

The CEFR Companion Volume with New Descriptors: Uses and Implications for Language Testing and Assessment

Vith EALTA CEFR SIG

Trinity College Dublin, 27 January 2018

Report compiled by
David Little
with contributions from Stergiani Kostopoulou,
Kate Plenter-Vowles and Ülle Türk

Contents

Welcome – Lorna Carson, Director of the Centre for Asian Studies, Trinity College Dublin.....	1
Introduction – Neus Figueras, co-ordinator of the EALTA CEFR SIG.....	1
Session 1.....	3
Updates to the CEFR – John H. A. L. de Jong, Language Testing Services.....	3
CEFR Companion Volume: first thoughts – Anthony Green, University of Bedfordshire.....	5
The CEFR Companion Volume – Elif Kantarcioglu, Bilkent University.....	7
What the new “can do” descriptors can do for classroom assessment – Armin Berger, University of Vienna.....	8
Discussion.....	9
Session 2.....	11
The Mediation Scales – Barry O’Sullivan, British Council.....	11
Situating mediation within an argument for test use – Spiros Papageorgiou, Educational Testing Service.....	13
Situating mediation within the classroom – Dina Tsagari, OsloMet – Oslo Metropolitan University (formerly HiOA).....	15
Discussion.....	17
Session 3.....	19
The pluricultural and plurilingual scales – Vincent Folny, CIEP (Centre international d’études pédagogiques).....	19
Pluricultural and plurilingual scales: responses from the classroom and standardized testing – Jamie Dunlea, British Council / Gudrun Erickson, University of Gothenburg.....	21
Discussion.....	23
Session 4.....	24
Group discussion.....	24
Group 1: Updates to the CEFR – rapporteur: Ülle Türk.....	24
Group 2: Mediation – rapporteur: Stergiani Kostopoulou.....	25
Group 3: The plurilingual and pluricultural scales – rapporteur: Kate Plenter-Vowles.....	26
Session 5.....	29
Reports from the discussion groups.....	29
Group 1: Update of 2001 scales – rapporteur: Doris Froetscher.....	29
Group 2: Mediation – rapporteur: Bárbara Eizaga.....	29
Group 3: The plurilingual and pluricultural scales – rapporteur: Scott Stroud.....	29
Close of workshop.....	30
Appendix 1: Programme.....	31
Appendix 2: List of participants.....	32

Welcome – Lorna Carson, Director of the Centre for Asian Studies, Trinity College Dublin

Lorna Carson welcomed participants to Trinity College Dublin; as someone who has long worked with the CEFR, she was delighted to be able to host the VIth meeting of EALTA's CEFR SIG. She explained that the Long Room Hub, in which the meeting was being held, is designed to bring together academics working in the humanities and to foster interdisciplinary research.

Introduction – Neus Figueras, co-ordinator of the EALTA CEFR SIG

Welcome to the VIth EALTA CEFR SIG, which will discuss and critically evaluate the possible uses and implications of the recently published *CEFR Companion Volume* for language testing and assessment.

The CEFR SIG is a relatively recent development in EALTA; it was first convened in 2015, coordinated by the late Sauli Takala and myself. However, the CEFR and other Council of Europe initiatives have been central to EALTA's concerns since its creation. EALTA's interest in European initiatives is present in its mission statement: "to promote the understanding of theoretical principles of language testing and assessment, and the improvement and sharing of testing and assessment practices throughout Europe". It is also reflected in EALTA's participatory status in the Council of Europe since 2008.

At the Sèvres CEFR SIG in June 2017, a number of participants suggested that the next SIG should allow for discussion of the new developments around the CEFR. Following this recommendation from EALTA members, and in order to engage with the Companion Volume at the earliest possible opportunity in order to generate feedback from the field of language testing and assessment, this SIG has a different structure from past SIGs. Today's programme is organized around three main sections, with position presentations by testing and assessment researchers followed by prepared reactions by respondents coming from different areas in the field. Each section introduces discussion by evaluating the implications for language testing and assessment of three main areas in the Companion Volume:

- Updates to the CEFR (changes to descriptors, new descriptors, new scales for literature, phonological control, online communication ...)
- Mediation scales
- Plurilingual/pluricultural scales

The aim is to stimulate contributions from participants during discussion time.

Participants in this SIG include professionals from many different countries – Austria, France, Spain, UK, Italy, the Netherlands, Estonia, the USA, and of course Ireland. Two local participants have a long history of involvement with the CEFR, David Little and Joe Sheils. We are also happy to welcome Brian North, who coordinated the updating of the original CEFR scales and the development of the new scales, and is one of the authors of the Companion Volume.

We all come from different backgrounds and hold different perspectives, and I am sure you all have a lot to contribute, so please be generous and share your thoughts and opinions. The discussions will be included in a report of the meeting which will be put together for extensive dissemination. I am very grateful to David Little, who has volunteered to act as rapporteur, and to Ülle Türk, Kate Plenter-Vowles and Stergiani Kostopoulou, who have agreed to help by taking notes all through the day.

Many colleagues and institutions have made this SIG possible. First and foremost, I need to thank the main speakers and the respondents. They are all highly qualified professionals who have used and researched into the CEFR and who have a strong interest in studying and analysing the

Companion Volume and its implications for the field. I also want to thank EALTA and the British Council for their encouragement and financial support, and Trinity College Dublin for allowing us to use this fantastic venue free of charge. Most especially, I need to thank Lorna Carson for her expertise, generosity and extraordinary energy, and above all for making everything seem so simple and easy.

I hope you all find this SIG both engaging and interesting and look forward to a very productive day.

Session 1

Updates to the CEFR – John H. A. L. de Jong, Language Testing Services

The CEFR *Companion Volume* (CV) is the result of an enormous amount of work, and for that the authors deserve congratulation. There are nevertheless grounds for criticism. The text of the CV clearly still needs professional editing; the many typos mean that the text is sometimes cumbersome to read. More seriously, the CV offers ample information on the number of people involved in its development and on some aspects of the design of the data-gathering process, but it totally lacks information on the actual data generated by the scaling procedures. There is insufficient information to check the data, far less to replicate the study. We are not even given difficulty values for the new descriptors. In other words, we have no means to seriously evaluate the work. With this in mind, it might have been wise to create two publications, one for teachers (a companion to the 2001 CEFR) and the other for researchers.

When the CEFR was first published, in 2001, there were those who said it embodied a European approach and should not be used outside Europe. John Trim's reply to this was that if people outside Europe like it, why should they not use it? The Japanese have adapted the CEFR as CEFRJ and are using it in English language teaching. The Chinese on the other hand have developed their own Chinese Standards of English (CSE), but the work they have done encounters difficulties in scaling. This suggests that it's better to work with what has been done already and build on that.

Jan Hulstijn criticized the 2001 CEFR on three grounds: (i) its failure to distinguish consistently between levels of language proficiency (static aspect) and language development (dynamic aspect); (ii) its confounding of levels of language proficiency and intellectual abilities (though language is about content – PISA shows that many teenagers have an inadequate level of reading proficiency in their L1); and (iii) the potential mismatch between L2 learners' communicative and linguistic competences (though such a mismatch doesn't undermine the system). Glenn Fulcher criticized what he called the "reification" of the CEFR, but what's the problem? The levels make it possible to talk about learners in a way that wasn't possible before. Other criticisms of the 2001 version included: the fact that mediation was distinguished as one of four modes of communicative language use but not elaborated in terms of descriptors and illustrative scales; a lack of balance (more descriptors/scales for speaking than for other activities); and unbalanced levels (A1–B2 contained most of the descriptors).

The CV responds to these last two criticisms. Speaking is still dominant, but there are many more descriptors on reading and writing (Figure 1). The balance of the levels is also much improved (Figure 2), especially when taking into account that levels A2, B1 and B2 each represent two levels (Figure 3); though often there is "no

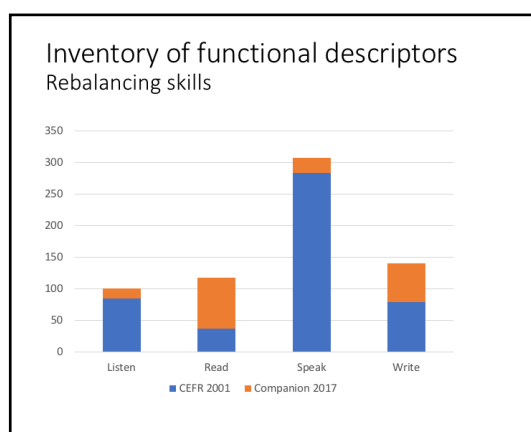


Figure 1

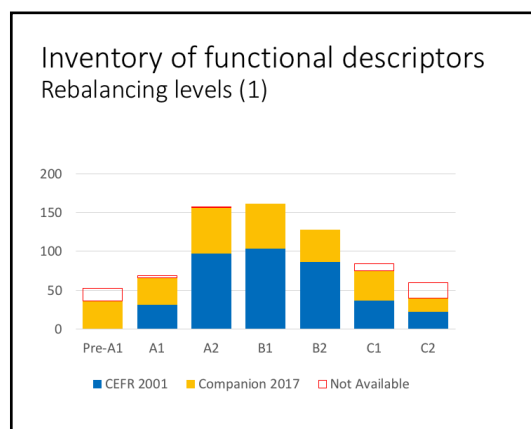


Figure 2

descriptor available” for pre-A1, A1 and the C levels (Figures 2 and 3). When descriptors are too long they are difficult for learners and teachers to use; on average the new descriptors contain five more words than the 2001 descriptors (Figure 4).

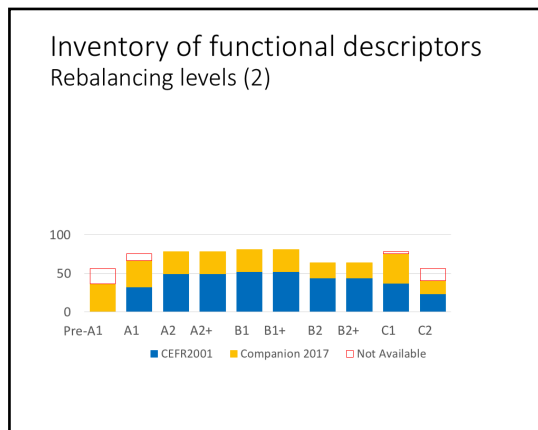


Figure 3

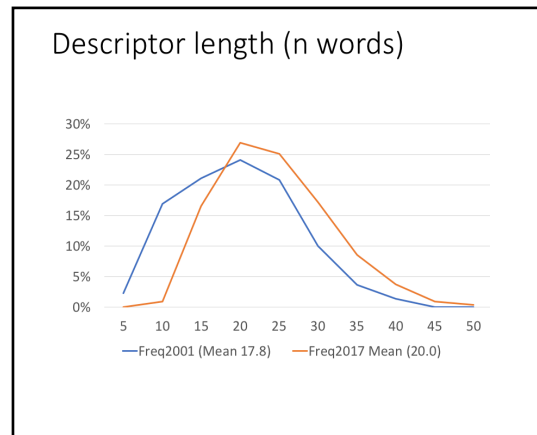


Figure 4

New scales have been developed for the areas of mediation and plurilingualism/pluriculturalism, and five new functional scales have been added: *Goal-oriented online transactions and collaboration*, *Online conversation and discussion*, *Reading as a leisure activity*, *Sustained monologue: giving information*, and *Using telecommunications*.

