignment).

- Avoid being too ambitious, aiming at too much.
- Make sure you have sufficient resources to complete the approach to specification you have selected (time; professionals who are familiar with the resource to be aligned to the CEFR and can help cross-check the analysis, etc.).
- Make it easier to complete Summary Forms by documenting decisions as you go along.

If you are a team coordinator, in addition to the above:

- Organize the team, defining clearly who will focus on what.
- Allow for sufficient discussion time.
- Do not rush analysts.
- Provide for cross-checking.
- Make sure you collect all relevant information you will need for the final report, including the Summary Forms at the end of each session.

CHAPTER 4: STANDARDIZATION

Standardization is the process of establishing that the main features of a given resource reflect a clear understanding of the relevant CEFR levels and descriptors. When two or more people are involved, the goal is to build a consensus regarding what a learner or test-taker can do at a given level and whether that corresponds to the level claimed by the resource. As with familiarization and specification, standardization procedures vary according to the resource, the purpose of the alignment, and the claim you want/need to make.

The process of standardization, what it entails, how to complete it and which instruments need to be used when aligning language exams or learner performances to the CEFR is described in detail in Chapter 5 of the *Manual for Relating Examinations to the CEFR (Council of Europe, 2009)* and is a prerequisite for standard setting. Our chapter here proposes an adaptation of the standardization procedures in the *Manual* to suit the needs for standardization if your resource is a curriculum or a coursebook.

This chapter is organized as follows:

By the end of this chapter, you should have a clear idea of:

- what standardization entails and how it links to the other activities in the handbook
- what approach is most relevant/useful for the context and for the participants
- how to carry out a standardization activity for your context
- how to document the activity.

4.1 What is standardization?

Standardization is the process of establishing that the main features of a new or existing resource – a curriculum, a coursebook or an exam – reflect a clear understanding of the relevant CEFR level(s). The participants use CEFR descriptors (or localized versions of CEFR descriptors) and illustrative examples (if possible, already aligned to the CEFR) to arrive at a common understanding of what a learner or test-taker can do at the CEFR level(s) in focus and how the level(s) can be operationalized for the resource in question. Participants in a standardization process should have prepared themselves by completing the familiarization and specification stages.

Standardization is a prerequisite for standard setting, but it may also be carried out at other times, for example, before writing learning or testing materials. In this way materials and item writers, editors and content managers achieve a shared understanding of the CEFR level their resource is aiming at. Alternatively, standardization may be carried out as part of a post-writing validation or evaluation. In this latter case, since the resource already exists and we are attempting to establish its alignment to the CEFR, we take a **bottom-up** approach: meaning we start with the resource and compare it to existing exemplar resources (e.g. test or learning tasks provided by the Council of Europe).

The situation with a resource either about to be developed or in the early stages of development is different, which is why we suggest that standardization can be usefully carried out early in the development process. In this situation (a **top-down** approach), at least one component of the resource needs to be developed in full before the standardization process can begin in earnest. The coordinator decides which component of the resource to develop (of those described in Specification Forms 3.4 and 3.5). Working in small groups (2–4 people), the participants develop that component. The groups then exchange their work and analyze it against Summary Forms 3.4 and 3.5 in order to answer the following question: **Does the component accurately reflect the CEFR level(s) it targets?**

In both the bottom-up and top-down approaches, participants first share their responses to the question in small groups. Then the coordinator opens the discussion to the whole group and encourages contributions with a view to building a consensus regarding which of the components developed best operationalizes the CEFR level(s) targeted. The knowledge gained from developing the component and building consensus will inform further work on this and future resources.

Usually, the standardization process focuses on one particular CEFR level, though in order to gain an understanding of the upper and lower limits of that level, it is necessary to take the adjacent levels into account. Where the context demands a broader proficiency range, standardization can embrace more than one level, which may include one or more ‘plus’ levels (A2+, B1+, B2+). It is important to bear in mind the CEFR’s profiling approach to language ability because different language activities and strategies are sometimes connected; for example, a high level of speaking ability may develop together with a high level of listening ability.

If you are working with a less widely used language or for a purpose other than testing, you will have more limited resources to draw on. But standardization can still be carried out effectively provided you make allowances to suit your context.

Standardization may be attempted individually, although in that case there will be no discussion and no consensus building. If you plan to work on your own, you should engage one or two colleagues or experienced professionals to act as critical friends throughout the standardization process and validate your conclusions. However, we would strongly recommend against doing this without at least a small team as the strength of claim that can be made for a single person is likely to be extremely low. It may be appropriate for low-stakes in-class materials or assessments, but not for any higher-stakes situation where the learner can be disadvantaged (e.g. end-of-year achievement test where promotion to the next level is based on test performance; preparation for an important examination where the materials are not targeting the correct level). For this reason, you will see that the remainder of this chapter assumes a larger group of participants.

4.2 Organizing the standardization process

Guidance and logistics play a key role in all phases of the alignment process, as is seen in Sections 2.3.1 and 2.3.2 of this handbook in the context of familiarization and in Chapter 5 in the *Manual* in the context of standardization.

The person responsible for coordinating the process will have to

- decide on the logistics and content of the session(s)
- prepare the materials for the successive activities
- write a report on the process,

and pay special attention to:

- the number of participants, their level of expertise and their relation to your resource
- the work you have already done in the familiarization and specification phases
- the identification/location of examples similar to your resource, if possible already aligned to the CEFR level(s) your resource is targeting
- the analysis of the main features of the identified examples in relation to the main features of your resource
- the discussions during the session(s) of your standardization event, which are key to building a common understanding of the CEFR level(s) as reflected in your resource
- the collection of evidence as the standardization activities are completed
- the documentation of the outcomes of the discussion(s).

4.3 Steps in standardization

The *Manual for Relating Language Examinations to the CEFR* (Council of Europe 2009, p. 37) lists four steps in the process of standardization.

1. Carry out CEFR familiarization activities.
2. Work with illustrative examples of materials, tasks or performances already aligned to the CEFR to achieve an adequate understanding of the CEFR levels. These tasks or performances are critical as all later decisions and claims will depend on how accurately they reflect the level.
3. Develop an ability to relate local materials, tasks and performances to the CEFR levels.
4. Ensure that this understanding is shared by all parties involved, and is implemented in a consistent fashion.

Before launching into the standardization process, we are assuming that the familiarization and specification stages will already have been completed. With this in mind, we suggest taking the approach identified in Table 4.1. This table can serve as a checklist of activities for the planners of the standardization process.

Table 4.1: Steps in the standardisation process

Step	Activity	Gloss
1	**Decide on a coordinator for the activity (the Stage Lead)**	Responsibilities include: ▪ decide on the logistics and content of the session(s) ▪ prepare the materials for the successive activities ▪ write a report on the process.
2	**Select participants**	Pay specific attention to: ▪ the nature and level of their expertise ▪ their experience in participation in similar activities ▪ the number of individuals overall (up to 20 for a high-stakes test, as few as 2 or 3 for a low-stakes (e.g., formative) test; 5–10 for curriculum or materials judging).
3	**Identify expected CEFR levels**	Curricula or materials will have been developed to reflect a specific CEFR level or range of levels. If a single level, extend standardization materials to one above and one below. If the curriculum was developed without the CEFR in mind, it is a good idea to separate this step into two stages: ▪ discovery (of the approximate range of levels) ▪ confirmatory (confirming or rejecting this – the originally intended activity).
4	**Identify CEFR illustrative examples**	A number of these are available (see Section 4.3.2). However, pay attention to: ▪ evidence to support the CEFR level alignment claim ▪ appropriateness of the activity to your context (if it is unlikely to be appropriate for use in your context, it should not be used here as a false understanding of level may result) – this work can be done by internal staff ▪ if no appropriate examples are to be found, you should develop your own by asking a sub-group of the most expert participants (internal and external) to review a range of your materials to identify appropriate examples

		▪ are the scales operationalized in the examples relevant to your context?
5	**Plan the event**	Before setting out on the standardization event plan it carefully: ▪ ensure that all materials are fully prepared ▪ ensure that the space is available and appropriate ▪ ensure that the time available is appropriate.
6	**Carry out the activities**	As shown in the following section, these focus on: ▪ re-familiarisation with the CEFR (may require little time depending on the participants) ▪ review the outcomes of the specification stage ▪ devise an agreed statement or description of expected learner achievements using the CEFR descriptors as a basis ▪ analyse illustrative examples. The key question to consider is: *Does the component accurately reflect the CEFR level(s) it targets?* If the answer is YES – you should proceed as planned If the Answer is NO – you should decide to either amend the resource accordingly or cease the alignment process and accept the limited claims made to date.
7	**Reflect on the outcomes**	Before completing the forms for this section, it is important to reflect on the process, taking into consideration: ▪ the report of the activities from the coordinator ▪ the feedback from the participants.
8	**Complete the forms**	This should be done soon after the completion of the activities (though some elements may be completed as individual activities are completed). The work can be done by one person if needed – the coordinator. However, it is best done by a small team reflecting internal and external participants. Following completion, a final review with all the participants will be useful feedback and increase the validity of the report and of the claims made.

4.3.1 Activities

Activities 1–5 in this section are aimed at the participants in a standardization process and should be completed in the order in which they are presented. You may need at least one day for each mode of communication focused on in the resource.

We recommend that participants keep a record of their individual responses to the activities for each mode of communication, using Standardization Form 4.1, which includes a question on the usefulness of the standardization activities and how they could be improved. We also recommend that you record the group discussions in audio or video. Both the participant forms and the recordings will be very useful to the coordinator when completing Summary Form 4.2 and writing the standardization report.

Activities 1 and 2: Review of familiarization and specification

All participants in the standardization process need to refresh their familiarity with the CEFR. This can be done by revisiting the forms already completed (Appendix, Forms 2.1–3). Especially when a larger group is involved, additional review activities may be appropriate; you may, for example, decide to provide the group with additional information or set some quizzes (see also Section 5.6.1 of the *Manual for Relating Examinations to the CEFR*).

The completed specification forms (Appendix, Forms 3.1–5) can be used to remind participants which CEFR scales are relevant to the resource they are working on. Those scales can provide further familiarization activities; for example, participants can sort descriptors into levels. It may also be useful to build consensus on which four or five scales are the most important and to establish that certain scales are more relevant to certain sections of the curriculum, coursebook or test under consideration. This helps participants to focus on the areas of language most directly targeted. In later activities, the participants can refer specifically to these CEFR scales.

Participants also need to review work done during the specification stage. One way of doing this is by creating a commonly agreed **statement of context** that includes the target audience and target level. The statement of context should respond to the questions in the box:

1. **What is the resource?**
2. **Who is the resource intended for? (Who is the resource *not* intended for?)**
3. **Which CEFR level(s) is the resource aimed at?**
4. **Which CEFR scales does the resource target?**
5. **Where and when will the resource be used?**
6. **What are the features of the resource that best exemplify the answers to questions 1–5?**

Activity 3: Draft a standardized description of expected learner achievements

Participants should use the statement of context and the relevant CEFR scales to write a draft description of what the target learner can do, using and amending wording from the CEFR descriptors. In preparing for some standard setting methods, “target learner” can be replaced by “minimally competent learner” or “minimally competent test-taker”. After performing this task individually, participants come together and share their descriptions first in pairs and then in larger groups until the whole group reaches consensus on a standardized description. For more on how to achieve and verify consensus, consult Sections 5.4.1 and 5.7.2 of the *Manual for Relating Examinations to the CEFR*.

Activity 4a: Analyse illustrative examples

The aim of this part of the process is to analyse and discuss the features of the illustrative examples in comparison with the features of your resource, so that you can check that your materials are an appropriate reflection of the target level(s), and that all participants in the session have a common understanding of the level(s).