The addition of a pre-A1 level is an important innovation because so much language learning doesn’t get to A1; tourists, for example, are happy with pre-A1. More work is needed on this new level, but we’re getting there.

The titles of five 2001 scales have been slightly revised, including the two that referred to “native speaker”. This term has rightly been removed from both titles and the nine descriptors that contained it in 2001, because with, for example, the large numbers of illiterates among first language users, they cannot be taken as the model for all language learners to strive for.

It is worth pointing out that relating the levels of the CEFR to the old system – beginner, elementary, pre-intermediate, intermediate, upper intermediate, advanced, proficiency – is not very helpful because B2 is already advanced (Figure 5) and, most certainly, language learning must begin before reaching level A1.

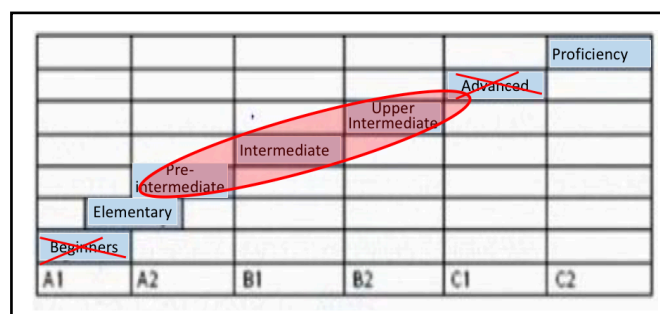


Figure 5: CEFR levels compared with traditional system

Altogether the CV runs to 229 pages and one wonders for whom it is intended. In my view the CV contains a lot of superfluous material (rehashing the 2001 CEFR publication), which makes it difficult to identify what is new. One way of addressing this would be to create a separate (shorter) volume for teachers focusing on what is new.

The appendices contain some useful material. Appendix 1, which summarizes salient features of spoken language for the successive CEFR levels (reproduced from the 2001 CEFR), is very helpful, as is Appendix 4, the written assessment grid. Appendix 5 gives an account of the development and validation of the new descriptors, but again there is a serious lack of numerical data. Appendix 7 lists the changes to 2001 descriptors, but we are given no data to show that the descriptors in question are at the same position on the scale as before. Appendix 8 gives the 93 sources of the 93 descriptors added to the 2001 scales, but without indicating which descriptors came from where. Finally, Appendix 9 offers descriptors for *Establishing a positive atmosphere* and *Visually representing information*. One wonders whether these are language abilities; and as they provide descriptors for levels B1 and B2 only, whether they actually constitute scales.

In conclusion I return to my earlier point. As it stands, the CV is a kind of (incomplete) project report. It should be split into two separate publications: (i) a teacher-friendly document with the scales and not much else, perhaps just a summary of the choices made about, for example, the reference to native speakers, and (ii) a scientific report that describes the methodology in detail and provides the statistical data. The discussion about what the CEFR is etc. should be left out: it is already present in the 2001 document.

Respondents

CEFR Companion Volume: first thoughts – Anthony Green, University of Bedfordshire

The CV highlights certain innovative areas of the CEFR for which no descriptor scales had been provided, especially mediation and plurilingual/pluricultural competence; this needed to be done and is helpful. The CV also builds on the successful implementation and further development of the CEFR, for example by more fully defining “plus levels”, introducing a new pre-A1 level, and introducing sign language. And it responds to the demand for more elaborate description of listening and reading and for descriptors for other communicative activities such as online interaction, using telecommunications, and expressing reactions to creative text and literature. In general, the description at A1 and the C levels, particularly C2, is enriched.

The first section of the CV discusses the CEFR’s role in language education and its action-oriented approach. It argues that the CEFR “does not prescribe any particular pedagogic approach” and is neutral in the sense that it “raises questions rather than answering them” (CV, p.27). At the same time, the CV insists that the CEFR is “not educationally neutral” because it implies that teaching and learning are “action-oriented” (ibid). What does this mean? That the CEFR raises questions but doesn’t tell us how to teach; that it is educationally committed and has to do with a particular approach to going about teaching and assessment. This approach treats the language user/learner as a “social agent”, acting in the social world and exerting agency in the learning process. The CV is, helpfully, more explicit than the 2001 CEFR in declaring that this implies a paradigm shift in both course planning and teaching, promoting learner engagement and autonomy.

The CV reminds us (p.28) that the CEFR is a tool to facilitate educational reform projects, helping us to think about what we are doing and why. It assists the planning of curricula, courses and examinations by working backwards from what the users/learners need to be able *to do* in the language, as illustrated in Figure 6. In this way it helps us to move away from syllabuses based on linear progression through language structures or predetermined sets of notions and functions, and it encourages us to abandon a *deficiency* perspective that focuses on what learners have not yet acquired. It helps us to move towards syllabuses based on needs analysis, oriented towards real-life tasks and constructed around purposefully selected notions and functions, taking account of who learners are and why they need to learn languages. And it encourages us to adopt a *proficiency* perspective guided by “can do” descriptors. Profiles of the kind illustrated in Figure 6 show what

people need to be able to do to perform particular roles, and the CV (p.42) offers two procedures for defining curriculum aims from a needs profile. Users need to look elsewhere, however, for elaboration and help with such a process.

The CV presents newly developed illustrative descriptor scales alongside existing ones; schematic tables group scales belonging to the same category (communicative language activities or aspects of competence); and a short rationale now precedes each scale, explaining the thinking behind its categorisation. We are reminded that descriptors are illustrative, open-ended and incomplete, and that they focus on aspects that are new and salient at the level in question and do not attempt to describe everything relevant in a comprehensive manner. The illustrative descriptors are one source for the development of standards appropriate to the context concerned; but they are not offered as either content or performance standards. As users of the CEFR we have to provide our own standards to suit language learning in our specific context. However, the CV gives only very limited suggestions on how such a process might be managed.

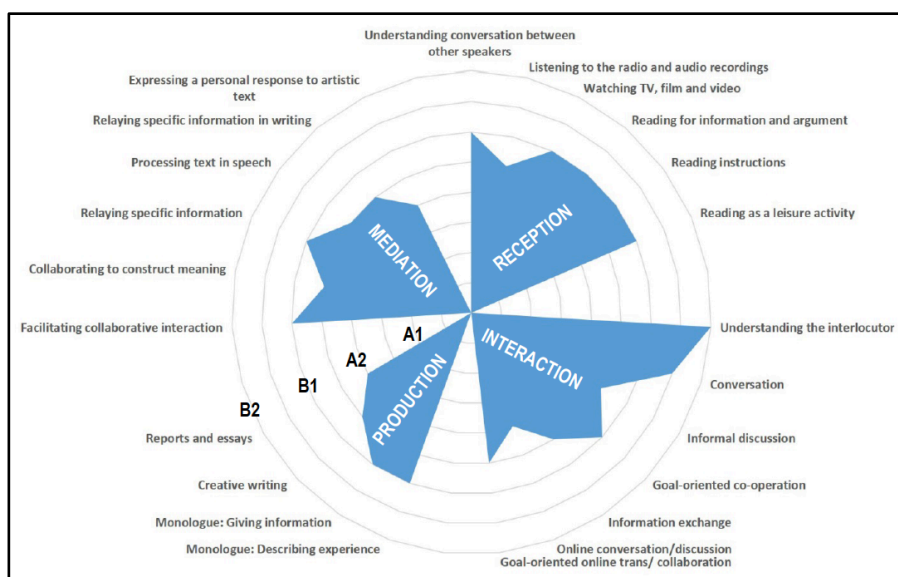


Figure 6: Invented profile of needs – lower secondary CLIL (CV, p.37)

Altogether, seven new scales have been added to those in the 2001 CEFR; pre-A1 is an entirely new level; and 15 of the 2001 descriptors have been adapted (seven at C2) across 12 scales. *No descriptor available* occurs in fewer scales than in 2001: four in C1 (previously nine) and 19 in C2 (previously 25). Fifty-six new descriptors have been added at C2 level (this total includes the new scales), and the 13 instances of “native speaker” have been eliminated. There has also been a scaling back on absolutes: we now have *Can understand with ease virtually any kind of spoken language* instead of *Has no difficulty with any kind of spoken language*, *Can understand virtually all forms of the written language* instead of *Can understand and interpret critically virtually all forms of the written language*, and *interacting authoritatively with effortless fluency* instead of *interacting authoritatively with complete fluency*.

Phonology has been subjected to detailed reconsideration. Intelligibility is now the key factor for discriminating between levels; descriptors focus on the effort required from the user/learner’s interlocutor to decode his/her message. Explicit mention is made of accent at all levels. Besides intelligibility, key concepts operationalized in the scale include the extent of influence from other languages spoken, control of sounds, and control of prosodic features.

Appendix 6 provides examples of the use of descriptors for online interaction and mediation activities in the personal, public, occupational and educational domains as shown in Figure 7. It would be useful to have this kind of elaboration for other scales.

Can explain (in Language B) the relevance of specific information given in a particular section of a long, complex text (written in Language A).			
PERSONAL	PUBLIC	OCCUPATIONAL	EDUCATIONAL
in an article, website, book or talk face-to-face/online concerning current affairs or an area of personal interest or concern	from presentations at public meetings, from public documents explaining policy changes, political speeches	a business report, article, regulation or workplace policy	an article, book, reference book or lecture/presentation

Figure 7: Mediation descriptor applied to different domains of language use (CV, p.175)

More generally, I feel it would be helpful to have further guidance on how to generate useful descriptors for local teaching and learning contexts. What design principles underlie the descriptors? If I'm going to elaborate on the CEFR for my own context, what needs to go into a descriptor? Given that its essential characteristics are definiteness, clarity, brevity and independence, what should a good descriptor provide for users? A task? A speech act, speech event or activity? Themes, topics, situations? An indication of the nature of input and output? Performance qualities and limitations? Restrictions, conditions and constraints? These features occur very unevenly in the descriptors we have, which also retain a lot of inconsistent wording, for example, *catch*, *recognize*, *follow*, *understand* in descriptors for reading and listening comprehension. The CV says nothing about how a descriptor might be elaborated, and this prompts the question: How can users of the CEFR carry forward the work provided here in their own working context and expand on what has been done in a way that is consistent with the principles of the CEFR itself?

The CEFR Companion Volume – Elif Kantarcioglu, Bilkent University

My reflections on the CV are from the perspective of a small-scale language tester: my university's proficiency exam is taken by five thousand students each year. Our experience of developing exams in relation to the CEFR is shaped by two considerations: (i) our exam construct is based on a theory of language, and (ii) learners' self-evaluation and teachers' evaluation of learner performance play a crucial part in exam validation. Institutions like ours should be the judges of whether the CEFR suffers from lack of theory, as some critics claimed when it was first published in 2001. In our institution, we have not observed any mismatches between the existing theory and what the CEFR offers.

By including more descriptors and scales, the CV allows us to better define test specifications (e.g. operations). Examples of this are new B2 descriptors for reading and listening. "Can recognize when a text provides factual information and when it seeks to convince readers of something" usefully supplements "Can understand articles and reports concerned with contemporary problems in which the writers adapt particular stances or viewpoints" (*Reading for information and argument*; CV, p.62); and "Can recognize speaker's point of view and distinguish this from facts that he/she is reporting" usefully supplements "Can follow the essentials of lectures, talks and reports and other forms of academic/professional presentation which are propositionally and linguistically complex" (*Listening as a member of a live audience*; CV, p.57).

The additional descriptors and scales also allow us to better specify expectations in marking schemes. This is the case, for example, with the new phonology scales, especially when they are read in conjunction with the report on their development, available online (<https://rm.coe.int/phonological-scale-revision-process-report-cefr/168073fff9>) but not included in the CV. The division of phonology into three scales – *Overall phonological control*, *Sound articulation* and *Prosodic*

features – is particularly helpful. An additional B2 descriptor for *Grammatical accuracy* also provides extra detail: “Has a good command of simple language structures and some complex grammatical forms, although he/she tends to use complex structures rigidly with some inaccuracy” is a useful complement to “Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding” (CV, p.132). The extension of the self-assessment grid with new descriptors and scales allows us to provide more detailed learner profiles, while the removal of “native speaker” references allows us to set more realistic goals and absolves us of the need to clarify which native speaker we are referring to.

On a more critical note, I like the descriptors included in the scale for *Reading as a leisure activity* but wonder about the relevance of this scale for testing purposes: the descriptors could be wholly relevant to some testing contexts, but the title prevents the scale being applicable to standardized assessment. Also, in the scale for *Vocabulary range* a new descriptor has been added to level B1: “Has a good command of a range of vocabulary related to familiar topics and everyday situations” (CV, p.131). But surely “range” in this descriptor requires a modifier, otherwise it may seem to be at a higher level than the B2 descriptor “Has a good range of vocabulary for matters connected to his/her field and most general topics”.

I conclude with two questions for reflection and discussion:

- What are the possible implications of the inclusion of “online interaction” in high-stakes exams?
- What mechanisms can be put in place to collect feedback and data from institutions engaged in high-stakes assessment in order to further develop scales that have been identified as incomplete in the CEFR?

What the new “can do” descriptors can do for classroom assessment – Armin Berger, University of Vienna

Implementation of the 2001 CEFR in the classroom has tended to be quite weak. This is not least due to a number of practical criticisms of the CEFR: it is not very accessible and user-friendly; many of the descriptors are vague and imprecise; additional sub-levels need to be defined; the descriptors are not applicable to young learners and unsuitable for higher education and teaching languages for specific purposes; and there’s a need for additional materials for teachers. Now that the new descriptors are available, the question arises: Is the CV likely to encourage a stronger implementation? Does it make us want to think again? The new descriptors certainly challenge our current practice in at least three respects: in terms of why, what and how we assess in the classroom.

Classroom assessment is primarily concerned not with the level of proficiency that has been reached but with what has been learnt/achieved. The additional descriptors serve to accentuate this distinction, strengthening the horizontal dimension of language learning. As the CV explains, “The reason the CEFR includes so many descriptor scales is to encourage users to develop differentiated profiles” (CV, p.36).

The illustrative descriptors commonly act as the *de facto* framework: what is covered by the descriptors is taught and tested, what is not covered is not; their illustrative nature tends to be forgotten. The new descriptors challenge the current *de facto* framework in various ways. For example, the scale for *Reading as a leisure activity* may encourage teachers to emphasize extensive reading, while the scale for *Online conversation and discussion* may remind us to develop curricula that take online communication into account. At the same time, the descriptors do not necessarily reflect the specifics of online communication or how the medium interacts with language use; and the relation between new and existing descriptors is not always clear, for example “Can express him/herself with clarity and precision in real-time online discussion, adjusting language flexibly and sensitively to context, including emotional, allusive and joking usage” (*Online conversation and*

discussion C2) and “Can use language flexibly and effectively for social purposes, including emotional, allusive and joking usage” (*Conversation C1*).

The CEFR describes learners as social agents, which implies the use of *meaningful* collaborative tasks and activities in the classroom. The addition of more descriptors that focus on the use of academic language at the lower levels is therefore a challenge to teaching traditions where content is merely a vehicle rather than an end in itself. The scale for *Listening as a member of a live audience* (CV, p.57) provides two examples: “Can follow a straightforward conference presentation or demonstration with visual support (e.g. slides, handouts) on a topic or product within his/her field, understanding explanations given” (B1) and “Can follow the general outline of a demonstration or presentation on a familiar or predictable topic, where the message is expressed slowly and clearly in simple language and there is visual support (e.g. slides, handouts)” (A2+).

Although the separation of skills is still common in classroom practice, the CEFR’s action-oriented approach encourages the development of integrated approaches to language learning, teaching and assessment in the classroom. We are reminded of this by two descriptors in the scale for *Goal-oriented online transactions and collaboration* (CV, p.95): “Can participate in complex projects requiring collaborative writing and redrafting as well as other forms of online collaboration, following and relaying instructions with precision in order to reach the goal” (C1) and “Can collaborate online with a group that is working on a project, justifying proposals, seeking clarification and playing a supportive role in order to accomplish shared tasks” (B2).

The CV describes the learner as “a ‘social agent’, acting in the social world and *exerting agency in the learning process*” (CV, p.26; emphasis added). This empowering vision implies that we should encourage learner engagement, foster learner autonomy and adopt assessment practices that accommodate learner agency. This entails making use of interactive assessment formats (Bell & Cowie 2001), procedures that require learners to exercise initiative and control (Black & Wiliam 2009), and dynamic forms of assessment (Poehner 2008).

These considerations suggest the following questions for reflection and discussion:

- In what ways can the new descriptors encourage CEFR implementation in a strong sense in classroom contexts?
- In what ways will the new descriptors shape the why, the what and the how of classroom assessment in different contexts?
- What could be done to make the CV more accessible to practitioners?

References

- Bell, B. & B. Cowie (2001). The characteristics of formative assessment in science education. *Science Education* 85(5): 536–53.
- Black, P. & D. Wiliam (2009.) Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* 21(1): 5–31.
- Poehner, M. E. (2008). *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting Second Language Development*. Berlin, Germany: Springer.

Discussion

Barry O’Sullivan disagreed with John de Jong’s suggestion that the CV should have been published in two separate versions, one for teachers and the other for researchers; he argued that everything needs to be together, the new work and information already contained in the 2001 CEFR. John de Jong wondered why, in that case, the Council of Europe had not prepared a second edition of the CEFR. Brian North explained that it’s a matter of status. The CV has not yet been launched, and a programme to roll out implementation and experimentation is planned for 2019–20, starting with a launch event in Strasbourg in May. The idea is that a set of studies will then be published in 2021 or 2022. After that one might consider whether or not to have a new edition of the CEFR.

Another participant suggested that whether or not the scales for *Online interaction* and *Reading for pleasure* are used for evaluation should depend on why these activities are being focused on. John de Jong argued that if you describe how an activity develops, that doesn't necessarily mean that you have to test it. He thought that the *Reading for pleasure* descriptors might usefully guide the development of graded readers, which are currently based on intuition. Tony Green agreed that the existence of a scale didn't necessarily imply the development of a test; he also pointed out that tests are not the only form of assessment.

Elif Kantarcioglu repeated that she likes the descriptors in the *Reading for pleasure* scale; her worry has to do with assessing what is defined as a leisure activity. Brian North explained that the authoring group had a lot of difficulty with the title of this scale; part of the problem was finding a formulation in English that had an obvious equivalent in French.

A participant who specializes in text linguistics pointed out that you don't need an elaborate vocabulary to talk about film; also that very little in the online descriptors is internet-specific. Beppie van den Bogaerde welcomed the inclusion of online communication in the CV because it plays a central role in Deaf communities and thus in Deaf education. In her view the new descriptors support the application of the CEFR's action-oriented approach to the classroom.

Jamie Dunlea thought it was time to consider developing a more dynamic approach to the use of real-time data on CEFR use.

Doris Froetscher noted that for some years the Austrian school-leaving exam has included the task type "blog", which turns out to be more motivating than other task types; the new scales will be useful in this area. Armin Berger argued that we need to be very clear what makes blogs different from other forms of written text; more research is needed in this area.

Tony Green pointed out that we shouldn't feel obliged to limit our tests to things that are scaled in the CEFR; we may need to develop our own scales.

Barry O'Sullivan argued that the mediation scales provide a rationale for existing test practice. The big question is not whether online interaction should be included but how you test it. What aspects of language do we need to focus on?

Brian North pointed out that the online scales imply digital literacy, which means that they do not focus exclusively on language. The mediation scales move in the direction of a holistic approach which implicates general tendencies in education that focus on dialogue and the co-construction of knowledge.

Session 2

The Mediation Scales – Barry O’Sullivan, British Council

Focus on the scales

Figure 8 provides an overview of the mediation scales. I began by thinking that they wouldn’t have much to offer large-scale assessment; now I’ve changed my mind completely.

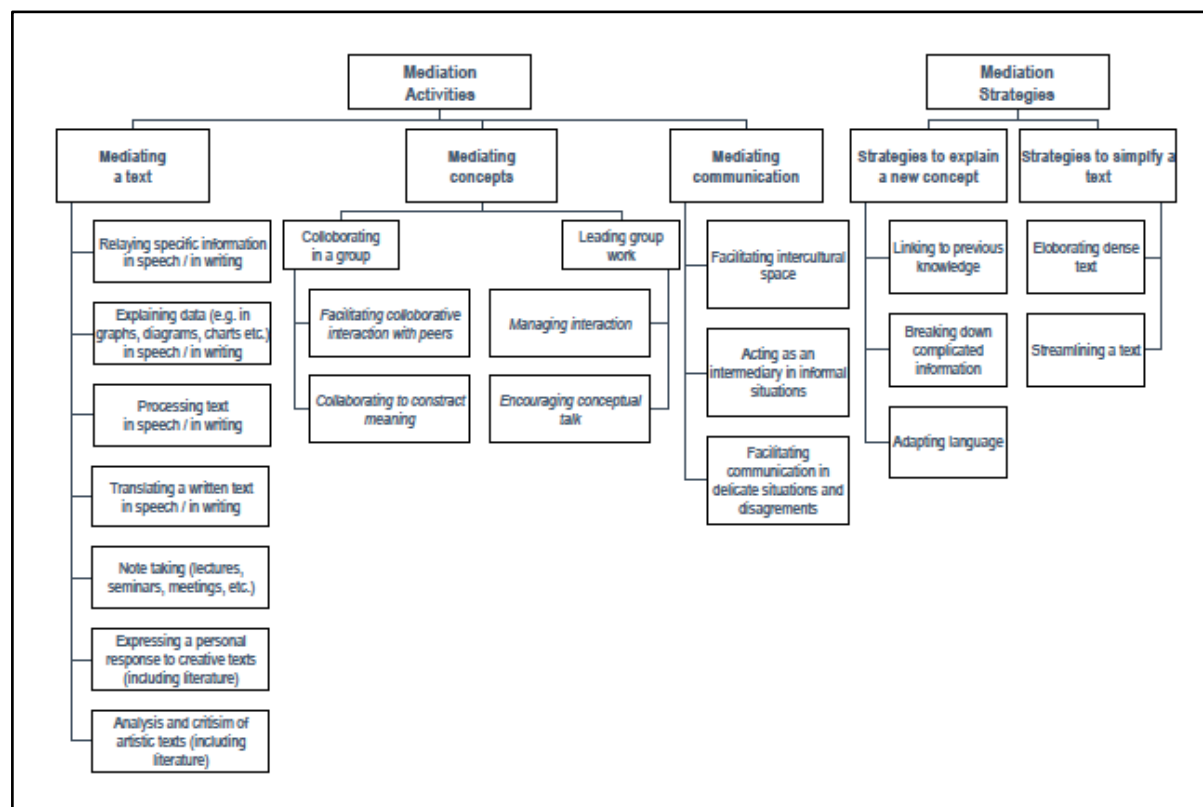


Figure 8: Schematic overview of mediation scales

While the 2001 version of the scale for *Processing text* focuses on the cognitive aspect and reflects Khalifa and Weir’s (2007) cognitive model of reading progression, the seven scales for *Mediating a text* focus on the social use of language, what we can do with a text. This reaffirmation of the social context of language use is welcome, though more work is needed on the intended recipient of the information in question. The text mediation scales introduce the concept of inter-language usage, whereas the original scales focused on intra-language usage (which is nevertheless retained: in any descriptor *Language A* can be the same as *Language B*). In these ways the new scales significantly broaden construct definition.