Before analysing the illustrative examples, participants can be asked to respond individually to the following questions and to note their responses and comments for later discussion:

1. **Are the illustrative examples relevant to the proficiency level(s) I am concerned with?**
2. **Are the examples relevant to the context and purpose of my resource?**
3. **Are the examples relevant to the mode(s) of communication, language activities and language strategies central to my resource?**

A more detailed analysis can then be undertaken using the forms already completed as part of the specification activity (Appendix, Forms 3.1 to 5). The participants match the examples to the content of the forms, keeping a record of their analysis by ticking the boxes in the forms or writing a short comment. It may be useful to do this as an individual activity before repeating it in pairs and larger groups and thus arriving at a consensus. A more holistic approach can also be adopted, with participants completing a customised version of Form C3 in the *Manual for Relating Examinations to the CEFR* (see Form 4.3 in the Appendix of this Handbook).

In order to standardize learner performances, follow Section 5.5.1 (pp. 42–44) of the *Manual* for speaking and Section 5.5.2 (pp. 44–46) for writing.

If illustrative examples are not available, Activity 4b provides an alternative way of achieving a common understanding.

Activity 4b: Benchmark your resource

In the absence of illustrative examples, this activity requires participants to analyse their resource and decide whether it is an accurate operationalization of the CEFR scales and descriptors it claims to reflect.

For an already existing resource (a bottom up approach to standardization), a small group of the most expert participants (a mix of internal and external ideally) will be asked to consider the standardized description agreed upon in Activity 3 and analyse their resource against Specification Forms 3.4 (overview) and 3.5 (detailed mapping). They can perform these tasks individually and answer the following question on the basis of their analysis:

1. **Does my resource accurately reflect the CEFR level(s) it targets?**

The remaining participants are then given time to discuss their responses in pairs or small groups – typically led by the individuals from the smaller group who made the judgments. After that, the coordinator opens the discussion to the whole group and encourages contributions with a view to building a consensus regarding how well the resource operationalizes the CEFR level(s) targeted and what changes, if any, need to be made.

Activity 5: Completing the forms and writing the standardization report

This activity is obviously undertaken when all previous activities have been completed, recorded and studied. The work is typically led by the coordinator and will involve a group of co-authors, typically internal participants from the earlier activity groups. The work should be completed as soon after Activity 4 as possible. It is advisable to first complete the forms and then complete a relatively brief reflection activity on the process, considering all aspects of the activities (participation, discussion, outcomes) and the possible strength of the claims that can be made. The completed forms and report should then be reviewed by a separate group, either independently or through a group sharing event. This allows the coordinator to confirm (or not) the claims made and the decision to move to the next stage, standard setting.

4.3.2 Tools

The following websites provide access to illustrative examples aligned to the CEFR. You will, of course, need to check whether the examples they offer are appropriate for your context:

- Cambridge Assessment English sample videos:
- Cambridge English YouTube videos (speaking samples in English)
- Speaking tests illustrating the six levels of the CEFR – examples in French, German, English, Spanish, Italian:
- Council of Europe speaking, reading, listening and writing tasks
- Council of Europe Writing Tasks: Pilot Samples – examples in English, French, German, Italian, Portuguese
- English Profile (English Vocabulary Profile, English Grammar Profile and Text Inspector) (for understanding the CEFR level of texts written in English)
- Merlin Platform
- Plan Curriculas Instituto Cervantes (examples for Spanish)
- Profile Deutsch (examples for German)
- Profilo Lingua Italiana (examples for Italian).

4.4 Guidelines for reporting on the standardization stage

The guidelines for reporting on the familiarization process (Chapter 2) also apply to standardization, and as in previous chapters, the forms completed during the process provide an overview of the activities completed and the order in which they were carried out. Remember that the reporting of the work undertaken forms the basis of your alignment claims, so you should carefully consider your target readers and the message you are trying to get across when putting your report for this and other stages together.

In order to show that the outcomes and interpretation of the standardization are valid, it is necessary to report **how and why** things happened throughout the standardization process. Each step in the process should be described and a rationale provided for each decision made. The following need to be included:

- a list of participants, with brief biodata
- the names and responsibilities of the people involved in the organization of the session(s)
- the schedule of the session(s).

Notes for your own implementation …

Completing standardization activities is very time-consuming, so before you begin you should decide on the scope of the exercise and know where you will obtain the all-important illustrative examples. Here is a final reminder of key essentials.

- Decide on the method and intensity of your standardization according to the needs of your context and available resources. Don't be over-ambitious but try to ensure that you are doing enough to support whatever alignment claims you wish/need to make.
- Ensure that you have sufficient resources, especially time and access to professionals who are familiar with the resource to be aligned to the CEFR and can help cross-check the analysis.
- Organize the team, defining clearly who will focus on what.
- Allow for sufficient discussion time.
- Provide for cross-checking with external experts and stakeholders.
- Make sure you collect all the information you will need for the final report, including the Summary Forms at the end of each session.
- Make sure you collect and store data securely and in conformity with legal requirements (e.g. GDPR in the EU). This may mean making clear to those who provide the data and to potential users the full range of uses to which it may be put.

CHAPTER 5:
STANDARD SETTING

Standard setting is rooted in the empirical process of gathering quantitative evidence to make appropriate *cutscore decisions* on *score scales*, something most closely associated with tests and assessments. The methodologies associated with standard setting were developed for aligning tests and assessments with standards, qualification levels, and frameworks defining criterial features of proficiency. Standard setting helps answer the question: *at what point on the score scale of a particular test can we conclude a test-taker has sufficiently demonstrated the proficiency described for a particular level within a set of standards, qualification levels, or framework of proficiency?*

This chapter thus has a stronger focus on testing and assessment than other chapters. This does not mean that standard setting is not relevant to the other resources addressed in this handbook. But it does mean that it will not always be practical to attempt to apply fully the methodologies relevant to tests and assessments to these other resources.

This chapter is organised as follows:

By the end of this chapter you should have a clear idea of:

- **the concept of standard setting**
- **how standard setting relates to the other stages of the alignment process**
- **the practical steps in the process of planning and carrying out standard setting**
- **the logistical and resource demands of standard setting**
- **the nature of typical data collection activities used for standard setting**
- **how standard setting relates to the different kinds of resources described in this handbook (assessments, curricula, textbooks etc.)**
- **the importance of documentation and reporting**
- **where to go to find more detailed technical information on standard setting methods and examples of their application.**

Note: All standard setting methods involve statistical analysis of quantitative data, but we will only describe these processes in outline. Section 5.6 will provide a list of resources for pursuing these technical aspects of standard setting in more detail.

A note on this chapter: setting expectations

As noted, standard setting was designed for use with tests and assessments, which explains why Chapter 5 takes a slightly different approach to the other chapters. The handbook focuses on the evidence we need to collect to support the alignment of various resources to the CEFR: curricula, teaching and learning materials, and assessments. Chapter 1 presents these resources as the components of a Comprehensive Learning System (see Figure 1.2 in Section 1.4). Because they are designed to perform different functions, these components do not share the same features, and we do not expect to implement the five stages of alignment described in Chapter 1 in exactly the same way for each part of the CLS triangle. Assessment is often the “hard edge” of how we evaluate student attainment of curriculum goals and the efficacy of curriculum delivery. It is thus important to look at standard setting specifically in relation to assessment to ensure the fairness of the judgments we make about students and the evaluations we make of other parts of the CLS through the use of assessment tools aligned to the CEFR. Accordingly, in Sections 5.1–3, we focus on standard setting in relation to assessment instruments, while in Section 5.4 we discuss how some of the evidence-based approaches to demonstrating alignment for assessments through standard setting can be adapted to add to the robustness of alignment claims made for other resources too.

5.1 What is standard setting?

5.1.1 Why do we do standard setting?

Standard setting is described by Cizek as “The proper following of a prescribed, rational system of rules or procedures resulting in the assignment of a number to differentiate between two or more states or degrees of performance” (see Section 5.6, resource 3). This definition reflects the origins of standard setting in the United States, particularly in large-scale standardized assessments. Arguably, however, in the two decades since its publication in 2001, the CEFR has inspired a great deal of innovative work on applying and adapting standard setting methodologies. The goal of such work is to identify the point on a test score scale at which a test-taker can be considered to have demonstrated the level of proficiency described in one of the CEFR levels. If an assessment is to report student scores in relation to CEFR levels then, as Kaftandjieva (2004) notes, “standard setting is at the core of the linking process”.

As noted in Chapter 3, the process of alignment is “top down” if you are designing a new assessment on the basis of the CEFR and “bottom up” if you are attempting to align an existing assessment which may not have been designed in relation to the CEFR. In either case, before beginning to plan the process of standard setting, you will have carried out the steps outlined in Chapters 3 and 4. You will thus have a detailed understanding, not just of the CEFR level(s) your assessment is related to, but of the specific scales and descriptors to which your tasks and items should be aligned. You will also have developed clear test specifications and have data on the technical performance qualities of your assessment, such as reliability etc. As noted in the *Manual for Relating Examinations to the CEFR* (Section 5.6, resource 4), “linking a qualitatively poor examination to the CEFR is a wasted enterprise which cannot be saved or repaired by careful standard setting”. It is important to remember that standard setting itself, “successful” or not, does not constitute validation of the assessment.

It is also important to emphasize that standard setting and validation are no less relevant to small-scale institutional and classroom assessment than they are to large-scale high-stakes tests.

If the alignment process is not to be a "wasted enterprise", assessment at all levels should follow principles of good test development: test specifications, clearly defined rating scales, and the analysis and evaluation of test results. There are many useful discussions of how good test development and validation principles can be applied to classroom assessments.

On the basis of the CEFR-related evidence you will have gathered in the specification and standardization stages (Chapters 3 and 4) and the technical evidence that confirms the quality of your test, you will already have a fairly clear idea what the cutscore(s)[4] for the CEFR level(s) you wish to report will be; standard setting should not tell you something completely new about the meaning of your test scores. Nonetheless, and although it is sometimes difficult for non-specialists to accept, tests are never absolutely accurate and precise instruments and always contain some measurement error[5]. Standard setting is thus required for two reasons. First, you need to identify where cutscores actually sit on the score scale by refining and adjusting your initial estimates; and secondly, you need to provide stakeholders with empirical evidence to justify how you decide to report test results in terms of CEFR levels. Although we noted above that standard setting is not in itself validation of the assessment, it does contribute to validation. If your standard setting evidence does *not* confirm that your assessment is aligned to the CEFR, you will nevertheless have gathered information that can be used to revise the design and operation of the assessment.

5.1.2 How do we carry out standard setting?

Standard setting is a way of collating quantitative, empirical evidence to determine scoring decisions in relation to the CEFR. But the quantitative nature of standard setting should not lead to superficial trust in the "numbers". It is important to emphasize that all major methodologies involve, to some degree, human judgments; and there is no one-size-fits-all methodology or approach. A number of standard setting methods are available, with a solid body of work now available on their implementation in relation to linking exams to the CEFR (for case studies, see Section 5.6, resources 5, 6, 7, 9, 10 and 11).

Your choice of standard setting method will depend on a number of factors. Some methods have been developed for, or have been shown to be most appropriate for use with, specific types of assessment. Thus, when choosing a method it is necessary to take the features of the assessment itself into account – receptive or productive skills, response formats, etc. Other factors must also be taken into account, including:

- available resources
- the experience and expertise of the person or team in charge of standard setting
- the availability of experts to take part in the standard setting process
- access to test content, test data and sample student performances on the test.

Standard setting is inherently practical, and careful attention must be paid to logistical arrangements before, during, and after the application of whatever method is selected and also to the more technical aspects of collecting and analysing the standard setting data itself.

---

4 A cutscore is the lowest possible score on a test or other assessment that can be considered a pass – in this case, it's the lowest point at which we can still claim a learner is at a specific CEFR level.

5 Measurement error refers to the difference between the observed value of a variable (here a learner's test score) and the true value of that variable (the actual ability of the learner).