The scales for *Mediating concepts* focus on group work and may be used to add depth to our current construct operationalization, where mediation is implied but not explicitly dealt with. Different roles within group work are defined, so operationalization may make it necessary to designate roles. Some aspects are potentially problematic and will require greater examiner awareness; for example, “taking a leadership role” may be misconstrued as “taking over”. Co-construction in groups is not addressed (though it may not be addressable). Overall, the new scales broaden construct definition and can be operationalized.

The scales for *Mediating communication* focus on activities that are unlikely to be found in standardized international proficiency exams. *Facilitating intercultural space* includes showing interest, demonstrating sensitivity and respect, and dealing with misunderstandings (the first two of which

don't necessarily require that we use language). *Acting as an intermediary in informal settings* includes informally communicating the sense of what speakers are saying, conveying important information, and repeating the sense of what is expressed. And *Facilitating communication in delicate situations and disagreements* includes exploring participants' differing viewpoints, elaborating on viewpoints, establishing common ground, establishing possible areas of concession between participants, and mediating a shift in viewpoint of one or more participants, to move closer to an agreement or resolution. The concepts contained in these scales are observable, and it's feasible to create test tasks to reflect the activities they describe, most likely for a formative test of language use for this specific purpose. But while the scales for *Mediating communication* again broaden construct definition, it's unlikely that they can be operationalized in a formal summative test.

There are five scales for *Mediation strategies*, three for *Strategies to explain a new concept* (*Linking to previous knowledge*, *Breaking down complicated information*, *Adapting language*) and two for *Strategies to simplify a text* (*Amplifying a dense text*, *Streamlining a text*). Strategies, of course, underpin language use and communication, but they are not clearly observable and any attempt to include them in assessment would introduce an element of subjectivity detrimental to reliability and accuracy. At the same time, it's important to note that mediation strategies are described in terms of their manifestation (e.g. "can clearly explain", "can paraphrase", "can simplify"). These scales do not necessarily broaden construct definition, but they are a useful attempt to operationalize the strategy dimension

Task design

The mediation scales are clearly of no practical value in designing tasks to test the receptive skills, but they will be useful in designing tasks to test productive skills, especially integrated skills tasks (e.g. listening and/or reading into speaking or writing), for which they reflect a richer construct definition. They will clearly help not only in task design but also in scoring responses. As pointed out in the CV, not all scales will be relevant to all tasks. However, not enough thought has been given to the audience or reader

Implications for the test industry

At Language Testing Forum 2017, Constant Leung pointed to the limitations of construct definition in tests. The CV's mediation scales suggest that this is more problematic than previously thought: construct definition (and test tasks) need to be reconsidered. Specifically, construct definition should engage more explicitly with the social dimension of language use. This is clear in the test tasks I have reviewed, where the audience/reader is insufficiently defined. Construct definition should also recognize that meaning construction in groups is not easy to pinpoint and should take into account inter- and intra-language variation as a parameter of language use.

As noted above, the mediation scales are of no practical value to the design of tasks to test receptive skills, but they are of limited (but clear) value to traditional writing (and speaking) test tasks such as IELTS Task 2. They are of greater relevance to integrated skills tasks such as iBT (e.g. reading into writing). This suggests that new task types (e.g. group and inter-language) are required to take fullest advantage of the mediation scales.

As for test interpretation, current test tasks do not allow us to interpret performance in relation to mediation as it is defined in the CV. To allow for this, test tasks may need to be more localized, e.g. taking into account the candidates' L1. Tests should also include tasks that are more group/team-oriented and/or offer a more consistent definition of audience.

As regards consequences, the mediation scales pose significant challenges to scoring as traditional models under-define task fulfilment. Tests designed to reflect the concept of mediation (as defined in the CV) will have a positive washback by broadening our understanding of language use to recognize inter- and intra-language usage, the relevance of audience (reader or interlocutor), and the importance of an individual's role within a group.

Concluding thoughts

The 2001 scale for *Processing text* reflects a more cognitive model than the CV and should be retained as it is. The inclusion of different languages is likely to be problematic in that the relationship between the two languages will not always be the same; e.g. *Language A Chinese* to *Language B Thai* will not be the same as *Language A French* to *Language B Spanish* or *Language A Chinese* to *Language B French*. Also, not all groups are equal. The number of participants can impact on individual participation, as can the relationships between participants and an individual's perception of the other participants.

Applying the mediation scales is going to be more complex than initially envisaged. Test developers who use them to broaden their construct definition will find that the resulting tests will be more locally appropriate than current instruments. This will be positive for test users but quite problematic for test developers.

References

Khalifa, H. & C. J. Weir (2007). *Examining Reading*. Cambridge, UK: Cambridge University Press.
Leung, C. (2017). *Constructs in educational assessment – A view from the classroom*. Plenary Address to the Language Testing Forum, University of Huddersfield, November.

Respondents

Situating mediation within an argument for test use – Spiros Papageorgiou, Educational Testing Service

Test scores are used to facilitate various decisions, including admission, placement and certification, so they can have a significant impact on individuals and institutions. The decisions that are taken go beyond how a test taker performed on certain tasks on a certain occasion and under certain conditions.

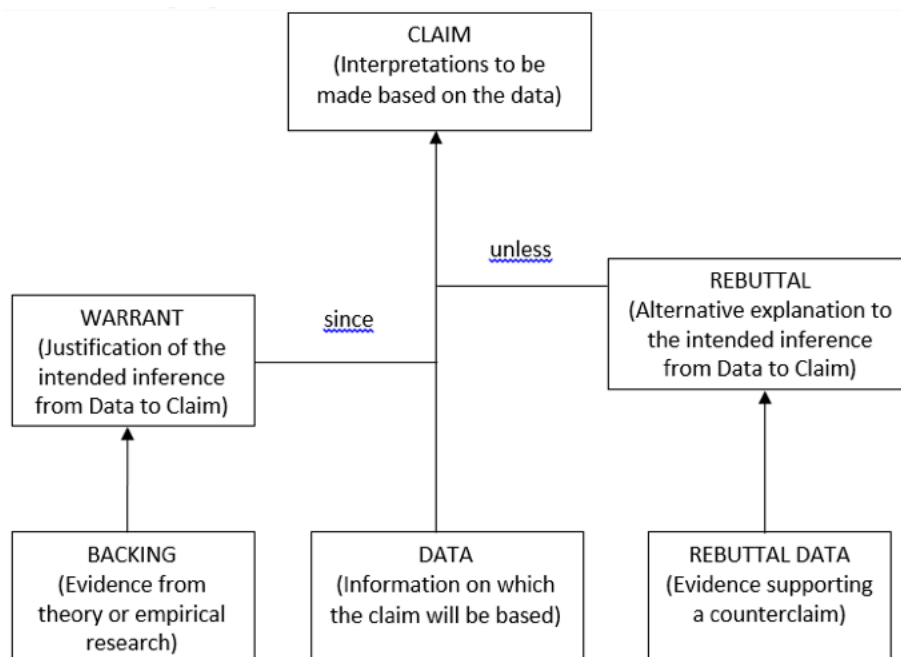


Figure 9: Core philosophy in an argument-based approach (adapted from Bachman 2005: 9–11)

Let me first explain the core philosophy in an argument-based approach. Toulmin's argument structure (Toulmin 2003) is at the core of argument-based approaches to validation (e.g. Chapelle 2008; Kane 2013; Bachman & Palmer 2010). The major premise of Toulmin's argument structure (see Figure 9) consists of providing support for claims based on data, that is, information or facts on which the claim is based. To provide this support, a warrant is stated, which in turn is supported by backings coming from theoretical or empirical evidence. Contrary to warrants, rebuttals function as alternative hypotheses which can challenge the claim. Data can be gathered to help support warrants or to help support, weaken, or reject rebuttals. An argument-based approach to validation has two components: (i) specification of the proposed interpretation and uses in some detail (the *interpretation/use argument*), and (ii) evaluation of the overall plausibility of the proposed interpretations and uses (the *validity argument*). Research evidence for an interpretation/use argument may relate to inferences about: test content representation and relevance to the real world; the accuracy of scores and the extent to which items measure the ability of interest; the reliability of scores – their consistency across test forms, administrations, etc.); the meaningfulness of scores – whether or not they are indicative of the language ability to be measured; whether scores predict performance in the real world; and the usefulness of scores for decision making and positive consequences.

This is how the CEFR defines mediation (2.1.3; Council of Europe 2001: 14):

In both the receptive and productive modes, the written and/or oral activities of *mediation* make communication possible between persons who are unable, for whatever reason to communicate with each other directly. Translation or interpretation, a paraphrase, summary or record, provides for a third party a (re)formulation of a source text to which this third party does not have direct access. Mediation language activities, (re)processing an existing text, occupy an important place in the normal linguistic functioning of our societies.

According to this definition, mediation is not only a cross-linguistic skill: mediation tasks may also be operationalized within one language. An example of a cross-linguistic mediation test task would ask candidates to read a newspaper report in one language and summarize it in an e-mail of 80 words in another language. Within-language mediation can be assessed in a variety of ways. Takers of an academic English test may be given a short written text, for example on the psychological concept of “flow”, and a video clip that provides an example, and their task is to explain “flow” and how the video example illustrates the concept. Younger learners may be shown an animated video in two parts. In the first a teacher explains what happens when blue paint is mixed with yellow; in the second a pupil explains that she missed class and asks the test taker to explain what happened (see Wolf et al. 2017).

In the context of large-scale language proficiency testing, we have to consider what score-based decisions need to be made that require mediation tasks: is mediation relevant to the target language use domain? Does cross-linguistic mediation raise fairness concerns? For example, is content comparable across different versions of the test? And can we be sure that we are not assessing non-linguistic competence? Is within-language mediation sufficiently operationalized in task design? Are we striking the right balance between situational authenticity and exploring innovative ways of enhancing the interactional authenticity of test tasks? Finally, what is the actual use of the text?

References

- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly* 2(1): 1-34.
- Bachman, L. F. & A. S. Palmer (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying their Use in the Real World*. Oxford: Oxford University Press.
- Chapelle, C. A., M. K. Enright & J. M. Jamieson (eds) (2008). *Building a Validity Argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 50(1), 1-73.
- Toulmin, S. E. (2003). *The Uses of Argument*. Cambridge, UK: Cambridge University Press.
- Wolf, M. K., A. Lopez, S. Oh & F. Tsutagawa (2017). Comparing the performance of young English language learners and native English speakers on speaking assessment tasks. In M. K. Wolf & Y. G. Butler (eds), *English Language Proficiency Assessments for Young Learners*, pp.171-190. New York, NY: Routledge.

Situating mediation within the classroom – Dina Tsagari, OsloMet – Oslo Metropolitan University (formerly HiOA)

My first thought on reading the mediation scales in the CV was that if we as language teachers or our students mediate to this level and with this degree of complexity, we're going to change the world! My reflections will focus in turn on the role of the L1 in L2 pedagogy, the status of dialects and varieties, the emphasis of the mediation descriptors on facilitation, classroom-based assessment, and the role of mediators.

The CV explains (p.103) that in the mediation scales, language A and language B may be “two different languages, two varieties of the same language, two registers of the same variety, or any combination of the above. However, they may also be identical.” The status of L1 and varieties to support and mediate an L2 is, however, somewhat contested in L2 classrooms in many educational contexts. For example, research carried out in Cyprus (Tsagari & Georgiou 2016; Tsagari & Gianikas 2017; Tsagari & Diakou 2015) shows that even though they believe they should use only the target language in their classes, teachers resort to the L1 for a variety of purposes. They use it in instruction: to translate words and sentences; to present and explain grammar and other target language forms; to check comprehension; and to draw contrasts between L1 and L2. They also use the L1 for classroom management: to give and explain instructions; to answer students' requests; to encourage students to participate in the L2; to discipline and reprimand; and to deal with administrative issues. And they use the L1 for social and cultural purposes: to give feedback and praise; to establish rapport; and for humorous and friendly talk. However, using the L1 in these ways makes teachers feel guilty, which prompts the question: Are teachers ready for mediation as presented in the CV?

With regard to the use of varieties, things are not very different. There are various contexts that are similar to the one I describe here. The 2010 Cypriot National Curriculum for Language (Ministry of Education and Culture 2010) capitalized on the fundamental constructivist assumption that competence in a language variety that is extraneous to the speech community, in this case Standard Modern Greek, can be developed by making use the students' native language, in this case Cypriot Greek, as linguistic capital while respecting their right to use their L1. In other words, teachers could use Cypriot Greek to mediate the teaching of Standard Modern Greek (Tsiplakou 2015, in press). However, political forces and other pressures led to the withdrawal of this curriculum, which prompts a second question: Are educational systems ready to support and use mediation?

According to the CV, “A person who engages in mediation activity needs to have *a well-developed emotional intelligence*, or an *openness* to develop it, in order to have *sufficient* empathy for the *viewpoints and emotional states* of other participants in the communicative situation” (p.102; emphasis added). But what if these features are missing (see, for example, Kliueva & Tsagari under review)? The CV also tells us that the scale for *Facilitating pluricultural space* “reflects the notion of creating a shared space between and among linguistically and culturally different interlocutors, i.e. *the capacity of dealing with 'otherness' to identify similarities and differences to build on known and unknown cultural features, etc. in order to enable communication and collaboration*” (CV, p.120; emphasis added). In other words, the language user/learner acts as a “cultural mediator”, though the capacity to deal with “otherness” raises political, psychological and cultural issues.

When it comes to classroom-based assessment, teachers are expected to be assessment literate, i.e. they should be skilled in choosing and developing assessment methods appropriate to their instructional decisions; administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods; using assessment results when making decisions about individual students, planning teaching, developing curriculum, and improving institutional quality; developing, using and evaluating valid student grading procedures; and communicating assessment results to students, educational decision makers and other relevant stakeholders. These considerations prompt a third question: How can teachers be helped to do “mediation” in the classroom as part of their other assessment mandates.

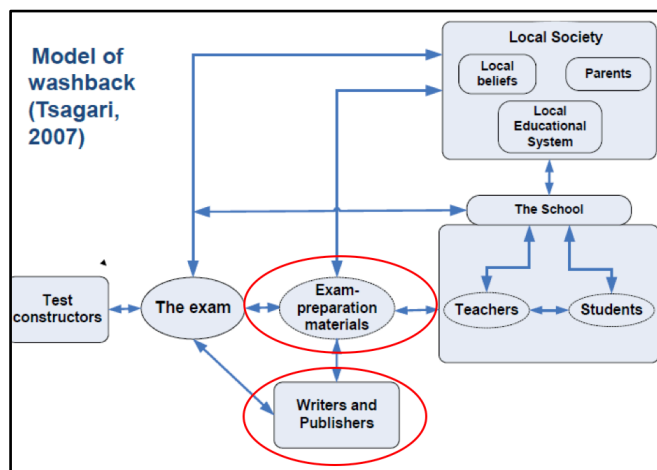


Figure 10: Model of washback (Tsagari 2007)

As in the processes of washback (Figure 10), which is mediated by, e.g., teachers, textbook writers/publishers, etc., so many stakeholders along with language users need to familiarize themselves with the nature/content of the mediation scales in the CV and know how to operationalize them effectively in their practice (e.g. teaching, assessment, textbooks).

Overall, consideration of the role of mediation in the classroom prompts four questions: Are educational systems and teachers ready to employ and sustain mediation? What kind of accommodations and conditions need to be in place for mediation (scale descriptors) to be implemented successfully? What levels of language assessment literacy (skills, knowledge, abilities) do teachers, textbook writers/publishers, and examination boards need if they are to implement the mediation scales? And, finally, what is the profile of the language user who engages in mediation?

References

- Kliueva, E. & D. Tsagari (under review). Emotional literacy in the EFL classroom: teachers' views and practices. *System*.
- Tsagari, D. (2007). Investigating the Washback Effect of a High-Stakes EFL Exam in the Greek Context: Participants' Perceptions, Material Design and Classroom Applications. Published PhD thesis. Department of Linguistics and English Language, University of Lancaster, UK.
- Tsagari, D. & C. Diakou (2015). Students' and teachers' attitudes towards the use of the First Language in the EFL state school classrooms. *Research Papers in Language Teaching and Learning*, 6(1): 86–108 Available at <http://rpltl.eap.gr/images/2015/06-01-086-TSAGARI-DIAKOU.pdf>
- Tsagari, D. & E. Georgiou (2016). Use of mother tongue in second language learning: voices and practices in private language education in Cyprus. *Mediterranean Language Review* 23: 101–126. DOI: 10.13173/medilangrevi.23.2016.0101
- Tsagari, D. & C. N. Giannikas (2017). To L1 or not to L1, that is the question: research in young learners' foreign language classroom. In Y. Bayyurt & N. Sifakis (eds), *English Language Education Policies and Practices in the Mediterranean Countries and Beyond*, 131–152. Frankfurt am Main: Peter Lang. <http://dx.doi.org/10.3726/b10989>.
- Tsiplakou, S. (2015). Διδάσκοντας διάλεκτο σε ένα παιδαγωγικό πρόγραμμα κριτικού γραμματισμού: η εκπαιδευτική μεταρρύθμιση της Κύπρου [Teaching dialect within a critical literacy pedagogical program: the education reform of Cyprus]. In M. Tzakosta (ed.), *Teaching Modern*

Greek Varieties and Dialects in Primary and Secondary Education: Theoretical Approaches and Pedagogical Implementations, 187–210. Athens: Gutenberg.

Tsiplakou, S. (in press). Which dialect? What literacy? *Proceedings of Modern Greek Dialects and Linguistic Theory* 7. Ρέθυμνο: Πανεπιστήμιο Κρήτης.

Discussion

John de Jong noted that PISA 2015 included a test of collaborative problem solving, while PISA 2018 will seek to assess global competence (defined by the OECD website as “the capacity to examine local, global and intercultural issues, to understand and appreciate the perspectives and world views of others, to engage in open, appropriate and effective interactions with people from different cultures, and to act for collective well-being and sustainable development”). He argued that although language encroaches on everything, we don’t have to include everything in teaching and testing language. In his view the mediation scales go too far in the direction of personality issues and ethics; people who don’t have empathy or emotional intelligence can still learn and use languages very effectively. We mustn’t lose sight of the fact that if we adopt a multi-trait, multi-method approach to validating a test, we still have to determine which traits form a construct independently of how it is measured.

Brian North wondered whether the mediation scales have implications for Bachman and Palmer’s concept of communicative language ability. He reminded John de Jong that the CEFR is a political project based on values: hence the nature of the mediation descriptors. Is it acceptable, given the current state of affairs in Europe, for teachers to decline responsibility for these values? Supporting this point, Joe Sheils explained that the mediation descriptors are a close relative of the Council of Europe’s recently developed Competences for Democratic Citizenship.

Barry O’Sullivan said that in the 60 years since Cronbach we’ve developed a much richer understanding of language and thus of the construct. The 2001 CEFR provided a common language for discussing levels; the CV helps us to broaden our understanding of the construct. Mediation is happening all the time, but how we operationalize it is a challenge. It’s really difficult to take the mediation scales out of their context in the CV; they fit together and thus broaden the construct. This was not the case with the original scales.

A student asked: What are you testing? Language itself or tools to test the language? Barry O’Sullivan replied that Cronbach was concerned with tools, and Spolsky’s observation that TOEFL was a very good test of *something* reflected the same tendency. Now the emphasis is much more on use of language, but we may need to remember that tools are important too.

Spiros Papageorgiou pointed out that there are multiple purposes for testing and the decisions we take should depend on the context in which we’re working. EMI contexts, for example, are very different to EAP in English-speaking countries.

Dina Tzagari explained that she’s not negative towards the mediation scales. The descriptors encourage us to reflect on what it means to be a European citizen in the twenty-first century. From another perspective, mediation between speakers of different languages/dialects/varieties as agents of “culture” may be more reflective of political decisions. Nevertheless, reflecting on the mediation descriptors and operationalizing them in the classroom is very challenging.

Barry O’Sullivan argued that mediation isn’t new either in classrooms or in tests. What is new is the way in which the descriptors and scales encourage us to think about mediation in relation to learning and assessment. Inclusion of the mediation scales in assessment emphasizes the need for localization.

John de Jong was concerned that the discussion was neglecting measurement principles, which we need to keep sane and clean, likewise our constructs.

Brian North suggested that one of the main reasons for the CEFR’s success is that it describes the functional competence that governments want but also has a values dimension. The skills that

are dealt with in the mediation scales are needed in the real world. It's also important to remember that the CEFR is all about profiling. It's possible to be C2 in many ways but not in mediation.

Tony Green recalled that in 1991 Geoff Brindley pointed out that in language teaching and testing we have a very limited view of communication; the mediation scales help us to take a broader view.

Session 3

The pluricultural and plurilingual scales – Vincent Folny, CIEP (Centre international d'études pédagogiques)

Plurilingualism is one of the defining features of the twenty-first century world. An example of “good practice” in culture and the arts is provided by the website of *arte*, the Franco-German television station, which exists in five versions: English, French, German, Polish and Spanish. A visitor to the English-language version can watch a video clip that has a German sound track with subtitles in English, and there are numerous other language combinations. The hairdresser’s window in Figure 11 shows a more trivial mixing of languages. The assumption is that prospective customers will be attracted by the combination of English and French, though elderly monolinguals may find it confusing.



Figure 11: Hairdresser’s window, Paris

The CV defines plurilingualism as “the dynamic and developing linguistic repertoire of an individual user-learner”; and in the 2001 CEFR plurilingualism is presented as an “uneven and changing competence” (CEFR 6.1.3.1, p.133) in which the user/learner typically has different resources in the various languages or varieties that make up his/her repertoire. The fundamental point is that plurilinguals have a single, interrelated repertoire that they combine with their general competences and various strategies in order to perform tasks. From this definition we can infer that language acquisition is an ongoing project that is both individual and collective; that language repertoires should be dynamic not static; that language acquisition is not an instantaneous process; that language attrition should be taken very seriously – it too is dynamic; and that our capacity to learn new languages is what enables us to develop and interact with society.