This chapter does not explain in detail how to apply the various methods available. However, we summarize some of the key features of the most commonly used methods. These features are relevant to the processes and logistical considerations which will be discussed in Section 5.1.3. In planning for standard setting, you should address the following considerations.

- It is usual to distinguish between ***test-centred*** methods, which make judgments about test items, tasks, etc., and ***examinee-centred*** methods, which make judgments about samples of student performance on the test or evidence of students' proficiency level before they take the test.
- Judgments are made by a panel of experts on the basis of a task defined by the method selected, and judges need to be trained in the use of that method.
- You need access to test content that is representative of the assessment to be aligned to the CEFR. In examinee-centred approaches, it may also be necessary to collect student performance samples, including the scores they have been allocated by whatever scoring system is employed by the test design.
- You need to determine procedures to analyse the data collected.
- You need a decision rule for setting cutscores based on the analysis of the data collected in accordance with the method selected.

In discussing the **how** of standard setting, it is worth noting that the other stages of alignment summarized in Chapter 1 – familiarization, specification, standardization and validation – are essential steps in the development and alignment of the resource in question prior to standard setting. To a greater or lesser extent they are also contained in microcosm in the standard setting process:

Familiarization is an essential element of training to ensure that the panels of experts fully understand the criterial features of the aspects of the CEFR that will be the focus of standard setting.

Specification is something that should be carried out by you, the person responsible for the alignment process, before you embark on standard setting. Specification is critical to understanding the intended alignment of your assessment or other resource to the CEFR. If you are developing a new assessment or resource, specification is a critical part of the development process. A supplementary specification activity can provide powerful evidence to triangulate the data collected from the quantitative standard setting process. Carrying out specification as a complement to standard setting is particularly important when you are aligning an existing resource that may not have been designed with CEFR alignment in mind.

Standardization is also a critical step in test development prior to standard setting, particularly for item writers and other key stakeholders involved in the development process. Some of the standardization activities described in Chapter 4 can be used in training for standard setting, particularly when collecting locally appropriate examples of performance benchmarked to CEFR levels. Finally, it is important to collect evidence that will help you analyse and evaluate your standard setting process and determine whether your cutscore decisions are plausible and defensible.

5.1.3 Standard setting practicalities: planning, organization, data capture and analysis

As noted above, standard setting is an essentially practical activity with significant logistical demands. We want to emphasize, however, that many standard setting methods can be adapted for use in resource-constrained contexts and on a small scale, for example, with in-house assessments developed within a particular programme or institution. Section 5.2 presents some examples of standard setting methods in relation to relatively straightforward language test formats and based on scenarios typical for medium to large organizations. While these examples can be relatively resource-intensive, in Section 5.3 we discuss some possible alternatives and adaptions for smaller scale situations.

We cannot overemphasize the importance of focusing on the practical aspects of the process when planning for standard setting. Table 5.1 lists some of the key steps and decisions which need to be taken when you are planning and preparing for standard setting. While the list may look daunting, in Section 5.3 we discuss what these aspects might mean for smaller-scale, resource-constrained contexts.

Table 5.1: Key steps in planning and preparing for a standard setting activity

Selecting a method
You will need to become familiar with the literature on standard setting. You may have the financial resources to commission an expert to carry out the necessary activities. Nonetheless, you will still need to have a grasp of the options available if you are to evaluate critically your expert's recommendations and justify your choices to key decision-makers. In many instances, you may be required to plan, lead, and carry out the standard setting yourself.
Identifying the resources required to carry out the standard setting
Physical: equipment (e.g. laptops, specialist software, sound equipment for playing listening files); physical space or online platform for standard setting meetings. **Financial:** remuneration for judges; costs of preparing and printing materials, renting laptops for judges, etc.; securing a physical space for standard setting activities etc. **Human:** administrative support for standard setting activities; training for judges; delivery of the process (e.g. handing out and collecting materials, assisting judges with technical issues etc.); analysis of the data.
Identifying relevant experts for the panel or panels of judges
You will need to establish criteria for recruiting and accepting judges. It can be a challenge to identify suitably qualified people who are both willing and available to take part. For the methods described in Section 5.2, 12–15 judges is considered appropriate. However, if you are planning on linking multiple tests, or carrying out standard setting across reading, listening, speaking, and writing, you may need to consider having more than one panel. See Methodology Box 3 in this chapter.
Ensuring access to the test materials required
For test-centred methods, this may be actual test content, i.e. test items. While this may seem straightforward, using real test items brings with it security considerations. For listening tests, it is necessary to access the actual recordings used for input, and for computer delivered tests, panellists will need to be able to view the test content in the same way test-takers do. The method may require the use of actual spoken or written performances by test-takers.
Ensuring access to test data required for analysis
Usually, it is important to have item difficulty data for the test items from typical test-taker populations. For examinee-centred methods that use actual test performances, you may require the ratings given by trained examiners.

Preparing communications
This will include letters of invitation for judges that outline the purpose of the project, as noted in Chapter 4, GDPR-compliant consent forms, and instructions on the logistics for participants. It may also contain a Non-Disclosure Agreement (NDA) where live test materials and performances are to be used.
Developing training materials for the CEFR
For the CEFR: Familiarization and standardization aspects of training as described in Chapters 2 and 4. **For the method:** Descriptions of effective training adapted from examples of the method you select.
Preparing the data collection instruments
Judgment forms: To avoid input errors, forms for individual judges' decisions need to be easy to use. It may be worth considering digital judgment forms. **Feedback forms:** Feedback from judges is an essential part of procedural validity evidence.
Preparing the tools for data collation and analysis
Many methods recommend carrying out more than one round of judgments and showing results from interim rounds to the judges for discussion. This requires you to collect, collate, and analyse the judgment data rapidly during the standard setting activities. For this purpose you should prepare formatted spreadsheets or other ways of collating individual judgments into data sets.
Preparing the physical space needed
It is important to use a location large enough to accommodate the panel comfortably. You will need to ensure that any equipment required is at the location, and you may need additional space for securely storing sensitive test material. Overnight accommodation, meals and refreshments may also be required.

While the examples of actual standard setting events presented in Section 5.2 take place over two or three days, it should be clear from the list of logistical considerations in Table 5.1 that planning and preparation need to begin well in advance of the actual event. As mentioned earlier, Form 5 is designed to help with the planning of the standard setting stage by offering you a checklist of key points to consider.

5.2 Standard setting examples for reception and production

In this section we present real-world examples of the sequence of activities at the heart of data collection for standard setting when applied to language tests for the purpose of alignment with the CEFR. The focus is on schedules and the order of activities, not the details of, for example, the selection of judges or the collection and analysis of judgment data. The methods selected are workable, defensible and practical and have been widely used, both in educational measurement generally and in relation to the CEFR. The methods adapted for use in the examples in this section are described in Section 5.6, resources 2, 3 and 4; detailed technical reports of the application of the particular combination of methods will be found in Section 5.6, resources 9 and 10. These resources deal with linking studies for locally developed standards not the CEFR itself. However, they contain appropriately detailed descriptions of data collection, statistical analysis procedures and results, procedures for determining cutscores, and validation evidence.

One reason for using the methods below as examples is that they contain many of the key steps and features which are shared by the most commonly used methods. They thus give a clear indication of the steps involved in a standard setting event, regardless of the method used and the time required to carry them out successfully. However, the fact that they are used as examples should not be taken to imply that they are the best methods for your particular context. Variations and alternatives are discussed in Section 5.3.

5.2.1 A standard setting example for reception

We present a schedule for a standard setting event for a listening test on the following page. It combines two commonly used **test-centred** methods: the **Basket method** and the **Modified Angoff method** (see Methodology Box 1). In our example, the Modified Angoff data is intended from the outset to be used for setting the cutoffs. The methods were combined to complement the strengths and counter the weaknesses of each. The Basket method was used as a warm-up activity to get the judges to think of the items in terms of CEFR levels. These initial judgments provide a starting point for range-finding for the probability judgments made using the Modified Angoff method, which reduces the cognitive complexity of this task.

Once again, our purpose in presenting these examples is not to suggest that this is the only or best way to approach standard setting but to emphasize that methods can, and where appropriate should, be adapted by you to suit your context. However, when doing so it is essential to have a clear rationale for your adaptations, to document them explicitly, and to validate them through analysis of the results: did the adaptations work?

Table 5.2: Sample schedule of activities for a listening test

Day 1	
9:30	Introduction to the scope and goals of the seminar: self-introductions
10:00	Overview of CEFR: principles, development methodology, structure
10:30	Break
10:45	Refresher: activities with key descriptors for listening
11:30	Discussion: reflections on the activities in the self-study booklet[6]
12:30	Lunch
14:00	Overview of the test to be used in the project
15:00	Standard setting with listening items: explanation and practice
16:30	Break
16:45	Final discussion (agree on a list of descriptors, and criterial CEFR features)
Day 2	
9:00	Review: descriptors, list of criterial CEFR features, and the listening test
9:45	Listening test (BASKET METHOD)
11:00	Break
11:30	Discussion, presentation of item difficulty data, (Panellists can make adjustments or re-listen to particular items by choice)
12:30	Lunch
14:00	Listening test (MODIFIED ANGOFF METHOD), Round 1
15:30	Break
15:45	Housekeeping and review plans for final day
16:00	Discussion, presentation of item difficulty data
16:30	Listening test (MODIFIED ANGOFF METHOD), Round 2

[6] This links back to Familiarization and consists of either a paper or digital presentation of a series of tasks adapted by the coordinator for self-access from the Council of Europe's exemplar tasks for the targeted CEFR level(s). The booklet should be completed by the panel members prior to the standard setting event.

It is necessary to elaborate on some of the key features of this schedule.

- The panel of judges are also carrying out standard setting for the reading component of the same test. It is easy to see that the time required for the activities quickly adds up, making it difficult for judges to be available for all the activities as they extend over several days. This is one of the most common logistical challenges in organizing a panel of judges. The solution, discussed further in Section 5.3, is to present many of the familiarization and standardization activities as self-access materials which the judges are required to carry out before the two-day standard setting event. This makes it possible to allocate most of the available face-to-face time for training in the method(s) to be used and actual standard setting.
- Note that for the main standard setting method which will be used to set the cutscores, two rounds of judgments are carried out, with discussion in between. This is a common feature of many methods. Because in this example, the Basket method is used for a different purpose, only one full round of judgments is carried out, although discussion is still important.
- During discussion, it is common to present the panel's judgments as a basis for identifying major areas of agreement or disagreement. Note that the break of 30 minutes between Round 1 judgments and discussion allows the facilitators to collect the judgments, collate them into Excel format, carry out preliminary analysis and transfer the results to graphic format for presentation to the judges.
- Before judges are asked to make their first judgments with the Basket method, they are asked to take the listening test under the same conditions as test-takers, listening to the recordings used in the test with the same time constraints and responding to the questions. They are then given the item key so that they can check their answers before proceeding to make their judgments. To assist them in this process, judges are given a transcript of the recorded element of the listening items. This saves time and helps them to switch their focus from test-taker to expert judge. An additional piece of useful evidence can be gathered by asking the judges to reflect on the accuracy of their decisions either through a brief interview or using a Likert type reporting tool (i.e., a scale of 1 to 5 with 1 not accurate and 5 highly accurate).

5.2.2 A standard setting example for production

On the following page, we present an example of a standard setting event for the speaking and writing components of a test. It employs a commonly used examinee-centred method, an adaptation of the **Analytic Judgment method** (see Methodology Box 2). The example deals with two test components in order to emphasize again how we need to adapt our approach for different situations. As in Section 5.2.1 our purpose is to illustrate some key common threads that will be applicable across methods, not to suggest that the method described or its application is the only or best approach. The schedule differs in important ways from the one presented in Section 5.2.1, again in order to show how adaptations can and should be made to suit the needs of different contexts.

Table 5.3: Schedule of activities for tests of Speaking and Writing

Day 1	
9:00	Introduction to the scope and goals of the workshop; self-introductions
9:30	Refresher on CEFR speaking/writing: activities with key CEFR descriptors for spoken and/or written production and/or interaction as required
10:30	Break
11:00	Finalize description of spoken production and/or interaction
12:00	Lunch
13:30	Refresher on CEFR writing: activities with key CEFR descriptors for writing ability
14:30	Finalize description of written production and/or interaction
15:30	Break
16:00	Overview of the tests to be used in the project
17:00	Standard setting with speaking items and writing: explanation and practice
17:30	Housekeeping and plan for Day 2
18:00	Finish
Day 2	
8:30	Review of distinguishing features of spoken production/interaction
9:00	Round 1 judgments of speaking performances (includes a reflection on accuracy of judgments)
12:30	Lunch
14:00	Presentation and discussion of panellists' judgments for common performances
14:30	Round 2 judgments of speaking performances (includes a reflection on accuracy of judgments)
17:45	Housekeeping and review plans for final day
18:00	Finish
Day 3	
8:30	Review of distinguishing features of written production/interaction
9:00	Round 1 judgments of writing performance samples (includes a reflection on accuracy of judgments)
12:30	Lunch
14:00	Presentation and discussion of panellists' judgments for common items
14:30	Round 2 judgments of writing performance samples (includes a reflection on accuracy of judgments)
17:45	Ensuring all materials have been collected
18:00	Finish

Again it is necessary to elaborate on some of the key features of this schedule.

- This example combines two skills in a single standard setting event. As noted above, the availability of judges is a constant challenge, so it is important to exploit potential synergies to make the most of the time available for data collection.
- In this example, the method used is the same for speaking and writing. This allows participants to become increasingly familiar with the method throughout the event, which lasts for three days.
- The example assumes that the judges are familiar with the CEFR and that the panel recruited for the receptive skills standard setting (the listening example above and also a separate standard setting event for a reading test) is also used for the productive skills event. This makes it possible to omit the general overview of the CEFR and the background to its development etc. from the schedule.
- As with the listening example, this schedule presumes that judges will have been given familiarization and standardization activities as part of a self-access training package. It is thus possible to focus on refresher activities for the CEFR followed by training for standard setting and data collection.
- Because the same approach and procedures are used for speaking and writing, we have put the refresher activities, description of the tests, and training in the method on day 1 for both skills.
- Note that both of our examples use two rounds of judgments, with the presentation and discussion of preliminary results between rounds. This is an important feature of most commonly used methods.
- However, the schedule for each round of judgment collection differs due to aspects of the method employed and the nature of the test responses used as the basis for judgments.
    - In this example, judges are given access to the speaking and writing performances on secure laptops specially provided for the event. This allows them to access the performances at their own pace, re-reading or re-listening as required. Thus, in this schedule a large block of time is allocated to each judgment round with no defined break. Judges are allowed to take breaks as needed to get refreshments, take comfort breaks, or get fresh air and clear their head. (This means making refreshments available on demand rather than at specific times).
    - In the listening example (5.2.1), 30-minute coffee breaks between rounds allow the facilitators to collate the judgment data and carry out preliminary analysis. In the speaking and writing example, given the longer block of time needed for judges to review the sample performances, a longer lunch break allows the facilitators to carry out the analysis and prepare the graphic presentation of results before the inter-round discussion.
    - Note that discussion between judgment rounds focuses on “common items”. To collect a sufficient number of ratings across a large enough sample of performances, judges are split into smaller groups of three or four. They are presented with a common set of performances and a set unique to their group. The full-group discussion focuses on the results for the common set as these are the only performances that all judges have considered.
- At the end of day 3, time is allocated to ensuring that all materials have been collected. In this example, performances from live examinations are used, and it is essential to protect the security not just of the test content but of the sample performances. Depending on the materials used, the nature of the test, and the location in which the event is held, this step may need to be carried out at the end of each day.

Methodology Box 1: Basket and Modified Angoff methods

In the **Basket** method, judges decide for each item the CEFR level at which a candidate will first be able to answer it successfully. They then put that item in the appropriate CEFR level "basket". The Basket method was quickly adopted for CEFR linking projects and at one point was the most widely used method for this purpose. It has, however, been the focus of criticism. Kaftandjieva (see Section 5.6, reference 9) has noted that the judgment task introduces inherent bias to the results. This is particularly problematic when a test is designed to focus on a single CEFR level and contains items which cover a very narrow range of difficulty within that level. At the same time, the method is easy to grasp and thus accessible to judges.

The **Modified Angoff** method, while also the subject of criticism for the cognitively demanding nature of the judgment task, has been shown to be statistically robust and has been widely used and researched. In this method, judges make judgments about the probability of a candidate at a defined level of ability successfully answering the item.

Note: Since introducing the two approaches can be challenging for panel members, it may be more useful to introduce the Basket method as a pre-standard setting activity and just use the Modified Angoff method on the day or days of the event.

Methodology Box 2: The Analytic Judgment method

For the Analytic Judgment method (AJM), the judgment task is conceptually similar to the Basket method, though the decision rules for setting cutscores differ, which avoids the bias issues mentioned in Methodology Box 1. Judges review actual performances on the speaking or writing test. It is essential that prior to the standard setting event, performances have been scored by trained raters using the scoring scales developed for the test in question. Judges are not informed of the scores the performances received. They allocate each performance to the CEFR level which best reflects the ability of a candidate likely to produce it.

After the judges have allocated all the performances to CEFR levels, various decision rules are available for setting cutscores. Essentially, they all boil down to the same thing: the facilitators use the pre-existing test scores to define the cutscore points on the test score scale that best differentiate between the CEFR levels. As noted in Chapter 4, Section 4.3, for this reason it is important to have a large number of performances, and to have multiple performances representing each step on the test score scale. This ensures that judges have to read or listen to multiple performances in much the same way as an examiner allocating marks would do.

Methodology Box 3: Key questions in selecting expert panel members

Question 1: What range of expertise do I need to identify?

1. Experts of the resource of interest, e.g., item writers and exam raters for a test; textbook writers and editors and classroom teachers for a textbook etc.
2. People who are most familiar with learners targeted, usually classroom teachers.
3. People who are experts on the standard (CEFR in this case).

Question 2: How can the panel represent a diversified view?

1. It is important to understand which subgroups can provide different valuable perspectives on the resource.
2. It is not a given that people from similar groups (e.g., experience, gender etc.) will act in a similar way so there's no point in forcing even numbers.

5.3 Challenges, solutions and alternatives

This section briefly considers how the scope and scale of standard setting activities can be adjusted to fit resource-constrained situations, e.g. a programme in a single school or institution. It also mentions some alternatives to the methods described in Section 5.2.

5.3.1 Adjusting scope and scale

While Cizek's definition of standard setting emphasizes the "proper following of a prescribed, rational system of rules or procedures", we believe it is not only acceptable but entirely sensible to adjust methodologies to suit the context. In test development generally, indices of technical quality, e.g. reliability estimates, inter-rater agreement, etc., are often discussed in terms of ranges of acceptable results depending on the "stakes" or importance of decisions to be made on the basis of the test and the ability of test users to make later adjustments to those decisions. Lower reliability estimates, for example, might be deemed acceptable for a teacher-designed classroom reading test. The teacher probably does not have access to an extensive item writing and pretesting system but will be able to monitor the impact of decisions made on the basis of the test results, such as the assignment of students to different ability groups for a project. The teacher will also be able to identify issues for particular students who have been misplaced and adjust group membership or support for students accordingly. We can apply a similar logic to standard-setting methodology.

One of the biggest challenges posed by standard setting is the need to secure a sufficient number of expert judges. If you are developing in-class progress tests with a colleague at your school, you probably lack the financial resources to recruit a panel of 15 experts. But you may be able to put together a small panel consisting of colleagues at your school. While the results of your standard setting project may need to be treated cautiously in terms of setting definitive cutscores, the processes of familiarization and standardization and providing standard setting judgments on actual test items will be an invaluable professional development experience for all involved.

In small-scale, in-house situations, you may not need to set a definitive cutscore on a complete test: your interest may be rather to ensure that you have a bank of appropriately targeted items to use for practice activities, homework, ongoing progress checks etc. In these circumstances, carrying out a Basket Method activity with a small number of fellow teachers could be useful. Because you do not have to set a cutscore for the test, the potential bias in the method need not be a concern. If consensus is achieved, however, the item level judgments will indicate whether or not the tasks measure the CEFR level abilities in focus.

Your colleagues' judgments on items you have developed can confirm that you are on target or highlight issues that are easy to miss when evaluating one's own test items. And working together allows for a reality-check on what is happening in your school outside your own class. *Are students in different classrooms with different teachers being presented with material that is targeted at the correct level? Is the coverage and difficulty of the items in the tests consistent?* Using standard setting techniques with even a small group of five or six colleagues can identify potential issues and give you and your colleagues confidence that you are working in the right direction – or help you to make adjustments if necessary. In such situations, you can also be much more flexible in how you carry out the activities and collect data than in the examples we provided in Section 5.2.

While it is still useful to collect judgments in a controlled and consistent way, when you are working with colleagues in-house it may be possible to carry out many of the activities individually and to come together only to discuss your judgments between rounds and to evaluate the final results.

5.3.2 Making adjustments in a defensible way

Before considering what adjustments you might usefully make, you need to ask yourself:

- What are you hoping to achieve by demonstrating the alignment of your tests with the CEFR?
- Who are the key stakeholders to whom you need to demonstrate this alignment?
- What kind of decisions will be made on the basis of alignment with the CEFR?
- And critically, what resources and expertise can you access to achieve your goals?

We do not recommend that you make changes to established methods lightly. But in a resource-constrained classroom context, we believe that adjusting methods to ensure that you have some degree of empirical alignment data is preferable to not attempting any standard setting because you cannot replicate a larger-scale standard setting event. However, before making changes, you first need to understand what you are changing and the impact your proposed modifications may have. Also, you should clearly document the rationale for your changes and what they involve.

5.3.3 Online standard setting

An important development in recent years has been the move to remote, online standard setting. This has happened in the field of educational measurement generally (see, e.g., Section 5.6, resource 3) and in standard setting in relation to the CEFR (see resources 16a and 16b – as these are unpublished, we suggest that you contact the author for details). Online standard setting allows the main data collection activities to be carried out without bringing the judges together in one place. This reduces the cost of large-scale projects because it is no longer necessary to rent a physical space for the several days required for the event or to reimburse judges' travel and accommodation expenses. A further benefit is the potential to widen the pool of judges by including individuals who for practical or cost reasons would not be able to participate if required to travel to the event. Of course, online standard setting also imposes its own logistical and other challenges, such as a stable internet connection for all participants, which need to be considered carefully. However, following the large-scale move to online teaching and assessment in 2020, it is likely that many people and organizations will be much better prepared and open to online standard setting in the future.

5.3.4 Other standard setting methods

Section 5.6 lists resources which give overviews of methods. It is important to engage with this literature in order to evaluate what will work best in your context. Here, we draw attention to some possible alternatives to the examples we gave in Section 5.2. In doing so, we again emphasize the importance of doing your own research, being flexible, and looking for what is feasible for your situation.

There are many alternatives to the three methods we described briefly in Section 5.2. Two alternative test-centred methods are the **Bookmark method** and the **Item Descriptor matching method** (Section 5.6, resources 4, 5 and 6). The Bookmark method, which utilizes statistically (IRT)-generated difficulty values while presenting panellists with a fairly straightforward judgment task, is now one of the most popular methods in the United States and has also been employed for linking tests to the CEFR. The Item Descriptor matching method has been less used in relation to the CEFR (for an example, see Section 5.6, resource 12). Its judgment task has similarities to the Basket method and has the benefit of explicitly involving CEFR descriptors. Resource 2 in Section 5.6 describes an interesting adaptation of the Basket method which has the potential to overcome some of the bias noted in Section 5.2 while maintaining the conceptually accessible judgment task which makes the method attractive.