The CEFR defines plurilingual and pluricultural competence as “the ability to use languages for the purposes of communication and to take part in intercultural interaction, where a person, viewed as a social agent has proficiency, of varying degrees, in several languages and experience of several cultures. This is not seen as the superposition or juxtaposition of distinct competences, but rather as the existence of a complex or even composite competence on which the user may draw”

(CEFR 8.1; p.168). From this we can infer that similarities and differences work together; skills, knowledge and adaptability are key for communication between social actors.

According to Eurobarometer 386, “Just over half of all Europeans claim to speak at least one other language in addition to their mother tongue. The majority of Europeans (54%) are able to hold a conversation in at least one additional language, a quarter (25%) are able to speak at least two additional languages and one in ten (10%) are conversant in at least three.” It’s important to note, however, that these figures were lower than in 2005, and more than 50% of the population can use at least two foreign languages in only eight EU member states: Luxembourg (84%), the Netherlands (77%), Slovenia (67%), Malta (59%), Denmark (58%), Latvia (54%), Lithuania (52%), Estonia (52%).

The CV includes three “pluri” scales, as shown in Figure 12.

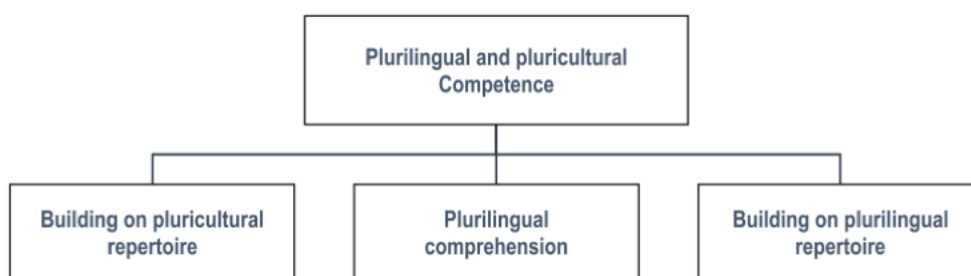


Figure 12: CV scales for plurilingual and pluricultural competence

A qualitative analysis of the descriptors demonstrates that the principal dimensions of the scale for *Building on pluricultural repertoire* are knowledge (14 references; e.g. “Can identify and reflect on similarities and differences in culturally determined behaviour patterns”), otherness (12 references; e.g. “Can identify differences in socio-linguistic/-pragmatic conventions”), reflexivity/meta-competence/mediation (10 instances; e.g. “Can deal with ambiguity in cross-cultural communication and express his/her reactions constructively and culturally appropriately in order to bring clarity”), and adaptiveness (six references; e.g. “Can deal with ambiguity in cross-cultural communication”). Descriptors in this scale imply an algorithmic structure, from action or cognitive activity (20 references; e.g. “Can interact in a multilingual context on abstract and specialized topics”, “Can explore similarities and differences between metaphors and other figures of speech”), to resources (16 references; e.g. “in the languages in his/her plurilingual repertoire”, “an utterance [or] an expression from another language”), to aims (16 references; e.g. “either for rhetoric[all] effect or for fun”, “in order to improve understanding in a discussion of abstract and specialized topics”), to conditions or context (11 references; e.g. “more familiar to the interlocutor(s)”, “between third parties who lack a common language”).

The principal dimensions of the scale for *Plurilingual comprehension* similarly imply an algorithmic structure, from action or cognitive activity (16 references; e.g. “Can use what he/she has understood in one language”), to resources (14 references; e.g. “a text in another language (e.g. when reading short newspaper articles on the same theme written in different languages”), to aim (11 references; e.g. “to find relevant information”). From that analysis, we can infer that the plurilingual scales are strongly linked to the action-oriented approach.

When we consider possible uses of the “pluri” scales, it’s important to take account of the difference between observation, assessment and measurement. We can observe the use of different languages by the same speaker, student or candidate; we can assess the same phenomenon, judging (for example) that communication proceeds without serious obstacles; and we can measure each speaker’s productive proficiency in the languages he/she uses. In this scenario it’s unclear, however, what we do about receptive proficiency.

It would be very difficult to apply the plurilingual scales to existing exams because we mostly measure competence in a single language in order to take decisions about people in a high-stakes

context, for which we need reliability. We want to encourage the development of plurilingual competence, but measuring it is a challenge that has not been resolved. Perhaps we should be thinking of taking joint action with test providers in other languages and/or seeking advice from the users of our tests. Plurilingual competence could usefully be assessed in a number of domains, for example: in bilingual education, various vocational contexts (hospitals, airports, international companies), and language teacher education.

The pluricultural scale can be used for assessment but not for measurement because there's nothing to measure. It's important to remember, however, that the scale touches on issues that are sensitive in many countries. It's also worth pointing out that although the word "culture" is used 52 times in the CEFR and 25 times in the CV (it collocates with "shared", "communication" and "democratic"), neither document contains a clear definition of culture.

International companies are likely to be interested in this scale, as are European areas where "different" cultures are "identified". When using the pluricultural scale, we must never forget that its focus is on facilitating communication; and we must take account of our own context and the policy of our country, school or company. The simplest way of applying the scale will be by focusing on cultural knowledge based on fact, though it will be important to select content with care. I can speak about trains in Japan and how they make daily life there different from daily life in America, and I can say something about what this reveals about Japanese and American values. But if we decide to focus directly on values ... good luck!

Respondents

Pluricultural and plurilingual scales: responses from the classroom and standardized testing – Jamie Dunlea, British Council / Gudrun Erickson, University of Gothenburg

We begin with three questions. If we are to test, what do we need to focus on? Should we test at all? If not, what are the "pluri" scales useful for?

If we are to test ...

According to the CV, "the fundamental point is that plurilinguals have a *single, interrelated, repertoire* that they combine with their *general competences* and various strategies in order to *accomplish tasks*" (p.28; emphasis added); and "the level of each descriptor in the scale *Building on plurilingual repertoire* is the functional level of the weaker language in the combination" (p.50). Figure 13 (Figure 8 in the CV, p.39) helps us to understand how this is meant to work in practice.

The CV also tells us that "the descriptor scales are [...] reference tools. They are not intended to be used as assessment instruments, though they can be a source for the development of such instruments" (p.41). Plurilingual comprehension usually involves activities like exploiting one's receptive ability in one language (however partial) to deduce the meaning of texts written in another language. Again, it is the minimum functional level needed in each of the languages concerned to perform these activities that the descriptor

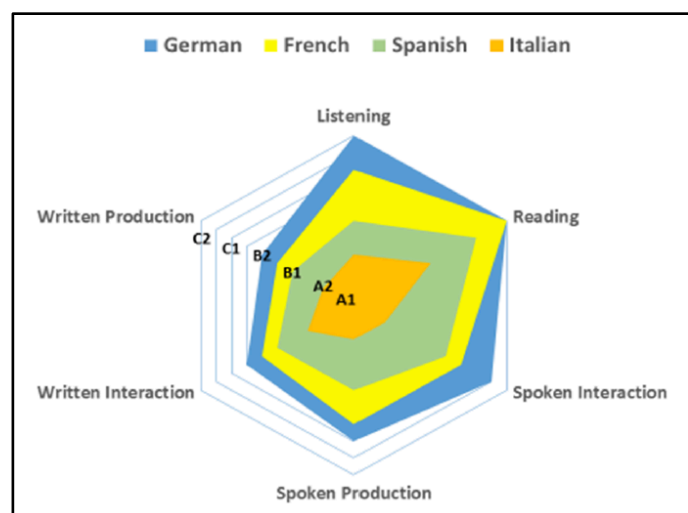


Figure 13: Example of a plurilingual profile

scaling refers to. In our view, we need more explicit links between the plurilingualism descriptors and the scales used as benchmarks of performance in the communicative activities and competences. While a lot of work has been done on the use and interpretation of CEFR scales for individual languages – localization in the form of European Language Portfolios, etc. – we still have little in the way of cross-linguistic comparison and validation. In this connexion it is worth recalling these words of Messick (1986): “One recommendation is to contrast the potential social consequences of the proposed testing with those of alternative procedures and even of procedures antagonistic to testing, such as not testing at all.”

Where are the “pluri” scales useful? Aspects to consider

Plurilingualism and pluriculturalism are tremendous assets at individual, educational and societal levels, not least in language learning. Defined and handled adequately, they may widen horizons, enhance communication, promote learning, and contribute to awareness-raising at all levels. They have an obvious role to play in teacher education; this applies particularly to the mediation scale for *Facilitating pluricultural space*.

The theory of constructive alignment (Biggs 2003; Biggs & Tang 2011) calls for coherence between curriculum, teaching and assessment. But does this mean that everything in the curriculum has to be assessed – and graded – at individual level? From an assessment perspective, how are we to handle observations that indicate less desirable attitudes and/or behaviour? Do we need to distinguish between learning and acquisition? Do we need to distinguish between classrooms for L1, L2 and FL? What construct does plurilingual and pluricultural competence belong to? And how is it related to the individual languages that teachers teach and assess/grade? Clearly, we must take account of contextual and individual differences, for example, more or less monolingual and monocultural contexts/schools.

The plurilingualism approach emphasises the interaction of the individual as language learner with all the social factors attendant on and intimately intertwined with the development and growth/change that take place as we not only learn a language but learn how to communicate and interact with other users of that language. It is the knowledge of how that language interacts with users and use that alters and expands our world view and our perceptions of ourselves.

The CEFR was developed as a non-language-specific framework and is therefore under-specified at the local and individual language level (it was always intended that this work would be carried out locally). Outside Europe the reverse approach has sometimes been taken. English language learning in one particular context was the starting point for the development of the Chinese Standards for English and the CEFR-J in Japan, and the use of the CEFR as benchmarks in Vietnam. The next step in those countries is to expand or transfer those frameworks to other languages. That process should include recognition of the rich linguistic diversity within their own borders, something that the plurilingualism scales can facilitate.

English is swamping the language education landscape, as the *European Survey on Language Competences* (European Commission 2012) confirmed: in all but two of the 15 participating countries/regions, English was the first L2 (the two exceptions were the Dutch and German-speaking regions of Belgium, where the first L2 was French). With this dominance in mind, the promotion of plurilingualism could have an impact on language policy at national levels. We are not suggesting that Europe should dictate or prescribe: we who have a wealth of experience of operationalizing and localizing the CEFR in contexts across Europe need to be prepared to engage with a wider audience. In many ways this engagement should reflect the principles underpinning the plurilingualism approach in the CEFR: we engage as equals and discuss and exchange information, and in so doing both sides will learn and grow.

Tentative conclusions

We need to think about how the plurilingualism descriptors link to and can interact with the more granular descriptions already in the CEFR; we also need to push forward cross-linguistic

comparability studies. The plurilingualism concept in the CEFR has immense potential for positive impact in and outside Europe at policy level and at the micro, classroom and teacher perception level. But we shouldn't limit its application to standardized testing. We should be using this opportunity to expand our approaches to evaluation and assessment beyond tests, and even avoid evaluation where unnecessary. And we need to be proactively engaged in reaching out to contexts beyond Europe in which the CEFR is being used in order to learn from one another.

Finally, it is worth returning to the two questions that underlie our reflections: Should we test at all? And if we shouldn't, in what ways are the "pluri" scales useful?

References

- Biggs, J. (2003). Aligning teaching for constructive learning. The Higher Education Academy. https://www.heacademy.ac.uk/system/files/resources/id477_aligning_teaching_for_constructing_learning.pdf.
- Biggs, J. & C. Tang (2011). *Teaching for Quality Learning*, 4th edition. Maidenhead, UK: McGraw Hill & Open University Press.
- European Commission (2012). *First European Survey on Language Competences*. Brussels: European Commission. http://ec.europa.eu/dgs/education_culture/repository/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf.
- Messick, S. (1986). The once and future issues of validity: assessing the meaning and consequences of measurement. ETS Research Report. Princeton, NJ: Educational Testing Service. <http://onlinelibrary.wiley.com/doi/10.1002/j.2330-8516.1986.tb00185.x/pdf>.

Discussion

Barry O'Sullivan argued that the "pluri" scales are quite Europe-centric and wondered whether they are the least exportable part of the CEFR. They may not be applicable to large-scale standardized testing, but they immediately become relevant when tests are localized. Jamie Dunlea agreed: the descriptors intrude on a very political and values-laden space, but they could be used to stimulate discussion of language policy, which is very widespread, especially in countries like India, where people are naturally plurilingual.

Brian North pointed out that European interest in plurilingualism is very recent; it was the feature of the CEFR that was noticed last. At the Language Policy Forum held in Strasbourg in 2007, John Trim gave a talk in which he said that, alas, plurilingualism was still for the future. Piloting and consultation carried out while they were being developed concluded that the "pluri" scales could be used for goal-setting, awareness-raising and self-assessment, but not for standardized testing – though Austria has recently introduced an optional Matura exam in mediation, using L2 and L3.

Joe Sheils recalled that the CEFR took as its starting point the European Cultural Convention, which implies respect for one's own plurilingual repertoire and the plurilingual repertoires of others. Forced reconfiguration of an individual's linguistic repertoire is inimical to the CoE's values. John de Jong noted that in Kazakhstan the language of schooling is either Russian or Kazakh; the latter is an imposed language, and in PISA Kazakh-medium schools are at a disadvantage.

Jamie Dunlea observed that a recent conference in China on educating the whole person involved "European" topics: critical thinking etc.

Brian North argued that the widespread belief that the brain has a limited capacity for languages helps to account for some of the conflicts in language policy. Many Swiss parents have this belief, which explains the opposition to increased language learning at school. The political challenge is to decide in which domains official languages should be spoken.

Session 4

Group discussion

Group 1: Updates to the CEFR – rapporteur: Ülle Türk

John de Jong launched the discussion by asking participants to say what they thought about the updates to the CEFR. Some of the views recorded here were expressed by several people, and some issues were discussed more thoroughly than others.

It was generally felt that the increase in the number of descriptors is positive because it strengthens the idea that the scales are illustrative and can be added to depending on the context. There are now many more scales to choose from, which is particularly useful for different teaching contexts. At the same time, the fact that there are so many scales may make it difficult for people to get the full picture of what is available.

There was a lot of discussion about the difficulty of ensuring that the scales are used properly, that the levels are not reduced to handy labels. It seemed to be agreed that though teachers probably won't read the CV, having two documents – one for teachers and the other for researchers – might have more disadvantages than advantages. It might result, for example, in fewer people understanding the principles behind the levels and scales. There was general agreement that the CEFR and CV should be mediated to teachers via teacher education.

This latter point led to a discussion of the extent to which the CEFR is actually used in teacher education. John de Jong suggested that someone should try to find out by approaching associations of vocational and university teachers.

There was broad agreement that the pre-A1 level is a very useful addition, but it was noted with concern that a lot of gaps remain. A need was felt for more descriptors at the lower levels generally (A1, A1+?). John de Jong recommended Pearson's global scale for English (<http://www.english.com>), which has many descriptors at lower levels both for language activities and for vocabulary and grammar.

Participants also seemed happy with the new phonology scales as they make it possible to map learners' strengths and weaknesses in a more differentiated way; some learners, for example, may have problems with pronouncing individual sounds, others with prosody.

The plurilingual and pluricultural scales were seen as very useful in contexts similar to that of Scottish Gaelic, where they could be used to show people that learning the language does not mean that their English will be damaged or reduced. But there was disagreement about whether these scales can or should be used for testing. Some people felt that in order to ensure that teachers focus on these aspects, students should have to take a test. This led to a more general discussion about the relationship between the CEFR, curricula, teaching and testing. No conclusion was reached but several people stressed that the CEFR is not just for testing and that not everything that should be taught can or should be tested.

The literature and reading-for-leisure scales were seen as another very positive addition, though again there was no agreement on whether they can be used in testing. It was noted that in Austria an attempt has been made to test literature in the writing component of the school-leaving exam.

There was some discussion of the length of the new descriptors. It was suggested that when descriptors are too long, people focus on some parts of them and neglect the others, thus distorting their meaning.

Another question was raised regarding the descriptors which have been changed: Did the changes lead to a change in their perceived level of difficulty? And have the descriptors in question been recalibrated?

Group 2: Mediation – rapporteur: Stergiani Kostopoulou

The group began by asking whether teachers are ready to use mediation. It was pointed out that teachers are mediating all the time, so the mediation scales should not be presented as something new or challenging. This prompted a second question: Do they teach their learners to mediate? Dina Tsagari argued that we need research to discover what exactly teachers need to do in the language classroom and to establish what mediation tasks textbooks include. A third question was then raised: What are the implications for teacher assessment? A participant who was a Cambridge examiner for speaking said: “On page 144 of the CV, I find things like ‘turn taking’ and ‘interactive communication’ and I recognize these. As a teacher, I have these descriptors in my head and I do them in my teaching.” It was suggested that perhaps both teachers and students need to be aware of what’s involved in mediation. It was also pointed out that teachers use mediation when teaching other subjects, and students read information online all the time and mediate it.

There was some discussion of the interaction between the sociolinguistic, pragmatic and mediation descriptors. It was agreed that it’s not easy to separate mediation from paralinguistic aspects of communication. Neus Figueras noted that when the CEFR was first published in 2001, it took a while to reach a common understanding of the descriptors and scales. The same is likely to be the case with the mediation descriptors: our understanding will deepen as we read them, go beyond the label “mediation”, and discover other meanings. What is involved in mediation is already taking place, but the scales help us to refine our operationalization

Barry O’Sullivan pointed out that in a standardized test like IELTS we have to avoid local knowledge to ensure the test is fair, whereas local knowledge can be included in a localized test. The mediation scales might be the beginning of theoretical support for localization. Pluriculturalism means different things in different contexts, and it means different things to different people. You can have pluriculturalism in a monolingual community. Joe Sheils agreed that cultures are diverse in themselves; he thought Mike Byram might even argue that culture should come before language.

Barry O’Sullivan argued that most of what is in the mediation scales is already included in tests, but it is not called mediation. If you look at any one descriptor, you’ll find that you can’t use it to create a task; you have to consider how it is linked to descriptors in other scales. What we have lacked is reflection on rating scales, which don’t take account of mediation. Mediation is common in L1/L2 testing; students who have Estonian and Russian do it in school exams.

Courses in English for international healthcare students address a plurilingual/pluricultural context. We should not see things exclusively from the perspective of language learning but should also take account of language use. We learn content with language, and mediation is still required even when you’re a native speaker (e.g. studying medicine). Agreeing with this point, Joe Sheils cited the example of migrants who fail to pass citizenship tests because they lack knowledge not of the language but of the host society and its culture.

One participant argued that teachers can’t teach everything. The CEFR is growing and it is hard for teachers to keep up. Who is going to mediate it to teachers? There are too many scales, they involve teaching/testing more than language, and they threaten to overwhelm teachers in much the same way as the requirement to write endless instructional objectives overwhelmed teachers in the US. Another participant disagreed, arguing that the mediation scales give us more information. The additional scales are useful, but it’s necessary to be selective, choosing the scales that are relevant to your context

Dina Tsagari wondered whether we want teachers simply to be aware of the mediation scales or to enable students to carry out the tasks contained in the descriptors. The mediation and plurilingual scales pose a challenge to language testers. Perhaps they should report on a student’s profile in three dimensions: proficiency, mediation and plurilingualism.

The question was raised: How context-specific are the descriptors? A B2 descriptor refers to the ability to “work collaboratively with people from other backgrounds”, but what if students are

able to work with people from some backgrounds but not from others? It was agreed that it's important to know how to interpret the descriptors in different contexts.

Barry O'Sullivan argued that mediation isn't just for L2 learners, it's for all language *users*. When we look at tests of culture, e.g. the UK and US citizenship tests, their focus is all wrong. They include questions about TV programmes that native speakers couldn't answer if they don't watch the programmes in question. We should be looking instead at the social norms within a specific country, and the operationalization of these scales would be different in different countries. Can we export the social/cultural norms that support them? Maybe this is a question that can be answered only on the basis of research.

Group 3: The plurilingual and pluricultural scales – rapporteur: Kate Plenter-Vowles

The group began by considering the relevance of the “pluri” scales for language policy. Brian North said that they were designed to bring together foreign language concerns and minority languages, linguistic identity and related political issues. He noted that the concept of translanguaging, mostly developed in relation to Hispanics learning English in the United States, has become very fashionable. Linked to identity politics, it's fundamentally a limited notion; the concepts of plurilingualism and pluriculturalism are more open, and the scales are intended to be used flexibly, taking account of the characteristics of specific contexts.

Jamie Dunlea wondered whether it's necessary to be proactive, stating from the outset that there is no intention to use the scales for testing purposes. Another participant thought it might be possible to test plurilingual capacities in certain contexts, acknowledging that people have different competences in different contexts. Jamie Dunlea linked this suggestion to Barry O'Sullivan's argument that the “pluri” scales imply localization.

In France, in the aviation context, teachers of languages other than English are happy to see themselves as promoting plurilingualism.

Brian North pointed out that the 2001 CEFR identifies plurilingualism as a key goal for language education. One reason for developing the “pluri” scales was to move the concepts of plurilingualism and pluriculturalism into the mainstream.

Vincent Folny argued in favour of promoting the learning of a third language at school; neuroscientists acknowledge that plurilingualism is good for the development of cognitive skills. Brian North agreed: there has to be a policy of “mother tongue plus two”, otherwise English will be the only L2 taught in most education systems. An Irish participant noted that in Ireland there's a tendency to think that English is enough, but those who learn Spanish or French gain access to a whole culture, a richness they would otherwise lack.

Brian North suggested that Brexit implies that Britain has not accepted foreign language learning and is reinforcing a monolingual culture. Jamie Dunlea noted that the Foreign Office does not exploit the fact that many people living in the UK speak languages other than English. Another participant pointed out that many senior teachers, academics and policy makers are not proficient in a foreign language. Their attitude tends to be: “I never took another language. Why would you want to?” They cannot be relied on to promote language learning and plurilingualism. A participant from Spain said that in Andalusia “mother tongue plus two” is now official policy; a participant from Croatia reported that the same is true there.

Brian North argued that learning foreign languages other than English is useful in real life and might be promoted using the argument that one language can be a gateway to several others. It's possible, for example, to buy a ticket from a Danish website using your knowledge of English or German; and if you're learning French, you should also be able to understand a bit of Spanish or Italian.