For many examinee-centred approaches, one of the biggest challenges is securing access to a sufficient number of examples of examinee performances, particularly if, as with the Analytic Judgment method described in Section 5.2.2, those performances need to have been previously scored by trained raters. For productive skills tests, using actual student performances is obviously desirable. But in cases where this is not possible – for example, when performances cannot be collected in advance for fear of compromising the security of the test content – applying test-centred methods may be an option. For example, the Item Descriptor matching method study (Section 5.6, resource 12) was applied to evaluating writing task demands.

Resource 6 in Section 5.6 describes an application of the **Contrasting Groups approach**, another examinee-centred approach. This approach asks teachers to evaluate the CEFR level of students they are familiar with, and then collects test data from the same students. Test scores for students placed at B1, for example, can then be compared to scores for students placed at B2 by their teachers, and cutscores determined. This can be a powerful way of triangulating standard setting results from other methods or the focus of standard setting itself and is particularly useful for engaging teachers and other stakeholders in the project.

5.3.5 Standard setting across languages

While the majority of published studies focus on linking a test or set of tests for a single language to the CEFR, two of the resources in Section 5.6 describe innovative approaches to aligning test tasks and performances across **different** languages. Resource 1 in particular was designed to provide confidence that different linguistic and educational contexts share a common interpretation of the CEFR, and there are many local and institutional situations where such an approach might be useful. For example, in a university modern languages department, where exams are developed for different languages by different groups of teachers, it is important to demonstrate that criterial features of the same CEFR levels are being targeted in a coherent and consistent way. Resource 1 describes an examinee-centred approach for spoken performances in English, French, German, Italian and Spanish. This study also includes two different judgment methods: "evaluating candidates both through comparison with one another (ranking: norm-referenced) and with reference to the CEFR (rating: criterion-referenced)" (p. 4). Resource 8 describes a test-centred approach that employs the two methods in Section 5.2 to place reading tasks from different tests for English, French, German, and Spanish on a common scale (as this is unpublished, we suggest you contact the authors for details). Both studies used Multi-Facet Rasch Model analysis to make comparisons at the task level rather than to set cutscores on full tests, so they are equally relevant to the standardization activities described in Chapter 4, especially the benchmarking of performance samples.

5.4 Adapting standard setting methods for resources other than assessments

As already noted, for resources such as curricula and teaching and learning materials, standard setting may not be the most appropriate alignment tool. This doesn't mean that standard setting is not relevant, however: it remains part of the overall system of alignment to the CEFR through its application to assessments that should be integrated with the curriculum and its delivery using appropriate materials. In gathering evidence for the alignment argument, particularly for curricula, the specification and standardization stages described in Chapters 3 and 4 are most useful. However, it is clear that judgments should not be made either by a single individual even if not directly attached to the team who developed the curriculum. Instead the decisions should be made by a team of people who form an expert panel similar to that described above for a test. The judgments this panel is making relate to the CEFR level or levels that reflect specific elements of the curriculum and can be collected and analysed in the same way as has been suggested above; in other words, a number of rounds of judgment and discussion are needed to come to the most appropriate decision. This panel should include individuals who are internal to the curriculum team (and have expertise in that document and in the rationale behind the decisions that went into its creation) as well as external experts (who are very familiar with the CEFR).

In the case of teaching and learning materials, tasks, activities, etc. may have some similarities to test items but are not usually intended to produce a score. Completion of the textbook is not dependent on achieving a numerical score derived from performance on a specified number of items and tasks. Additionally, teaching and learning materials may be more broadly focused by design in order to encourage flexible and creative use. Nonetheless, it may be useful to adapt some of the approaches in Section 5.3.1. Setting a test-level (or resource-level) cutscore may not be the aim, indeed, it may not be possible or desirable. But it is worth obtaining empirical evidence that the items and tasks used as learning or progress-checking activities are in fact targeting the CEFR level they claim to target. The kind of item-level judgments obtained by standard setting methods can clearly contribute important evidence in support of the claim that a textbook or other resource is aligned to the CEFR. The adaptation of standard setting methods to contexts of classroom testing, discussed in Section 5.3.1, may also be relevant for teaching and learning materials.

5.5 Guidelines for reporting on the standard setting stage

As noted in previous chapters, it is unlikely that the same report will be equally relevant or useful for all stakeholders. Important questions we need to ask are:

- Who are we preparing the reports for?
- What is the purpose of the reports?
- What level of expertise and background knowledge can we assume in the intended users of the reports?

As well as clearly communicating your results, reports provide an opportunity for you to present validity evidence to support your recommendations. Thus **validation**, one of the five phases of alignment summarized in Chapter 1, also occurs in microcosm within the activity of standard setting. Validity evidence for standard setting is sometimes split into three categories.

1. Procedural: clear documentation of the method used and how it was implemented; feedback from judges to show that they thought the training provided was adequate etc.
2. Internal: technical evidence on the data analysed, e.g. the degree of agreement between judges, etc.
3. External: evidence that the results are consistent with other sources of evidence, e.g. evidence from other standard setting events, the number of students likely to be classified into CEFR levels with the cutoffs, etc.

It should be clear that validation, and reporting generally, cannot wait until after the event. We must decide in advance what kind of evidence we need to report and ensure that we have built the collection of this evidence into our plans and schedules. For example, to elicit procedural validity evidence, we must design a questionnaire to collect judges' feedback and we must ensure that we allow time for them to answer the questionnaire. After the event, we can rarely collect data we have missed or re-collect data which might not have been properly saved.

As noted in Chapter 4, full documentation of all aspects of our standard setting, from planning to results, is also an essential part of institutional memory. As well as providing the foundations for any future reporting, during the process of standard setting practical issues often require us to make adjustments to the original plans. If we don't document what these changes are, and the reason for them, it will be difficult to clarify at a later date what happened during the event, and to justify that to external stakeholders.

For an excellent example of a comprehensive study from all perspectives we recommend resource 15. Resources 11 and 17 offer a valuable overview of alignment studies undertaken following the publication of the pilot version of the *Manual for Relating Examinations to the CEFR* in 2003.

5.6 Resources

Many of the resources highlighted in earlier chapters will, of course, be relevant for carrying out standard setting, for example, links to the CEFR searchable descriptor bank and the 2001 CEFR and 2020 Companion Volume online. This section lists some resources that will help you to investigate standard setting methodology further. It does not aim to be comprehensive, but it provides a useful starting point and gives further information about the specific examples we have used in this chapter. As noted earlier, whether you conduct standard setting yourself or are responsible for commissioning experts to do so, some degree of familiarity with the literature is necessary if you are to make informed choices.

1 Breton, G., Lepage, S. & North, B. (2008). Cross-language benchmarking seminar to calibrate examples of spoken production in English, French, German, Italian and Spanish with regard to the six levels of the Common European Framework of Reference for Languages (CEFR). CIEP, Sèvres, 23–25 June 2008. Strasbourg: CIEP/Council of Europe,

2 Brunfaut, T., & Harding, L. (2014). Linking the GEPT listening test to the Common European Framework of Reference. LTTC-GEPT Research Reports RG-05. Taipei: Language Training and Testing Centre.

3 Bunch, M.B. (2020). Digital Module 14: Planning and Conducting Standard Setting. ITEMS Module, NCME.

4 Cizek, G. (ed.) (2012). *Setting performance standards.* New York: Routledge.

5 Cizek, G., & Bunch, M. (2007). *Standard Setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks: Sage Publications.

6 Council of Europe. (2009). *Relating language examinations to the Common European Framework of References for Languages: Learning teaching, assessment*. Strasbourg: Language Policy Division.

7 Dunlea, J., & Figueras, N. (2012). Replicating results from a CEFR test comparison project across continents. In D. Tsagari and I. Csepes (eds.), *Collaboration in language testing and assessment* (pp. 31-45). New York: Peter Lang.

8 Dunlea, J., Spiby, R., Malone, M., & Folny, V. (2021). *An innovative approach to evaluating claims of alignment to the CEFR for tests of different languages*. Paper presented at the 17th EALTA Online Conference.

9 Dunlea, J., Motteram, J., Spiby, R., Jiang, R., Spoettl S., Zehentner, M., Kremmel B., & Holzknecht, F. (In press). Aligning Tests of Language Proficiency and Numeracy to the Workplace Literacy and Numeracy Framework. *British Council Validation Series*, VS/2021/001. London: British Council.

10 Dunlea, J, Spiby, R., Wu, S., Zhang, J., & Cheng, M. (2019). China's Standards of English Language Ability (CSE): Linking UK Exams to the CSE. *British Council Validation Series*, VS/2018/001. London: British Council.

11 Figueras, N., & Noijons, J. (eds.) (2009). Linking to the CEFR levels: Research perspectives. Arnhem: Cito-EALTA.

12 Harsch, C., & Kanistra, V.P. (2020). Using an Innovative Standard-setting Approach to Align Integrated and Independent Writing Tasks to the CEFR. *Language Assessment Quarterly*, 17:3, 262–281, DOI: 10.1080/15434303.2020.1754828

13 Kaftandjieva, F. (2004). *Standard setting*. In Council of Europe, Reference supplement to the pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF). Strasbourg: Language Policy Division.

14 Kaftandjieva, F. (2010). Methods for setting cut-off scores in criterion-referenced achievement tests: A comparative analysis of six recent methods with an application to tests of reading. Arnhem: CITO and EALTA.

15 Kantarcıoğlu, E. (2012). Relating an Institutional Proficiency Examination to the CEFR: a case study. Unpublished PhD Thesis, The University of Roehampton, London.

16a Kollias, C. (2017). Virtual standard setting in language testing: Exploring the feasibility and appropriateness of setting cut scores in synchronous audio and audio-visual environments. Unpublished PhD thesis. Lancaster University, Lancaster.

16b Kollias, C. (2021). *Variations in setting cut scores: How comparable are cut scores across media, methods, time, and instrument length?* Paper presented at the 17th EALTA Online Conference.

17 Martyniuk, W. (ed.) (2010). Aligning tests with the CEFR: Reflections on using the Council of Europe's draft manual. *Studies in Language Testing* 33. Cambridge: Cambridge University Press.

18 Verhelst, N., Fugueras, N., Prokhorova, E., Takala, S. & Timofeeva, T. 2019. Standard setting for writing and speaking: The Saint Petersburg experience. In Ari Huhta, Gudrun Erickson & Neus Figueras (eds.), *Developments in Language Education : A Memorial Volume in Honour of Sauli Takala* (pp. 278–301). University of Jyväskylä Centre for Applied Language Studies & EALTA. https://jyx.jyu.fi/bitstream/handle/123456789/65608/978-951-39-7748-1.pdf?sequence=1&isAllowed=y

Notes for your own implementation …

Standard setting requires technical expertise and is resource-intensive. This chapter does NOT provide detailed support for planning and delivering a standard setting project, but it IS a useful launch pad for finding out what you need to know and where to start looking for that information.

Some important points to keep in mind:

- You will need to engage with technical and academic literature.
- As noted in all the chapters, you should engage critically with that literature and adapt methods to your resources, project needs, and your aims regarding alignment to the CEFR.
- You should document all steps in the process: the rationale for selecting a method, the adaptions you make to fit your context, the process of preparation and data collection itself and the analysis, results, and validation evidence.
- Take time to carefully consider the resources you have, and the resources you will need, in order to carry out standard setting appropriately for your context.

Finally:

- Planning the logistics and practical aspects of standard setting events is as important as selecting the right method and having the right technical expertise for the project.
- While standard setting is an empirical process requiring statistical analyses of quantitative data, at its heart it is also fundamentally about people:
    - the judges and the group dynamics during standard setting events
    - you and your team delivering the training and facilitating the judgments
    - the test-takers and stakeholders who will be impacted by the decisions made based on the evidence you gather.