Jamie Dunlea thought that it might be unrealistic to emphasize plurilingualism outside Europe; he doubted whether it would find much resonance in Asia. Another participant repeated the point that people think they don't need to learn another language because everyone speaks English, as

seems to be the case, for example, in Holland. A third participant pointed out that if English is the only language people learn, they miss an opportunity to gain access to other cultures and broaden their mindset. Jamie Dunlea suggested that the “pluri” scales will be useful for promoting foreign language learning while still valuing migrants’ linguistic abilities in their home languages.

Referring to the slide in Vincent Folny’s presentation that showed a decline in plurilingual abilities across Europe, Brian North noted that in some areas of Switzerland people are becoming less bilingual than their parents; nowadays young Swiss people speak English rather than another of the languages of their country.

Vincent Folny observed that many French citizens are not able to read in their language of origin, usually an African language or a variety of Arabic. There are many multilingual communities with lots of language skills, so we must ask: Is our society doing the right thing with language, or are we missing something in terms of policy?

A participant pointed out that instead of admiring Macron’s plurilingual abilities, people in France sometimes criticize him for not speaking French.

Scott Stroud asked whether the group had anything to say about pluricultural skills. His language centre, for example, places a lot of emphasis on intercultural communication; students are taught languages, but they are also taught how to communicate in those languages. Then students object: “I’m an engineer learning English, why do I have to learn intercultural skills?” Another participant objected that it’s impossible to separate language and culture. Scott Stroud wondered how sensitive learners of Chinese were likely to be to Chinese culture; in his view the pluricultural scales will encourage teachers of Chinese to address this issue. Another participant pointed out that people tend not to see that culture and language are interconnected.

Brian North agreed that the pluricultural scale is intended to address this issue, though care was taken to avoid the highly contentious term “intercultural competence”; the descriptors refer rather to intercultural encounters. The concept of pluriculturalism has to do with openness, being able to engage with people from all over the world.

Another participant noted that policy and curricula tend to link language and culture and the need to promote awareness of the “otherness” that was mentioned in an earlier session. Language and culture also tend to be linked in teacher education for elementary schools.

Vincent Folny said that when culture is discussed in France, there’s a tendency to think that everyone is different. Brian North added that this is the issue with the concept of intercultural competence: we immediately start talking about differences. Scott Stroud said that his students are told to look for similarities, and Jamie Dunlea agreed that it’s important to look for common ground: by focussing on differences we build walls. Brian North explained that the pluricultural descriptors seek to focus on positives and common ground, playing down misunderstanding and differences.

Recalling that in his presentation he had suggested that companies would be interested in the “pluri” scales, Vincent Folny said that their interest would be driven by economic considerations: they understand how important it is to understand the cultural factors that influence markets.

Scott Stroud noted that in his experience if one person at a company meeting speaks English, the meeting will be in English; and if a meeting comprises mostly French and a few German speakers, the meeting will be in French. Sometimes in multilingual business groups, each participant speaks his or her own language and everyone understands. Brian North pointed out that such behaviour is an illustration of what the CEFR understands by “partial competence”.

Jamie Dunlea asked about language policy in Ireland. An Irish participant responded that Irish is the main language according to the government, but then in secondary school the parents say, “Ah, don’t bother”. But learning Irish does help students to learn another language.

There was a brief discussion about the extent to which it’s possible to present one’s personality through a foreign language: Are you the same person when you speak another language, or does it change your personality? Many people in Japan believe that learning English will change their personality, so that language learning becomes a matter of personal growth.

It was suggested that in order to teach a language, it's necessary to be able to speak another language besides your mother tongue so that you have some understanding of the language learning process.

There was some discussion of the different policies countries and regions adopt to broadcasting foreign films: some insist on dubbing soundtracks into their own language, while others are content to use subtitles.

Brian North argued that plurilingualism is both an educational and a societal goal; it has to be assessed in some way otherwise students won't take it seriously, but it shouldn't be an assessment goal. Jamie Dunlea agreed, emphasizing that it's important to promote plurilingualism in the right way: not testing for plurilingualism itself, but using the idea of plurilingualism to help with assessment.

Summing up, Scott Stroud described plurilingualism as an umbrella ability to include in discussion of language policy and a way of defending languages that are disappearing. The "pluri" scales are not intended to be used to test everyone but rather as a way of promoting plurilingualism. How this is done will depend on local circumstances.

Session 5

Reports from the discussion groups

Group 1: Update of 2001 scales – rapporteur: Doris Froetscher

The group discussed what aspects of the updates to the 2001 descriptors and scales were most useful; several participants were fans of the pre-A1 and phonology scales. It was suggested that there are not enough descriptors for A2+ (John de Jong pointed out that additional descriptors are available from Pearson at <http://www.english.com>). The usefulness of the new descriptors for classroom-based assessment was noted, but it was pointed out that language use involves more than the descriptors capture. The “pluri” scales were thought to be particularly helpful in contexts where two languages are in use. The group considered whether the mediation and “pluri” descriptors are useful for testing, concluding that it depends on the context and function of the test. The group also recalled that the CEFR serves multiple functions, in relation to policy, curriculum, teaching and testing. It was agreed that teachers need more training in the use of the CEFR to help them grasp the underlying principles. It would be interesting to get in touch with teacher education institutions and associations to find out how well the profession knows the CEFR.

Group 2: Mediation – rapporteur: Bárbara Eizaga

The group agreed that research is needed to find out what features of mediation could be used in the classroom; also, what is and what should be included in teaching materials. Should teachers as well as learners be assessed in mediation? Teachers mediate all the time, but usually they don't teach learners to mediate. It will take some time for people to understand what “mediation” means, including the headings of some scales; theoretical support would be a welcome addition to the CV. Mediation might be used as a way to foster plurilingualism, especially as the mediation scales have a strong role to play in learning. The number of CEFR scales is growing, and this prompts the question: Who is will mediate between teachers and CEFR descriptors? Some doubts were expressed about whether/how to use the mediation scales in assessment. Mediation cannot be assessed in a generic way, so localization will be crucial; it's essential to consider how descriptors are related to specific cultural contexts.

Group 3: The plurilingual and pluricultural scales – rapporteur: Scott Stroud

A recurrent feature of the discussion was the belief that the plurilingual scales are a way of making people aware that languages are being lost and of promoting the teaching of multiple languages. The existence of the scales doesn't mean that they should be used for testing; they are probably best used in a local situation. Most countries need a policy of “mother tongue plus two other languages”; this is happening in some regions of Spain. The group discussed how we can develop people's ability to learn other languages. If you can speak several languages, you don't need a high level in all of them; much can be achieved by partial competences. Intercomprehension also helps, for example between Spanish and Italian, or Dutch and German. In Switzerland more and more people are learning and speaking English. Some countries, e.g. France and Spain, still tend to dub films, while others use subtitles, which helps to make more languages available. Andalusia has recently invested heavily in broadening its television programmes. The dominance of English is now such that the teaching of other languages is in danger of disappearing from schools and universities. There was not much discussion of the pluricultural scales, though it was noted that learners should be encouraged to focus on similarities as well as differences.

Close of workshop

Bringing the workshop to a close, Neus Figueras thanked the presenters, respondents and participants for a stimulating day that provides the SIG with a useful starting point for further work. Lorna Carson thanked EALTA for holding the meeting in Trinity College, Neus Figueras for all her preparatory work, the student volunteers for providing essential organizational support through the day, and the British Council for its sponsorship.

Appendix 2: List of participants

Elena Arenales	ele.filologia@gmail.com
Christopher Bacon	baconc@tcd.ie
John Baird	johnbaird@rcsi.ie
Rosemary Benard	rosemary.benard@enac.fr
Armin Berger	armin.berger@univie.ac.at
Suzana Bitenc Peharc	suzana.bitenc@ric.si
Eveline Boers-Visker	eveline.boers@hu.nl
Manuela Calzini	manuela.kelly.calzini@trinitycollege.ie
Lorna Carson	carsonlorna@gmail.com
Nicola Carty	nicola.carty@glasgow.ac.uk
Bronagh Ćatibušić	bronaghcatibusic@gmail.com
Daniel Chavez	dchavez@tcd.ie
Joaquín Cruz	joaquin.cruz@ujaen.es
Lidija Cvikic	lidija.cvikic@ufzg.hr
John de Jong	jong70@kpnmail.nl
Maria Dimitropoulou	dimitrom@tcd.ie
Jamie Dunlea	Jamie.Dunlea@britishcouncil.org
Bárbara Eizaga	barbara.eizaga@uca.es
Yanhong Fei	feiy@tcd.ie
Jim Ferguson	jcferguson33@gmail.com
Ana Fernández Torres	aftorres.eoi@gmail.com
Neus Figueras	nfiguera@xtec.cat
Colin Flynn	colin.flynn@dcu.ie
Vincent Folny	Folny@ciep.fr
José Francisco Frías	jffrias.eoi@gmail.com
Doris Froetscher	doris.froetscher@bmb.gv.at
Rita Green	r.green2@lancaster.ac.uk
Tony Green	Tony.Green@beds.ac.uk
Susan Hackett	shackett@qqi.ie
Kristof Hegedus	kristof@euroexam.org
Tingting Hu	huti@tcd.ie
R. J. M. Joynt	r1171156@gmail.com
Elif Kantarcioglu	kutevu@bilkent.edu.tr
Krista Kerge	krista.kerge@gmail.com
Elif Kir	KIRE@tcd.ie
Stergiani Kostopoulou	stergiani.kostopoulou@ucd.ie

Ermioni Kounadi	kounadie@tcd.ie
David Levey	david.levey@uca.es
David Little	dlittle@tcd.ie
Ronan McGuinness	ronan.mcguinness@macmillan.com
Conor McNally	clpmcn@gmail.com
Olga Melnychuk-Magee	melnycho@tcd.ie
Josep Maria Montcada	jmoncada@xtec.cat
Tereza Mytakou	terezaaa.mmm@gmail.com
Brian North	bjnorth@eurocentres.com
Anna Nunan	anna.nunan@ucd.ie
Barry O'Sullivan	Barry.O'Sullivan@britishcouncil.org
Spiros Papageorgiou	spapageorgiou@ets.org
Cristina Perez-Guillot	mcperez@upv.es
John Pill	j.pill@lancaster.ac.uk
Kate Plenter-Vowles	k.plenter-vowles@rug.nl
Oliver Ramos Miranda	oliver.ramos@gmail.com
Laura Riera Grau	Laura.Riera@uab.cat
Noelia Ruiz Madrid	madrid@uji.es
Alex Runchman	alexander.runchman@ucd.ie
Joseph Sheils	joseph.sheils@gmail.com
Dilys Silva	dilys.silva@macmillan.com
Oscar Soler Canela	oscar.soca@gmail.com
Richard Spiby	richard.spiby@britishcouncil.org
Scott Stroud	scott.stroud@enac.fr
Cathy Taylor	cathy.taylor@trinitycollege.com
Dina Tsagari	Dina.Tsagari@hioa.no
Ülle Türk	ulle.turk@ut.ee
Beppie van den Bogaerde	beppie.vandenbogaerde@hu.nl
Odette Vassallo	odette.vassallo@um.edu.mt
Carolina Williams McDonnell	cwilliamsmcdonnell@gmail.com
Elisabeth Wippel	elisabeth.wippel@bmb.gv.at
Norman Wright	normaleewright2016@gmail.com
Ying Wu	yiwu@tcd.ie
Natalie Xie	Xiesenat@gmail.com
Julia Zabala	juzadel@upv.es