CHAPTER 6: VALIDATING YOUR CLAIMS

Chapter 6 focuses on how to go about validating the claims made following the whole process and presenting the validation report to stakeholders.

This chapter is organized as follows:

By the end of this chapter, you should have a clear idea of:

- **what we mean by validation in this context**
- **the importance of building an evidence base for the claims you make about alignment to the CEFR**
- **the structure and content of the validation report.**

6.1 What we mean by validation

The commonly accepted understanding of validation in educational measurement is that validation is established by building a clearly structured argument to support the drawing of specific inferences within a specified context of use – thus the assessment-use argument approach of Kane (1992, 2006, 2010)[7].

Since this handbook is not meant to offer the reader a detailed overview on how an assessment can be validated, we will consider validation from a somewhat different perspective. This manual has set out a series of steps through which the developers of a particular target resource (curriculum, materials, teacher education programme or test) can reach a point where they are in a position to make a substantiated claim that this activity is linked in some specific way to the CEFR. Therefore, when we refer to validation, we are referring to the act of collecting and presenting a body of evidence to support our claim of a link between our target activity and the CEFR.

Before moving on to outline what the evidence above might look like and how it might be presented, let's first consider two very important limitations of the validation of our claims.

6.2 Key limitations of validated alignment claims

We cannot meaningfully argue that a target is valid in its own right. As you've seen in Section 1.4, a set of classroom materials devised for one context (with clear links to a given curriculum) cannot be expected to automatically work in a different system. This means that the target activity can only be valid when used for a particular purpose in a particular context (normally the purpose and context for which it has been developed). Any change to either purpose or context will mean that new evidence will be needed to support a validation claim. So, validation is directly linked to the use of the target activity within a given context.

In addition, it is unwise to think that a target resource can be either valid or invalid. Since it is difficult to envisage a situation where the evidence gathered is so overwhelming (as in "a smoking gun") either way, it is better to think of validity as a matter of building an evidential basis through which you will attempt to convince the user/reader that the evidence in support of your claim of a link to the CEFR is sound, accurate and trustworthy.

---

[7] Kane, M. T. (1992). An Argument-Based Approach to Validity. *Psychological Bulletin, 112*, 527.
Kane, M. T. (2006). Validation. In R. Brennan (ed.), *Educational measurement, 4th ed*. (pp. 17–64), Westport, CT: American Council on Education and Praeger.
Kane, M. T. (2010). *Validity, Fairness, and Testing*. Paper presentation at the invitational conference: Educational assessment, accountability, and equity: Conversations on validity around the World. New York: Columbia University. https://www.ets.org/c/18486/pdf/19633_ets_tc_validityconference_Kane%202012_03_28.pdf

6.3 Building a validation evidence base

By carefully documenting the various stages of the process as outlined in this manual, the developer/validator can put together the kind of evidence referred to above. Figure 6.1 makes it clear that the act of validating the process essentially means bringing together the evidence gathered during the various stages of the process to form a coherent and convincing argument in support of your claim.

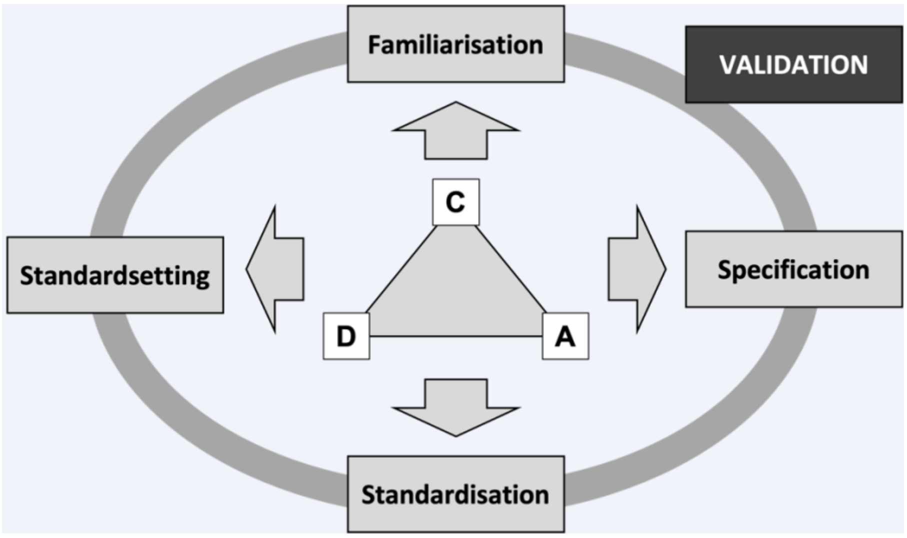

Figure 6.1: Linking claim validation overview
[C – Curriculum; D – Delivery; A – Assessment]

We do not see this as an all-or-nothing activity. As can be seen in Figure 6.2, depending on the quality of the evidence gathered at the various stages, different degrees of strength will be associated with the claim of a link to the CEFR. Note that even here, there is not a dichotomy in play – even the "strongest" claim does not mean that the argumentation cannot be disputed. Nor does it mean that the link is "proven". Obviously, more and stronger evidence will lead to a stronger argument that is ultimately more convincing.

In this figure, the darker shaded arrows represent the ideal progress through the alignment process. The lighter arrows indicate a less than favourable outcome, where the evidence of an alignment claim being justified is either weak or non-existent. For example, no claim can be made after familiarization; similarly, if the evidence is poor or missing then no claim can be made following specification (following the lighter arrows). If specification is successful and we progress to standardization (indicated by the dark arrow) only to find that the evidence here is poor or missing, we are only expected to make the weakest of claims of alignment and cease the process. If the evidence is strong enough, we progress to standard setting.

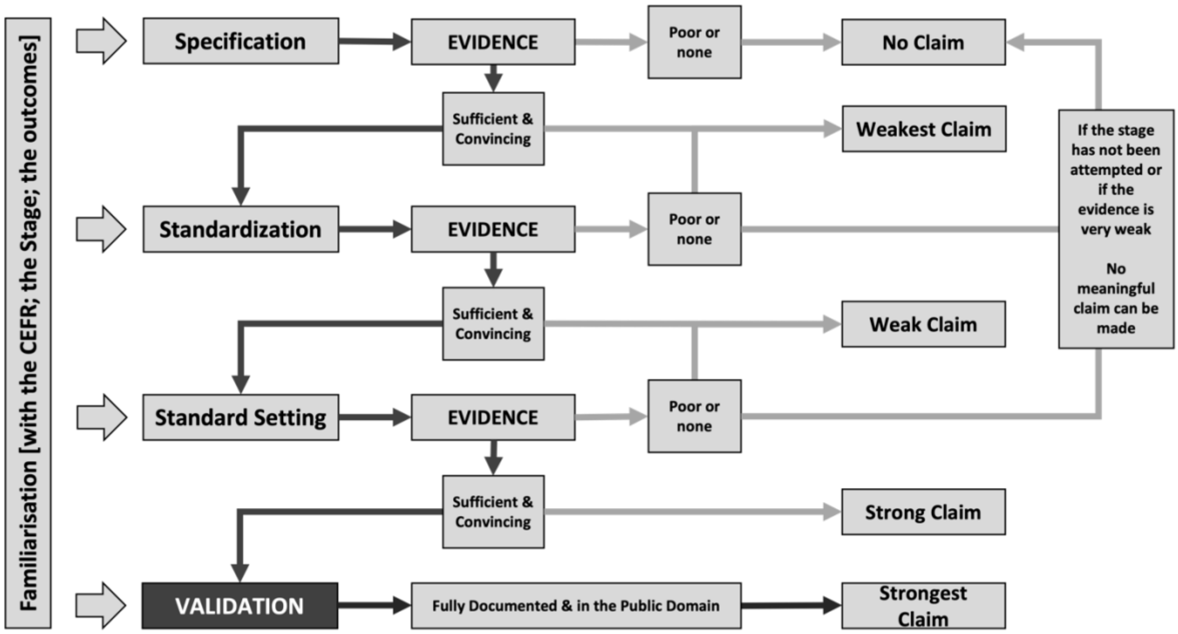

Figure 6.2: Towards a strong claim

6.4 A note on the structure of the report

As we have pointed out, we do not envision that the validation of the alignment process should fully reflect the quite rigidly structured arguments that are currently holding sway in the area of educational assessment. While a degree of formality is certainly expected, it should be remembered at all times that the validation argument in the educational assessment world tends to be focused on a single narrowly defined set of stakeholders – typically expert academic theorists. Our vision of the validation argument is quite different in that we recognize that there are likely to be a whole range of stakeholders to be convinced – after all, the only reason to present an argument is to convince a reader of the veracity of a specific claim. Different stakeholder groups will have very different expectations as to the form taken by this argument. For example, an academic may have an interest in a particular theoretical approach to argumentation and expect a complex structure such as the adaptation of the Toulmin[8] structure suggested by Kane (2010) for example. On the other hand, teachers or materials writers are unlikely to have the luxury of time to engage with a long and complex argument, so might expect a more practical and targeted overview. A parent or a learner, however, might expect that the evidence is presented in a more accessible format, for example through an online blog or even social media.

[8] The basic Toulmin argument is comprised of the grounds, the claim and the warrant. It is possible to add another three, if and when required: backing, qualifier and rebuttal. These elements can be organized to create a structured argument.

6.5 Guidelines for reporting on the validation stage

With regard to the construction of the final report, we feel that a relatively straightforward approach is likely to be appropriate. This essentially means building up a body of evidence by completing the forms provided in this manual, reflecting on the process and on the quality and quantity of the evidence at each stage and reaching an evidence-based conclusion with regard to the strength of the claim you feel the evidence allows.

The conclusion to each stage will also allow you to present your argument for progression to the following stage. We would expect that where the evidence is considered appropriate for a reasonable claim, then a move to the next stage will be automatic. In order to deflect counter claims (e.g., from external bodies or competitors), we would advise you to seek independent external review of the documentation (forms, reflection and claims).

Once we have completed all stages of the process as outlined in this handbook, it is now time to put the validation report together. As noted in the previous section, we are likely to have a broad range of stakeholders so you may need to think about how to communicate your evidence to them in different ways and with different amounts and types of evidence. However, it's critical that the underlying message is both consistent and accurate. For this reason, the preparation of the validation report would benefit from the support of communication experts. Table 6.1 suggests a broad structure for the process of developing the report.

Table 6.1: Suggested approach to writing the validation report

Stage	Activity	Gloss
1	**Purpose**	Outline the main purpose of the validation argument. Identify the target audience and the points to be made. This is akin to writing a theory of change for your alignment project.
2	**Structure**	Agree on the final structure of the various arguments to be put together – remembering that different stakeholders will require different approaches.
3	**Evidence**	Describe the evidence collection procedures adopted. Note: the main focus in this chapter has been on *internal validation* evidence, meaning that it is target resource-centred. In the case of tests, it is also advisable to collect *external* evidence – for example, teacher estimates to compare with categorizations based on the linking project.
4	**Write**	Construct the argument based on all available evidence – this will reflect the chapters in this handbook, employing the tables presented here together with the suggestions for those chapters where tables are not available.
5	**Claim**	At the end of each section (i.e., corresponding to the chapters in this handbook), summarize the evidence to support the level and strength of the claim you are making. Then provide the rationale for moving to the next stage. In the final section, you can bring all your claims together to make your final claim relating to level and strength.
6	**Communicate**	Make the final argument/report available to the target stakeholder groups (note: there may be a number of different reports for different groups)

Table 6.2 offers a quick overview of the specific details to be addressed in the report. This table is meant only to suggest what evidence goes into the argument and does not attempt to suggest how the argument itself is constructed.

Table 6.2: Overview of the stages and potential claims

<table>
<tr><th>Stage</th><th colspan="3">Validation evidence expected</th></tr>
<tr><td>Familiarization</td><td colspan="3">Identify participants and document selection process
Complete forms (Ch. 2)
Reflective evaluation and support for claim and/or move to next phase</td></tr>
<tr><td rowspan="3">Specification</td><td colspan="3">Identify participants and document selection process (if changed)</td></tr>
<tr><td>Curriculum: Complete forms (Ch. 3) amended as appropriate (with rationale)</td><td>Textbook/materials: Complete forms (Ch. 3) amended as appropriate (with rationale)</td><td>Test/Assessment: Complete forms (Ch. 3)</td></tr>
<tr><td colspan="3">Reflective evaluation and support/rejection of claim and/or move to next phase</td></tr>
<tr><td>Standardization</td><td colspan="3">Identify participants and document selection process
Document and rationalize selection of materials used (complete forms in Ch. 4)
Report on all activities undertaken
Reflective evaluation and support/rejection of claim and/or move to next phase</td></tr>
<tr><td>Standard setting</td><td colspan="3">Select the method to be used
Identify the resources needed (physical, financial and human)
Identify the relevant experts for the panel (document recruitment process)
Select and document test materials to be used
Identify and document the test data to be used in the event
Prepare all communications (specifically for the expert panel but also set out a dissemination strategy)
Identify and prepare all training materials (links to Familiarization)
Prepare all data collection forms/materials
Organize and prepare the physical space to be used
Report on all activities undertaken – including results of any statistical analyses
Reflective evaluation and support/rejection of claim and/or move to next phase</td></tr>
<tr><td>Validation</td><td colspan="3">Define the target stakeholder group
Outline the main purpose of the validation argument
Describe the structure of the argument
Describe the evidence collection procedure adapted (note: the focus in this chapter has been on internal, that is resource-centred, evidence. In the case of tests, it is also advisable to collect external evidence – for example, teacher estimates to compare with categorizations based on the linking project)
Construct the argument based on all available evidence
State the level and degree of final claim (this can be done at each stage, culminating in the final claim at the end of the process)
Make the final argument/report available to the target stakeholder group (note: there may be a number of different reports for different groups)</td></tr>
</table>

Many of the processes described in this handbook were either based on the earlier Council of Europe *Manual for Relating Examinations to the CEFR* (2009) or reflect practice in educational assessment. We feel that the forms contained in Chapters 2–4 are likely to be held in common across the three broad areas we have described (curriculum, delivery and assessment). The form presented for Chapter 5 is different to the others in that it is not designed to form a part of the validation report. Instead, it is designed to be a checklist for the planning of the standard setting stage of the process. Of course, the highlighted headings can be used to inform the contents of the report of the standard setting stage. We expect that the report on this stage will differ significantly depending on the resource being aligned. While the chapter focuses mostly on aligning tests to the CEFR, many of the steps will be repeated for other resources though the details will change (see Table 6.3. for an overview).

The selection of the method to use differs in that the absence of data related to curricula or materials means that the only option open is to directly compare the actual contents of the resource with relevant CEFR descriptor scales. This means that these standard setting events will be different to those for tests in that there will not be any measurement/score data to compare judgments with. However, the making of judgments and the related analysis of these judgments, together with the discussions that follow the rounds of judgment-making, will be quite similar to the approach taken for tests.

Table 6.3: Overview of the steps to be taken for the different resource types

	Assessment	Curriculum	Materials/Textbook
Selecting a method	Choice (test or examinee centred)	Resource–centred only	Resource–centred only
Identifying the resources needed	Physical, financial and human	Physical, financial and human	Physical, financial and human
Identifying relevant experts	Number required Recruitment	Number required Recruitment	Number required Recruitment
Ensuring access to test materials	Identify and prepare materials	Identify and prepare materials	Identify and prepare materials
Ensuring access to data	Identify and prepare test data	N/A	N/A
Preparing communications	Invitation, consent, joining instructions	Invitation, consent, joining instructions	Invitation, consent, joining instructions
CEFR training materials	Identification or creation and preparation	Identification or creation and preparation	Identification or creation and preparation
Data collection materials	Judgment and feedback forms for expert panel Data analysis plan	Judgment and feedback forms for expert panel Data analysis plan	Judgment and feedback forms for expert panel Data analysis plan
Preparing the physical space	Identify, book and prepare venue	Identify, book and prepare venue	Identify, book and prepare venue

Finally, the original *Manual for Relating Examinations to the CEFR* (Council of Europe, 2009) contains significant discussion of the concepts of *internal* and *external* validation evidence. As stated in Table 6.1, the primary focus of the evidence gathering activities described in this handbook is *internal*. In other words, we are looking at evidence to support the claims made that is based on the documentation of the process, and associated reflection. The independent review at each stage offers a degree of external evidence, and strengthens the claims made at these stages. When it comes to the standard setting stage, the types of external evidence will vary depending on the element in question, so:

- curriculum – review of process, evidence and claim by an external group, comprised of at least two independent experts
- materials – similar approach
- test – similar approach, but for additional support, we would advise the collection of additional data-driven evidence. This might consist of score data from a similar assessment to corroborate the findings of the expert panel, or the use of teacher judgments of their own learners' CEFR level compared with the CEFR levels based on performance on the target assessment.

The activities discussed in this handbook all relate to the estimation of how well the target resource aligns to the CEFR in terms of its design and development. In a panel discussion at the Chinese Association for Language Testing and Assessment (CALTA) conference in October 2021, we were reminded by Fan Jia from Southwest Jiaotong University that this does not necessarily provide convincing evidence that the resource is actually operating at the expected level. This is particularly the case with estimates of the alignment between a curriculum or textbook and the CEFR where it is critical to consider exactly how the resource is being interpreted and used in the real world (the standardized nature of tests means that they are most likely to be consistently administered, avoiding such interpretability issues). Clearly, this research cannot be undertaken during the development phase of a curriculum or textbook, but it should form the basis of a systematic monitoring post launch.

In addition, it is important to consider that most school-bound textbooks will be expected to support individuals in their learning for a relatively limited stretch of time. In recognition of the fact that it takes a long time to develop a sensitivity for cases and then a certain degree of mastery of case use and case endings, the textbook publisher of an A1 textbook for German, for example, should make it clear that this particular textbook only covers a stage in the longer-term development of these specific aspects of language proficiency.

6.6 A note on aligning tests which use AI-supported scoring systems

The increasing use of artificial intelligence (AI) auto-scoring systems in tests of speaking and writing has yet to be adequately dealt with in terms of validation theory and practice (O'Sullivan, Breakspear & Bayliss, in press). Since claims of an alignment to the CEFR (or any other language framework or standard) are essentially a validation issue, we feel that it should be addressed in this handbook. The proposal by O'Sullivan, Breakspear and Bayliss is that test developers who rely on AI models for scoring whether through a machine only or a machine-supported system should provide evidence on both the original model development and on the specific validation of the model for use in their test. When we are dealing with a "black box" system, where the scoring algorithm is not transparent and is essentially unknown to either the test developer or test user, there is a real issue for any alignment claim. This can be addressed through the publication of a **Datasheet** (Gebru et al., 2021)[9] a **Model Card** (Mitchell et al., 2019)[10] and a **mixed methods** (quantitative and qualitative) study that specifically targets the alignment claim.

The Datasheet represents a specification for the basic or underlying AI scoring model and outlines how it was developed while the Model Card is a short, accessible (as in relatively non-technical) validation report indicating how the scoring model was tested for use in a specific context (e.g. where potential sources of bias within that context are identified and the model is trained using an appropriate dataset to eliminate or at least limit potential bias).

It is no longer acceptable to simply use an AI-driven or supported scoring system without providing evidence that it is demonstrably appropriate for use with regard to specific decisions in specific contexts. Neither is it acceptable to make a claim of alignment to a language framework or standard without combining such evidence with an appropriate alignment study such as suggested in this handbook.

---

[9] Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daum III, H. & Crawford, K. (2021). Datasheets for Datasets. *Communications of the ACM (CACM),* Volume 64, Issue 12: 86–92. https://doi.org/10.1145/3458723

[10] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D. & Gebru, T. (2019). *Model Cards for Model Reporting*. FAT* '19, January 29–31, 2019, Atlanta, GA, USA. Retrieved from https://arxiv.org/pdf/1810.03993.pdf (3 August 2021)

Notes for your own implementation …

The whole purpose of undertaking a linking project is to gather together the evidence to support your claims of a link between your curriculum, materials, assessment activities or tests and the CEFR. The evidence consists of various strands linked explicitly to the concepts that lie behind the chapters of this manual. It is not expected that you will follow all of the suggested processes outlined here. You may decide on a different approach, though you will need to consider offering a convincing rationale for doing so in your documentation. When brought together, the various pieces of evidence will form the basis of your validity report.

Some important things to keep in mind are:

- It is important to recognize from the start that the linking project cannot provide definite proof that your work is precisely linked to the CEFR. Instead, it should be seen as the process of gathering together enough evidence to build a convincing argument that the claim you are making is strong.
- Consider who your intended readership is before commencing the writing of the report. It is likely that you will have a number of audiences in mind so build a communication plan before writing the final report so that different (though consistent) versions can be considered from the start.
- Where feasible, include independent external evidence to strengthen your claims. For example, instead of gathering judgments only from colleagues in your organisation try to include a number of externals – this applies to standard setting for tests as well as judgments for materials and curriculum elements.
- For the evidence in support of a claim to be considered fully “sufficient and convincing” (see Figure 6.2) some level of independent external input is expected.
- Alignment is between a specific resource (test, material or curriculum) for use in a specific context (e.g., secondary schools in Malaysia). Where significant changes are made to the resource or to the context (e.g., secondary schools in Spain), then it may be that the alignment claim is no longer supported. A full or partial re-alignment may be required.
- Include a reflective element with both internal and external perspectives as this can help you develop a sensible, yet detailed, argument that can then be adapted for use in different contexts, for example in targeting different stakeholder groups as indicated in the previous point.

APPENDIX:
FORMS TO SUPPORT THE ALIGNMENT PROCESS AND ALIGNMENT CLAIMS

Forms for Chapter 2: Familiarization

Form 2.1: Generic familiarization

This Summary form shows that you have completed the activities proposed and your responses can be used for future reference, for validation and in reporting.	
Name of user	
Date	
Name of institution	
Title of resource	
Type of resource	☐ Curriculum ☐ Course material ☐ Textbook ☐ Exam ☐ Other: ________________________ (Please specify)

	ACTIVITIES	Responses
1	**Chapters 1 & 2** - What does alignment to the CEFR mean? - What can be the implications for me? And for my context?	
2	**Section 2.4** - What are the competences, activities and strategies most relevant in my context? Why? - Are they all equally relevant? If not, how can I prioritize them?	
3	**CEFR CV, Appendices 1 & 2** - What level(s) will I mainly be working with? - Which are the main aspects in that level and in the adjacent ones? - Which aspects are most relevant in my context for each mode of communication?	
4	**CEFR CV, Figure 11.** Reception activities & strategies - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	CEFR CV, Figure 12. Production activities & strategies - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	

	CEFR CV, Figure 13. Interaction activities & strategies - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	CEFR CV, Figure 14. Mediation activities & strategies - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	CEFR CV, Figure 15. Plurilingual and pluricultural competence - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	CEFR CV, Figure 16. Communicative language competences - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	CEFR CV, Figure 17. Signing competences - Which are the most relevant scales in my context? Why? - Are they all equally relevant?	
	What is the rationale for my decisions above? (this is key as it will support later claims)	
	On a scale of 1 to 10, how satisfied am I with my generic familiarization with the CEFR? **1 2 3 4 5 6 7 8 9 10** **What extra familiarization do I need?** …… ……	

Form 2.2: Specific familiarization, Session 1

This Summary form shows that you have completed the activities proposed and your responses can be used for future reference, for validation and in reporting.

Name of user		
Date		
Name of institution		
Title of resource		
Type of resource	Curriculum	Course material
	Textbook	Exam
	Other: ________________________ (Please specify)	

ACTIVITIES	Responses [includes rationale]
PART 1	
1. Introduction and Review What features in the CEFR levels are most relevant in my context?	
2. Step 1: CEFR CV, Appendix 2. Self-assessment grid: Reception/ Production/ Interaction What key words best identify the following (consider action(s), type of information, conditions & limitations)? • oral comprehension • reading comprehension • oral production • written production • oral interaction • written & online interaction	
3. Step 2: CEFR CV, Appendix 2. Self-assessment grid: Mediation What key words best identify the following (consider action(s), type of information, conditions & limitations)? • mediating a text • mediating concepts • mediating communication	

PART 2	
4. Activities: XXXXXXXXXX scales **Scale name X** Why is this scale relevant in my context? What key words best identify my purpose(s)? **Scale name X** Why is this scale relevant in my context? What key words best identify my purpose(s)? **Scale name X** Why is this scale relevant in my context? What key words best identify my purpose(s)?	

Form 2.3: Overall summary of familiarization process

<table>
<tr><td colspan="2">This Summary form shows that you have completed the activities proposed and your responses can be used for future reference, for validation and in reporting.</td></tr>
<tr><td>Name of user</td><td></td></tr>
<tr><td>Date</td><td></td></tr>
<tr><td>Name of institution</td><td></td></tr>
<tr><td>Title of resource</td><td></td></tr>
<tr><td>Type of resource</td><td>Curriculum
Course material
Textbook
Exam
Other: _________________________ (Please specify)</td></tr>
</table>

ACTIVITIES	Responses (includes rationale)
1. Introduction and Review - What scales and descriptors discussed in Specific Familiarization 1 are most relevant in my context? - Why? - What aspects in the scales to be discussed in this session are most useful in my context?	
2. Activities: XXXXXXXXXX scales **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? **../..**	

3. Activities: XXXXXXXXXX scales **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? **Scale name X** - Why is this scale relevant in my context? - What key words best identify my purpose(s)? ../..	
On a scale of 1 to 10, how satisfied are you with Specific CEFR familiarization Session ?? **1 2 3 4 5 6 7 8 9 10**	

Forms for Chapter 3: Specification

Form 3.1: Description of the alignment and the resource

GENERAL INFORMATION		
Name of user		
Date		
Name of institution		
1. Title of resource		
2. Type of resource	Curriculum Textbook Other: ________________________ (Please specify)	Course material Exam
3. Nature of alignment	New production	Retrofitting
4. Target language		
5. Name of the resource material		
6. Brief description of the resource including its purpose and objectives		
7. Scope of the resource	International National	Regional Institutional
8. Target population	Young Learners Lower Secondary Adults	Uni/College students Upper Secondary
	Other: _____________________ (Please specify)	
9. Principal domain(s)	Public Occupational	Personal Educational
10. Intended CEFR Level(s)	A1 B1 C1	A2 B2 C2
11. Salient features Write down the salient features of the resource in relation to the intended overall CEFR level(s).		

Form 3.2: Language activities, strategies, and competences

Name of user		
Date		
Name of institution		
Title of resource		
Communicative language activities and strategies targeted		Oral Comprehension
		Audio-visual Comprehension
		Reading Comprehension
		Reception Strategies
		Oral Production
		Written Production
		Production Strategies
		Oral Interaction
		Written Interaction
		Online Interaction
		Interaction Strategies
		Mediating Texts
		Mediating Concepts
		Mediation Communication
		Mediation Strategies
Plurilingual and pluricultural competences targeted		Plurilingual Comprehension
		Building on Plurilingual Repertoire
		Building on Pluricultural Repertoire
Communicative language competences targeted		Linguistic Competence
		Sociolinguistic Competence
		Pragmatic Competence
Signing competences targeted		Linguistic Competence
		Sociolinguistic Competence
		Pragmatic Competence

Form 3.3: Graphic profiling of alignment

Name of user	
Date	
Name of institution	
Title of resource	

C2										
C1										
B2										
B1										
A2										
A1										
	Oral	Reading	Oral	Written	Oral	Written	Online	Text	Concept	Communication
	Reception		Production		Interaction			Mediation		

Form 3.4: Illustrative descriptors of language activities and strategies

The following form should be completed for each type of language activity, strategy, and competences targeted in the resource material.

Name of user	
Date	
Name of institution	
Title of resource	

	Short description and/or reference
Name of the language activity / strategy / competence	
Which communication themes are learners expected to be able to handle? *The lists in CEFR 4.2 can be of help as a reference (2001: p. 52)*	
Which communicative activities and strategies are learners expected to be able to handle? *Chapter 3 of the CEFR Companion Volume can be of help as a reference (2020: pp. 47–122)*	
Which text types are the learners expected to be able to handle? *The lists in CEFR 4.6.2 and 4.6.3 can be of help as a reference (2001: pp. 94–95)*	
After reading the scale for the related language activity, indicate and justify at which level(s) of the scale learners are expected to perform.	

Form 3.5: Detailed mapping

Name of user	
Date	
Name of institution	
Title of resource	

Component (activity, item, objective)	Communicative Activity / Strategy / Competence	Domain(s)	Theme(s)	Text Type(s)	CEFR Scale	CEFR Level	CEFR Descriptor

Forms for Chapter 4: Standardization

Form 4.1: Standardization summary form – participant

STANDARDIZATION SUMMARY FORM: Overview of the Standardization Process

This summary form outlines the standardization activities completed and can be used as a guide for reporting. Different forms will need to be completed for the different areas targeted.		
	Name of user	
	Date	
	Name of institution	
	Title of resource	
	Area	☐ Reception (Listening) ☐ Reception (Reading) ☐ Written Production ☐ Spoken Production ☐ Other: ____________________ (Please specify)
ACTIVITIES		**RESPONSES/DETAILS**
1	Familiarization	Review of prior activities (Ch.2) ☐ Additional activities (briefly list) ………… …………
2	Specification	Review of prior activities (Ch.3) ☐ Statement of context (description) …………………………………………………………………… ……………………………………………………………………
3	Standardized description of expected learning achievements	Description:
4a	Analysis of illustrative examples	
	1. Are they relevant to the proficiency level(s) I am concerned with?	Yes ☐ No ☐ Comments:

	2. Are they relevant to the context and purpose of my resource?	Yes ☐ No ☐ Comments:
	3. Are they relevant to the mode(s) of communication, language activities and language strategies central to my resource?	Yes ☐ No ☐ Comments:
4b	Benchmarking	
	a). Existing resource Does my resource reflect accurately the CEFR level it targets?	Yes ☐ No ☐ Comments:
	b). New resource Does the component reflect accurately the CEFR level it targets?	Yes ☐ No ☐ Comments:
5	Resource completion	Describe process of writing/editing briefly:
	a) Existing resource	
	b) New resource	

On a scale of 1 to 10, how satisfied are you with this standardization session?

1 2 3 4 5 6 7 8 9 10

How could the effectiveness of standardization sessions be improved?

Form 4.2: Standardization summary form – coordinator

STANDARDIZATION SUMMARY FORM: Overview of the Standardization Process

This summary form outlines the standardization activities completed and can be used as a guide for reporting. Different forms (4.1.1, 4.1.2, 4.1.3, …) will need to be completed for the different areas targeted.		
	Name of user	
	Date	
	Name of institution	
	Title of resource	
	Approach taken	☐ Top Down (new resource) ☐ Bottom Up (existing resource)
	Area	☐ Reception (Listening) ☐ Reception (Reading) ☐ Written Production ☐ Spoken Production ☐ Other: ____________________ (Please specify)
ACTIVITIES		**RESPONSES/DETAILS**
1	Familiarization	Review of prior activities (Ch.2) ☐ Additional activities (briefly list) ………… …………
2	Specification	Review of prior activities (Ch.3) ☐ Statement of context (description) ………………………………………………………………… …………………………………………………………………
3	Standardized description of expected learning achievements	Description:
	Illustrative examples used (names and sources) a…………………………….. b…………………………….. c……………………………..	Justification/results of advance analysis a…………………………….. b…………………………….. c……………………………..
4a	Analysis of illustrative examples	Yes or No? Add comments.

	1. Are they relevant to the proficiency level(s) I am concerned with? 2. Are they relevant to the context and purpose of my resource? 3. Are they relevant to the mode(s) of communication, language activities and language strategies central to my resource?	1. 2. 3.
4b	Benchmarking a). Existing resource Does my resource reflect accurately the CEFR level it targets? b). New resource Does the component reflect accurately the CEFR level it targets?	Yes or No? Add comments. a) b)
5	Resource completion a) Existing resource b) New resource	Describe process of writing/editing briefly: a)…………………………………………………………………… b)…………………………………………………………………………

Trialling and consultation

Has your resource been trialled with learners to ensure that it works as intended?

Where?

Who with?

How?

Who has been consulted re. the final draft of your resource?

On a scale of 1 to 10, how satisfied are participants with this standardization session?

1 2 3 4 5 6 7 8 9 10

How could the effectiveness of standardization sessions be improved?

Form 4.3: Analysis of illustrative examples. Holistic form

Given the diversity of resource characteristics, this is just an example form, which needs to be customized.

Skill	**CEFR level assigned**	**Comments**
Sample/Task 1		
Sample/Task 2		
Sample/Task 3		
Sample/Task …		
Sample/Task N		
This is an example of a single rating sheet which requires the participants to give one global judgment about the level of each sample or task. This rating sheet can be used to rate either performances (written or spoken) or test items.		

Form for Chapter 5: Standard setting

<table>
<tr><td colspan="5">Below is a checklist of key questions to take into consideration in planning for standard setting.</td></tr>
<tr><td>Name of user</td><td colspan="4"></td></tr>
<tr><td>Date</td><td colspan="4"></td></tr>
<tr><td>Name of institution</td><td colspan="4"></td></tr>
<tr><td>Title of resource</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Type of resource</td><td colspan="2">Curriculum</td><td colspan="2">Course material</td></tr>
<tr><td colspan="2">Textbook</td><td colspan="2">Exam</td></tr>
<tr><td colspan="4">Other: ________________________ (Please specify)</td></tr>
<tr><td rowspan="2">Selecting a method</td><td></td><td>Test-centred</td><td>Specify</td><td></td></tr>
<tr><td></td><td>Examinee-centred</td><td>Specify</td><td></td></tr>
<tr><td>What is the rationale for the method(s) selected?</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Identifying the resources needed
(For each indicate if you already have access to these resources)</td><td>Physical</td><td colspan="3"></td></tr>
<tr><td>Financial</td><td colspan="3"></td></tr>
<tr><td>Human</td><td colspan="3"></td></tr>
<tr><td>Identifying relevant experts
How many panellists will I convene? Why this number?</td><td colspan="4"></td></tr>
<tr><td>Details on how experts are identified and contacted</td><td colspan="4"></td></tr>
<tr><td>What are the minimum requirements for panellists? (experience, qualifications etc)</td><td colspan="4"></td></tr>
</table>

Ensuring access to test materials What materials will I require?		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
Ensuring access to test data Will I have access to test-taker performances and / or empirical data		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
Preparing communications		Invitation letter ready	sent
		Consent form ready	sent
		Joining Instructions ready	sent
CEFR Training Materials (if the materials do not exist already, indicate how you will produce them)		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
Data collection materials What format will I produce judgment data collection forms in?		Spoken Interaction	Details:
		Spoken Production	Details:

		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
		Feedback Forms	Details:
Do I have feedback forms for panellists?		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
What kind of data analysis will be required, including software requirements		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details:
Preparing the physical space Where will the sessions take place?		Spoken Interaction	Details:
		Spoken Production	Details:
		Written Interaction	Details:
		Written Production	Details:
		Reception – Listening	Details:
		Reception – Reading	Details: