Trolls, unicorns and the CEFR: Precision and professionalism in criticism of the CEFR

Brian North

(co-author of the CEFR and the CEFR Companion Volume)

This article starts by recalling the reasons that have been given for the CEFR's success, for example its neutrality, the way it encourages the constructive alignment of planning, teaching and assessment and helps educators to fulfil both instrumental and educational goals. It then reviews and responds to some of the main criticisms that have been made of the CEFR over the past twenty years concerning the relationship of the CEFR to linguistic theory, the compatibility of the CEFR descriptors with research in second language acquisition and corpus linguistics, the development methodology and formulation style of the descriptors, the intended scope of the CEFR itself and its relationship to socio-political power. It points out that many of these criticisms are based on misunderstandings or misrepresentations and underlines that a sustained constructive engagement with the CEFR is necessary if criticism is to inform future revisions. The article also draws attention to some of the innovations brought by the CEFR, which have tended to be overlooked, and which are reinforced and further developed in the recently published update to the CEFR, the CEFR/CV, which has just in its definitive form.

1 Introduction

The publication of the definitive version of the CEFR Companion Volume (Council of Europe 2020) is perhaps a moment to consider the criticisms that have been made of the CEFR over the years. This is particularly the case since, in reviewing the Companion Volume, authors seem compelled to repeat what have become standard criticisms (e.g., Bärenfanger, Harsch, Tesch and Vogt 2018; Deygers 2019; Quetz and Rossa 2019), but do not take account of answers to them (e.g., North 2008, 2014), and sometimes misrepresent the point made by a more discerning previous critic. Authors sometimes assume they know the CEFR and the criticisms of it, but misrepresent either or both when they write themselves. In fact, the CEFR seems to invite a kind of familiarity that sometimes leads to careless assumptions (e.g., "As is commonly known, the framework distinguishes five proficiencies (speaking, listening, reading, writing, and interaction) and describes six levels of these proficiencies with regard to one language" Backus et al. 2013: 191) or article titles that are, to say the least, unusual (e.g., "One framework to unite them all?[1]" Deygers et al. 2018).

The CEFR is published by the Council of Europe (CoE), whose remit is the promotion and protection of human rights and social justice. The CEFR is in fact the CoE's second most consulted document, coming on the list directly after the Declaration of Human Rights itself[2]. The CEFR was produced as part

---

1. This title echoes the "one ring to unite them all," the ring forged by Sauron, the personification of evil, in J.R.R. Tolkien's trilogy *The Lord of the Rings*. That title inspired the title of this current article.
2. 2018 CoE web statistics.

of a project to develop European citizenship and is part of a sustained commitment to promote quality inclusive education for all, particularly plurilingual and intercultural education. The significance of the CEFR for curriculum and assessment has been widely recognized both within and beyond Europe. It has been described in a state-of-the-art article on language curriculum as "[o]ne of the most important curriculum publications in the last decade" (Graves 2008: 148) and "[p]erhaps the most widespread example of backward design using standards [working backwards from goals defined with 'Can Do' descriptors] (Richards 2013: 26). Several surveys of the implementation of the CEFR in different countries are available (e.g., Byram and Parmenter 2012; Foley 2019; O'Dwyer et al. 2016; Piccardo, Germain-Rutherford and Clement 2011). Byram and Parmenter's edited volume documents some reasons for the success of the CEFR: the positiveness and clarity of the 'Can Do' recognition of modest achievement and related promotion of self-assessment; the extra-national, neutral non-prescriptiveness of the scheme; and the fact that it addresses both instrumental/functional and humanistic/educational aims of language learning. As Porto, one of their contributors, explains in more detail, the CEFR helps language policy makers to marry, in their local educational standards, (a) the needs of their governments to promote instrumental functional goals in English, the language of international communication and business, with (b) broader goals that she describes as: "Progressive Education, the main tenets of which are education for active citizenship, for social justice and for the protection of local languages, celebrating the students' interests and participation" (Porto 2012: 135).

Fundamentally, the CEFR offers the means to align planning, teaching and assessment and involve all stakeholders in what is effectively a quality cycle of 'plan, do, check, reflect and act' at the levels of the individual, the class, the programme, and the institution. A recent project from the ECML (European Centre for Modern Languages), CEFR QualiMatrix (www.ecml.at/CEFRqualimatrix), provides a practical online self-evaluation tool to assist in the planning or evaluation of CEFR-based innovation. It also provides some 35 examples of CEFR-based best practice in different contexts as illustrations of such innovation. In fact, of the two main aims of the CEFR, (a) to provide common reference points and a metalanguage to help language professionals situate their efforts, network, and compare, and (b) to stimulate educational innovation and more effective language learning, the second aim has always been predominant. This was confirmed again by the 47 member states at the Language Policy Forum called to take stock regarding the CEFR (CoE 2007). This aim is the reason the CEFR 2001 was set out as a thesaurus, inviting users to review and perhaps consider developing their current practice in the light of other options, with 'reflection boxes' at the end of each section to help them to do so. The CEFR is a reference work not a standard to be picked up and applied. The authors made this very clear in the foreword: "We have NOT set out to tell practitioners what to do or how to do it. We are raising questions not answering them. It is not the function of [the CEFR] to lay down the objectives that users should pursue or the methods they should employ" (CoE 2001: iv). The Companion Volume explains why the CEFR descriptors are consistently described as 'illustrative.' They are meant to be adapted to context and supplemented; North (2014) illustrates some of the ways in which this can be done.

In this article, I therefore discuss what are perhaps the six main misunderstandings concerning the CEFR and its descriptors. These are namely: (a) the relationship to theory; (b) the relationship to research on learner language; (c) the methodology through which the descriptors were developed; (d) the formulation of the descriptors; (e) the intended scope of the descriptors and indeed the CEFR itself, and finally (f) the status of the CEFR and its relationship to socio-political power.

2 The relationship to theory

There is sometimes an assumption that the CEFR has no theoretical framework. In fact, the theoretical framework was laid out in considerable detail in North (2000) and related publications (e.g., North 1997a). The CEFR move from the four skills to the four modes of communication (reception, production, interaction, mediation) was inspired by a series of criticisms of the inadequacy of the four skills model (Lado 1961) to describe actual language use (e.g., Alderson and Urquhart 1984; Breen and Candlin 1980;

Brumfit 1984; Stern 1983). It was also influenced by Halliday's (1989) precisions on the true distinction between spoken and written language; Swales' (1990) analysis that all genres derive from chat (*interaction*) and then storytelling (*production*), which created the reciprocal mode of *reception*; insights about long and short turns spoken by young people (Brown et al. 1984); and the distinction between basic interpersonal communication and more academic language (Cummins 1980). The model of communicative language competence is closely related to Bachman and Palmer (1996) and Celce-Murcía, Dörnyei and Thurrell (1995); it rejected Chomsky's competence-performance distinction and reflected the interpretation of competence in communication studies (e.g., Wieland and Backlund, 1980) and the world of work (see Richer 2017). The CEFR envisages a strategic cycle of planning, execution, evaluation and repair following Færch and Kasper (1983), with some categories for strategies that were inspired by: Tarone (1983) on interaction strategies; Barnes and Todd (1977) on cognitive and collaborative strategies in small groups; Kramsch (1986) on turn-taking, and Burton (1980) on "challenging" for clarification.

The theoretical model behind the updating of the CEFR descriptive scheme in the CEFR Companion Volume (CEFR/CV) is explained in North and Piccardo (2016) and Piccardo and North (2019). Many key aspects that were already implicit in the CEFR 2001, though not developed, are made explicit. These 'hidden aspects' of the CEFR include a Vygotskyan recognition of the social origin of learning and a focus on agency, with the learner seen as a *social agent*; an *action-oriented approach* implying collaborative, situated co-construction and learning (e.g., 2001 descriptors were provided for *goal-oriented collaboration* and *cooperating*); the introduction of the concept of *mediation*; and a detailed exposition of *plurilingualism,* which anticipated what is sometimes referred to in Anglophone literature as 'the multilingual turn' (Conteh and Meier 2014; May 2013). Many researchers and practitioners have welcomed the clarifications and further development in the Companion Volume (see, for example the report on the meeting called by EALTA[3]: Little (2018). Yet Deygers (2019: 3) considers that "since mediation must logically include at least two other communicative activities to take place, it is conceptually superfluous (Wittgenstein 1922 on Ockham's razor)". However, one could of course say the same about interaction. In fact this linear, Cartesian perspective—that the pieces make up the whole—which still held sway in 1960's structural linguistics, the context in which Lado (1961) proposed the four skills, is simply not tenable given the complex, ecological, paradigm in which we work today (see Larsen-Freeman 2011, Van Lier 2010).

3 The relationship to research on learner language

Let us now turn to the second, empirical, aspect of the relationship to research, which concerns the descriptors. At a recent colloquium in Gießen, Reimer (2019) repeated a common assumption that the progression shown in the descriptors was incompatible with SLA research, citing Hulstijn (2007) and Wisniewski (2017). In fact, Hulstijn (2007) said that there was no need to abandon the CEFR 'house' whilst secure (SLA-based) foundations were built and co-founded SLATE (eurosla.com) to provide that underpinning. In the first volume reporting SLATE results, Hulstijn, Alderson and Schoonen then wrote that "[t]he production of the scales was … an extensive empirical exercise … It is fair to say that the resultant scales are probably the best researched scales of foreign language in the world" (2010: 14-15). One might add that the work of the SLATE group and others has tended to confirm the progression in the CEFR scales. Findings seem to confirm CEFR suggestions that control of grammatical accuracy becomes a feature around B2 (e.g., Díez Belmar 2018; Forsberg and Bartning 2010; Martin, Mustonen, Reiman and Seilonen 2010; Thewissen 2013); Tono 2013), that vocabulary range increases steadily through the levels (Milton 2010), and that explicit markers for cohesion/coherence increase to B2 and then are substituted by more subtle means at the C-levels (Carlsen 2010).

Most of this research is actually CL research. The largest such projects are English Profile and the related Cambridge Learner Corpus (Harrison and Barker 2015) and the CEFR-J/JEFLL corpus (Tono 2013) and their reports do not display any particular contradictions with the CEFR. These studies are

---

3.   European Association for Language Testing and Assessment: www.ealta.eu.org

supplemented by other CL projects such as the Greek Integrated Foreign Language Curriculum (IFLC) project (Dendrinos and Gotsouilia 2015) and smaller scale work by Díez Belmar (2018), concerned with defining the errors of Spanish learners of English. Both these projects supplement the rather generic CEFR descriptors with data-based locally relevant detail. Finally, there is the work of Wisniewski (2017) with regard to German, to which Reimer referred. Wisniewski found what she states to be problems with the Vocabulary Control Scale and Fluency Scale in her data—though she says: "The fluency scale generally led to more convincing results than the vocabulary scales (Wisniewski 2017: 242). However, this data was from a single test task with a corpus based on only 38 learners.   In larger scale work (258 scripts in the MERLIN project) she points out possible weaknesses of the B2 descriptor on the Vocabulary Control Scale but concedes that it "captures observable, yet not exclusively typical behaviour" (Wisniewski 2017). She reminds the reader that: "The CEFR levels are not claimed to correspond to a developmental hierarchy in an SLA sense, either. All this is clearly stated in the CEFR itself and in pertaining publications (North 2000, 2014)" (Wisniewski 2017: 245).

In other words, this criticism about the lack of a basis in SLA/CL for the descriptors actually represents a caveat not a fault, and with this one possible exception, such research as exists actually supports the progression suggested. Furthermore, the range of SLA and CL research is very limited: both are concerned with linguistic features (predominantly grammar and vocabulary), often described as 'critical features' that distinguish between levels through their presence and the degree of accuracy in using them. Thus, SLA and CL research could in any case only inform refinement of the 13 scales for communicative language competences (aspects of linguistic, sociolinguistic and pragmatic competence). SLA and CL research is little or no help in relation to the vast majority of the scales, which are for communicative language activities and strategies (c 40 in 2001; c 65 in 2018/2020). I have always been very open about the fact that SLA research could not provide an adequate basis for the CEFR; I doubt that it ever will. After all, I made this point in first presenting the research and descriptors (North 1997b) and, as Wisniewski says, have repeated it constantly since: for example: "What is described [in the descriptors] is teachers' perceptions of language proficiency (appropriate for a common framework of reference), not validated descriptions of SLA processes …" (North 2007: 657). Unfortunately, misinterpretations of that 2007 statement by writers less careful than Hulstijn or Wisniewski, suggesting that the calibration is based upon teacher impressions and lacks a basis in empirical research, have unfortunately been passed on from article to article (or presentation). The CEFR descriptors, new and old, are in fact based upon a rigorous research methodology that captured and objectified collective professional wisdom, which brings us to the next point.

4 The Development Methodology

Despite the fact that the original research is described in the CEFR itself (Appendix B) and published in articles (e.g., North 1995; North and Schneider 1998) and books (North 2000; Schneider and North 2000), there are some remarkably persistent misconceptions. Firstly, as Alderson and Hulstjin (2010) pointed out, the 1993-96 work was based on the decades of experience in the profession with language proficiency scales. It was only descriptors for communicative language strategies that were written from scratch. Secondly, just because the descriptors were subjected to a rigorous validation and calibration process with the Rasch model, there is in some quarters a curious perception that the design of the scales was conceptually random, with the decision as to which of the initial pool of 2,000 descriptors should survive being made purely on the basis of statistical data. In fact, in both years of the 1993-6 Swiss National Science Research Council project that produced the 2001 descriptors, the *intuitive* development phase of just over a year was followed by a lengthy *qualitative* validation phase. This involved 32 workshops with teachers over the course of a second year in both the 1994 project for English and the 1995 follow-up for English, French and German. In this phase in both 1994 and 1995, teachers evaluated and suggested improvements to the descriptors in the initial pool. They were asked to identify which category descriptors belonged to and whether they were clear, pedagogically useful and related to real world language use. This methodology was later used by Eichelmann (2015) and Vogt

(2011) in CEFR-related projects, Eichelmann systematising it into a form in which it was then reused in the 2015-2016 data collection in the project to extend the CEFR descriptors (North and Piccardo 2016). On each occasion, 1994, 1995 and 2015-2016, qualitative data from workshops was used to whittle down a huge initial pool of potential descriptors to a set of really good descriptors that covered the intended categories across the intended levels (approx. 300 in 1994 and in 1995; 426 in 2015).

It was those—already validated—descriptors that were forwarded to the third, *quantitative* phase in what nowadays would be described as a sequential, mixed-methods research design (Creswell and Plano Clark 2018). Not only that, but in all the workshops of the qualitative phase in 1994 and 1995, the categories of the metalanguage used by teachers to discuss proficiency were studied using grounded theory (Byrant 2014). In each of the workshops, the discussions of pairs or small groups of teachers were recorded separately. That bottom-up analysis complemented the top-down analysis in the CEFR Authoring Group[4] in a process that decided the CEFR descriptive categories.

A related misunderstanding (repeated in, for example, Wisniewski 2017) is that in the development there was no relationship to actual learners and their language, the data to calibrate the descriptors coming from a task in which teachers sorted the descriptors into piles by level. Such a task is in fact a traditional method to derive scale difficulty values for descriptors, further developed by Smith and Kendall (1963) in an early example of data-based scale development. Such a sorting task was in fact only used in the final workshop in 1994 and in 1995 in order to (a) eliminate descriptors on which there was wide disagreement regarding the level of difficulty, and (b) check that descriptors were approximately the level intended so that they appeared on an appropriate data collection questionnaire. These questionnaires were to be used for classes of different school years in secondary education because one of the official aims of the project was to provide a snapshot of the range of achievement at the end of each year of the different sectors of the Swiss secondary school system. The results provided data as a basis for a proposal for national targets (Lüdi 1999). (For a graphic showing those results, see North 2000a: 319-33; Schneider and North 2000: 321).

The 50-item questionnaires used to collect the data for calibrating the descriptors were actually the precursors for the checklists of the European Language Portfolio, which the project was also charged with developing. On the questionnaires, teachers rated a structured sample from two of their classes. The same 0-4 rating scale was used to assess the extent to which each learner could do what was described in each descriptor. Three thousand five hundred learners were assessed with the descriptors by their teachers in this way (1,000 in 1995; 2,500 in 1995) and it is that—teacher assessment—data that was analysed with the Rasch measurement model to calibrate the descriptors.

A very similar approach was used on a larger scale in 2014-2017 for the development of the new scales on mediation and related areas for the CEFR/CV. The main data collection took place in 2015, following a year of preparation. This time there were 140 workshops in which approximately 1,000 informants, working in pairs, were involved in seeing whether the descriptors fit the category they were said to describe, as well as evaluating their clarity, pedagogical usefulness and relation to the real-world language use—as well as the actual formulation of each descriptor. In a second series of 189 workshops, some 1,300 informants, again in pairs, discussed the descriptors and judged their CEFR level. Finally, the definitive calibration came from an online survey conducted in English and French in which participants used the 0-4 rating scale from 1994/1995 to assess whether a person they were thinking of could do what was described in the descriptor. The process is described briefly in an appendix to the CEFR/CV and in more detail by North and Piccardo (2016). The descriptors took their more or less definitive formulation only at the end of that whole process. There then followed a process of slimming down the number of descriptors, and some final polishing during the consultation phases, which lasted up till May 2019. The last steps, in 2019, were to make the formulations 'modality-inclusive (i.e., also suitable for sign languages) and, as far as possible, gender neutral.

---

4. The CEFR Authoring Group consisted of John Trim, Daniel Coste, and Brian North; Joe Sheils, the project coordinator from the Council of Europe secretariat, joined the group later on by writing Chapter 7 on tasks. The revision for publication in 2001 was carried out by John Trim and Brian North.

5 Descriptor formulation

Another misunderstanding about the descriptors is that there is no systematic development in the progression up the scales and that the content found at different levels is somehow random or 'subjective,' because the same aspects are not treated systematically at every level. This criticism was addressed by North (2008) using the same subscale, *Understanding an interlocutor*, that had been criticised as an example by Alderson et al. (2006). A number of charts were also included as appendices in the manual for relating examinations to the CEFR to help readers see this systematicity and a selection of other charts are included in North (2014). Essentially there are two points behind this criticism, apart from the question of mere presentation. Firstly, there is a misunderstanding of the deliberate choice made in developing the CEFR descriptor scales—which are *not* rating scales, but rather curriculum orientation aids. Secondly, some language testers have an unrealistic expectation for detail and precision that is not appropriate in a common framework, intended to be used for different educational contexts in relation to different languages. The opportunities, challenges and limitations of what a common framework can provide for linking assessments are well explained by Harsch (2019).

To focus on the first point, the descriptors belong to a tradition in applied psychology that defines, and then calibrates mathematically to a scale, target behaviours at ascending levels of difficulty, that is to say important learning aims. This approach was pioneered with trainee nurses (Smith and Kendall 1963). Each descriptor is an independent criterion statement, which illustrates a 'salient feature' of behaviour at a particular point—or band—on the scale. This approach came into language education through language for specific purposes (e.g., ELTDU, 1976). The alternative 'systematic' approach, often used in language testing, is to describe exactly the same features at each level. The distinctions between levels are then made by juggling with qualifiers like 'some' 'a few' 'many' 'the majority of' etc. This approach is still very common in even recently published rating scales in the language field, even though it has been heavily criticised for a long time (e.g., Champney 1941; Alderson 1991). Because the prime objective of the CEFR is to provide curriculum aims, the intergovernmental Symposium that recommended the CEFR (Council of Europe, 1992) unanimously rejected the 'systematic approach,' instructing that the 'salient features approach' should be used for both CEFR and European Language Portfolio, ensuring coherence between the two.

Tracy (2017), however, suggests that the CEFR descriptors take precisely this 'systematic' approach: making relative distinctions between levels just with adjectives and adverbials:

> Despite the remarkable career of the CEFR, there is room for improvement. Many 'Can Do' statements contain among their descriptors quantifying ('large', 'small', 'short', 'limited', etc.) or qualifying expressions ('relatively simple', 'elementary', 'complex'). Descriptors refer to vocabulary or other features the test-taker appears to be 'more' or 'less familiar' with, is 'more' or 'less likely to encounter', or to terms and tasks which are 'more or less related to everyday experience'. There is also reference to what interlocutors can 'easily' or 'partially' understand. (Tracy 2017: 49).

In actual fact, only three of the expressions she mentions are used extensively in the 2001 CEFR descriptors, namely: 'short' (54 occurrences), 'complex' (36 occurrences) and 'limited' (10 occurrences). There are precisely zero occurrences of 'small,' 'relatively simple,' 'more familiar,' 'more likely to encounter,' 'less likely to encounter,' 'tasks more or less related to everyday experience,' 'easily understand,' or 'partially understand.' One does find 'simple' (101 occurrences) as opposed to 'straightforward' (21 occurrences) and 'complex' (36 occurrences); one also finds 'familiar/unfamiliar' (61 together) as well as one occurrence of 'less familiar.' There is one 'large' and one 'elementary.'

In other words, there are binary distinctions made, and there is a simple/straightforward/complex distinction, but this claim is exaggerated. There is also the fact that what is simple for an eight-year-old and a twenty-eight-year-old are not the same thing. Texts that are straightforward for me in my profession may not be so for you, if you specialise in a different field. In other words, the CEFR descriptors

"are themselves simply guidelines and we are encouraged to adapt them and rewrite them according to the context in which we work. … [The CEFR] is a *framework*, not a set of stone tablets; it exists primarily to help language professionals and language learners achieve their goals more successfully, to help us to think about how and what we teach and learn" (Frost and O'Donnell 2015: 4). This fact does rather tend to get forgotten.

6 The Scope of the CEFR descriptors

This leads us to the next misunderstanding—or unrealistic expectation. The CEFR is *deliberately* open-ended. This is because it is intended to be used in a wide variety of different contexts: for different languages, for different age groups, for different types of learning goals, in different pedagogic traditions. It is a generic, common reference point. It is not a turnkey, off-the-shelf system. The array of descriptor scales (c. 50 in 2001; c. 80 now) is intended to suggest selection, needs analysis. It is unlikely that all of the descriptor categories are priorities for any one group. The CEFR/CV reminds readers how scales can be used to create needs profiles for different groups and two (fictional) examples of graphic profiles are given (CoE 2018: 37-38, 2020: 38-39).

If the CEFR provided all the details on language exponents and text types, etc., for all the languages one might want to teach, it would become prescriptive—as well as a gigantic instrument. That detail is provided separately in the reference level descriptions (RLDs)[5] for the different languages. If the CEFR provided descriptor scales for each target situation/genre of each aspect of each of the four domains of language use (public, private, vocational, educational) it would again be in danger of becoming prescriptive. Therefore, the approach taken is generic, macrofunctional (see CoE 2018: 3-31; CoE 2020: 33-34). Users are invited to adapt descriptors and elaborate new ones that will fit the needs in their context. This adaption may mean adding linguistic detail that takes account of the educational context and the learners' linguistic repertoires (Díez Belmar 2018; Dendrinos and Gotsouilia 2015). It may mean adapting the descriptors themselves to a different age group—as done with the descriptors collated in Szabo and Goodier (2018).[6] It may mean adapting and/or developing descriptors for a particular academic or professional context; North (2014, Section 4.2.3) gives tips for doing so. It may mean analysing descriptors in order to specify text types, text features and microskills for listening and reading tests in a particular context (North and Jarocz 2013), or going a step further to develop and validate local listening or reading tests like for example Shackleton (2018). Shackleton developed tests with CEFR-based specifications and then followed the procedures to link scores on the test to CEFR levels recommended by the Council of Europe. All these are examples of sensible adaptation and extension of a common reference framework to the local context. None of these researchers expected the CEFR to be targeted specifically to their context.

One of the more surprising criticisms of the CEFR, therefore, is one by McNamara, Janne Morton, Storch and Thompson (2018) who talk about "the poverty of the CEFR construct for the assessment of EAP [English for Academic Purposes] readiness and progress" (McNamara et al. 2018: 17). They report that other scholars have been critical of the conceptualization of academic writing—in the CEFR and, they add, in tests specifically developed for academic writing in English like IELTS and TOEFL—as an "autonomous set of skills that once mastered can be used across contexts. Academic writing, from an academic literacies perspective, is fundamentally situated in particular disciplinary cultures …" (McNamara et al. 2018: 18) with at times even "variation between teachers within a discipline". In their project, they compare the construct in the three CEFR scales for written production to the perception of 13 first-year international students regarding their academic writing. They record the way that experience

---

5. https://www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions
6. https://www.coe.int/en/web/common-european-framework-reference-languages/bank-of-supplementary-descriptors

over the year of the demands of different disciplines and tutors led the students to move from that traditional 'set of skills view' of EAP writing to an appreciation that it is thoroughly situated. However, rather than perhaps criticising the tests created for their own discipline, which also follow that view, the authors focus on the CEFR. This is despite the fact that they concede that the process of studying the CEFR scales and undertaking this research, "ironically" helped them to reflect on their current practice and "will contribute to the ongoing fine-tuning of our Table of EAP competencies" (McNamara et al. 2018: 25). This process of reflection on current practice is actually, as they even mention themselves, precisely what the CEFR is intended to encourage.

This is not to say that CEFR descriptors are not useful in the context of teaching English at university level. Frost and O'Donnell (2015) document using them successfully to involve students in the process of tracking their progress in spoken production over the course of their three years of study, using the descriptors in adapted form for teacher, peer and self-assessment. Idris and Raof (2017) also report on learners using CEFR Table 3 (the six levels defined for range, accuracy, fluency, interaction, and coherence) for self- and peer assessment of spoken ability. Academic writing, however, with all its varying socioculturally-determined and genre-related expectations, obviously requires contextually-specific criteria for any assessment.

7 The CEFR, status and power

Criticisms like those of McNamara et al (2018) reflect misinterpretations of the aims and status of the CEFR. McNamara himself (2011) perceives the CEFR as an instrument of power: a universal language—spreading in the same way that English is spreading as an international language. In fact, as we will see below, it is in fact the combination of those two trends—the appropriation by the spreading ELT industry of the CEFR levels—that is the problem.

As mentioned at the beginning of this paper, the CEFR itself has two fundamental aims: The first is to provide common points of reference for national education systems in Europe. The second, most fundamental aim is to assist reflection on current practice and stimulate development and reform, including the promotion of plurilingual and intercultural education. Talk of the "reification" of the descriptors scales into a rigid system imposed on unwilling professionals has proven to be unfounded: there has never been a "strong political agenda" to standardise the language of assessment across Europe as Fulcher (2004) opined. Fulcher (2008) presents the CEFR as a vehicle for centralised planning, the removal of academic freedoms and the introduction of personal financial accountability for teachers. Fulcher proposes, on the basis of no evidence whatsoever, that the Council of Europe—or perhaps he meant European Union—intends to enforce harmonisation that will stifle teacher creativity and that makes any resistance "genuinely futile" (2010: 230).

One sees here, in a specialist form, a forerunner of the kind of misunderstanding and misrepresentation of the European project that led to Brexit. In fact, the adoption of the CEFR by European governments is done by a policy recommendation (Council of Europe and Council of Ministers 2008), not a resolution, let alone a treaty. The European harmonisation project in education is in reality the Americanisation of higher education in the Bologna process. The real linking of teacher freedoms and salaries to results on standardised tests that operationalise a standard is an Anglo-American vogue unconnected to Europe, let alone the CEFR. Indeed, the CEFR provides the basis for an alternative to standardised tests from the language testing industry. Common reference points independent of the industry leaders give at least the possibility of diversity and context-relevance in assessment. In any case, there is little doubt that ALTE and EALTA have both substantially contributed to raising language assessment literacy in Europe, at least partly thanks to the CEFR.

Any use of the CEFR *outside* Europe itself does, however, raise the question of "validity creep" (North 2014: 44): validity in relation to the CEFR "is an ongoing and, theoretically never-ending, process" (Council of Europe 2001: 22) and validation is always context-dependent. One can understand concerns from

language professionals outside Europe at what could be perceived as a reductive use of the CEFR to assist what could perhaps be regarded as neo-colonial expansion by the English language testing industry and associated ELF (English as a Lingua Franca) movement. As Savski points out, the "CEFR has mainly been interpreted as a language standard in Asian contexts, a view which is generally testing-oriented and largely excludes learners from being able to interpret the framework" (Savski 2019: 649). The motivation in these countries for the adoption of the CEFR, he reports, appears to be mainly neoliberal. He reports the juxtaposition of "CEFR" and "PISA" in documents, and even in tables of comparative levels in documents. In Malaysia, at least, this recontextualization of the CEFR extends to replacing local textbooks with ELT industry ones designed for CEFR levels.

In the current spread of the CEFR to South East Asia (see Foley 2019 for a review), others are concerned that it appears to be seen by governments as a silver bullet that will magically improve the effectiveness of language teaching, without provision of adequate opportunities for teachers to improve their level of language proficiency or pedagogic knowhow, and without adequate resources (e.g., Aziz and Uri 2017, in relation to Malaysia again). In Thailand, from a survey of 120 teachers, Franz and Teo (2018) conclude that, because of the way in which it is implemented, the perception of the CEFR "was first and foremost [as] a test" (Franz and Teo 2018: 9). As regards the meaning of the abbreviation CEFR: "'Cambridge' and 'Communication' were repeatedly cited for the letter 'C', and 'English' was more often cited for the letter 'E' than the actual European'" (Franz and Teo 2018: 11). Not that the teachers in these countries are necessarily against the type of teaching that the CEFR promotes, as all these authors (Aziz and Uri 2017; Frank and Theo 2017, Savski 2019) mention. The problem is a repetition of the naïve belief of twenty years ago in Europe that the introduction of the CEFR as a standard would automatically raise teacher efficiency and student achievement, beliefs rudely shattered for most by the results of the European Survey of Language Competence (European Commission 2012). In this respect, a study comparing use of the CEFR in Switzerland and Canada (Piccardo, North and Maldina 2019) suggests that, to achieve effective change through the CEFR, stakeholders at all levels (administrators, researchers, teachers) need to be involved in designing an on-going in-service teacher education programme spread over a considerable period of time, in which practitioners can be involved in the development of CEFR-related tools and materials and try them out in their classes.

North (2014) discusses many of the issues discussed above and then summarises with the table reproduced as Table 1 below.

Table 1. *Claims and counter-claims concerning CEFR normative influence (North 2014: 43)*

	Claim	Counterclaim
National level	Gives authorities a ready-made instrument to apply simplistically in language policy.	Empowers institutions and associations by providing the means to develop differentiated, local standards and assessments appropriate to context, yet linked to international standards. Thus helps avoid a takeover by multi-national high-stakes testing agencies.
Test providers	Forces test providers to align tests to the CEFR and to adapt the content of tests to the CEFR scheme in order to stay in the relevant market	Empowers new, smaller providers by giving a metalanguage and methodology to enable them to validate their product and explain it to users. This helps them to enter the market on equal terms, leading to a wider choice of validated assessment services. The CEFR provides a branching system of levels and categories that makes it easy to describe the profile of any language examination.

	Claim	Counterclaim
Teachers	Removes academic freedom and offers authorities a tool for increasing accountability for results (= levels) achieved	Empowers reformers by providing the means to challenge entrenched, inappropriate practices, where a CEFR-related curriculum is claimed.
Learners	Encourages monolingual approaches that fail to take account of the learner's plurilingual profile, developmental route and differentiated needs.	Through its face to learners, the Portfolio, encourages the concept of a plurilingual profile including mother tongue, plus the concept of course content determined by needs and priorities of the learners.

Finally, in the context of migration and citizenship, Krumm (2007) voiced fears that the CEFR levels are used to set barriers, rather than the hierarchy of descriptors being used in a differentiated manner to identify a reasonable profile for a context-appropriate standard. In fact, the 1996 and 1998 consultative versions of the CEFR showed such a profile, but it was removed in the 2001 edition as it was felt to be too complicated. Examples of such profiles have now been included in the CEFR/CV as previously mentioned. Unfortunately, as a recent survey demonstrates, the vast majority of member states still insist on a blanket level for all skills including writing (Rocca, Hamnes Carlsen and Deygers 2019). The CEFR was not intended to be used in this way, as stated in the preface to the CEFR Companion Volume:

> The CEFR is intended to promote quality plurilingual education, facilitate greater social mobility and stimulate reflection and exchange between language professionals for curriculum development and in teacher education. Furthermore, the CEFR provides a metalanguage for discussing the complexity of language proficiency for all citizens in a multilingual and intercultural Europe, and for education policy makers to reflect on learning objectives and outcomes that should be coherent and transparent. *It has never been the intention that the CEFR should be used to justify a gate-keeping function of assessment instruments.* (CoE 2020: 11, my emphasis).

8 Conclusion

In this article I have reviewed what seem to me to be the most common issues on which the CEFR has been criticised over the last twenty years. Other criticisms have of course also been made, for example the obvious one that the 2001 text is not exactly an easy read. This is an issue that the CEFR/CV makes a conscious effort to address, saying: "With this new, user-friendly version, the Council of Europe responds to the many comments that the 2001 edition was a very complex document that many language professionals found difficult to access" (CoE 2020: 21). The CEFR 2001 also showed signs of having been written in different styles by different authors, with a certain lack of balance—for example, why wasn't intercultural competence better developed when there was a background study on it undertaken at the time? There were also political compromises in the 2001 text, required to be a comprehensive compendium rather than promoting a viewpoint. This obscured some of the main innovations of the CEFR to the extent that even many people who have worked extensively with it did not take them on board. Here I am thinking of the move on from the four skills, the user/learner seen and treated as a social agent, the action-oriented approach as a classroom philosophy, mediation both within and across languages, and plurilingualism/pluriculturalism—let alone the connection between those concepts, which add up to a theoretically-grounded, ecological, pedagogic model (see Piccardo and North 2019). In many respects this opaqueness was difficult to avoid at the time because those concepts were all cutting-edge notions in the mid-1990s, which had not yet themselves been fully theorised.

At the same time, as I have tried to show in this article, one of the fundamental problems in relation to the CEFR is actually the nature of a lot of the criticism of the CEFR itself. To return to the ground I covered in the article, it is simply not true that the CEFR lacks a basis in theory; even if this basis was not spelled out in what is after all a language policy document and not an academic monograph, it was there in accompanying literature for those interested. The CEFR presentation of communicative proficiency was very sophisticated for its time, avoiding the rather static, componential, list-like nature of most contemporary models (See Piccardo and North, Chapter 2). It is in fact remarkable the extent to which the basic CEFR theoretical model, summarised in a paragraph in CEFR Chapter 2, did not need to be updated twenty years later for the CEFR/CV. That model allowed for the incorporation of later insights from the sociocultural theory, complexity theory, theory of action and agency and ecological theories of 'affordances' (See Piccardo and North 2019, Chapter 3). The conceptualisation of plurilingualism has more than stood the test of time and been justified by neurolinguistics research (see Piccardo, German-Rutherford and Lawrence forthcoming, especially the chapter by Riehl).

The criticisms in relation to research on learner language (SLA and CL) are also exaggerated, as we saw. The vast majority of research that has been undertaken supports the progression in the CEFR scales. The revision of the descriptors, 20 years on, offered the opportunity to incorporate any new insights. When it came to updating the 2001 scales in the CEFR/CV project, there was plenty of good material in relation to communicative language activities. But for communicative language competences and strategies, the sources were disappointing. The sum of the contribution from accessible SLA and CL research was the suppression of one example in one descriptor in the scale for grammatical accuracy at A2, at the suggestion of Belén Díez Belmar. In fact, for revision of the descriptor scales for aspects of communicative language competences, there was only some vocabulary work from the Finnish AMMKIA scale and some description of aspects of pragmatic competence, mainly from rating scales used by Cambridge Assessment. When it came to communicative language strategies, there was nothing at all for interaction or production, only descriptors for reception strategies from the REFIC framework produced in the MIRIADI intercomprehension project (De Carlo and Garbarino, forthcoming). In order to provide CEFR-informed contextualized descriptors, and to enhance curriculum innovation inspired by the CEFR, we need solid research that produces informed, constructive criticism that comes from a sustained engagement with the CEFR, as with SLATE, EALTA, ALTE, UNICert, as well the work of individual researchers like Díez Belmar and Wisniewski. But if it is to inform future revisions, this work needs to be reported in a manner in which it can be fed into new descriptors, or revision of existing ones.

Producing good descriptors is not a simple process because, even assuming that you know more or less what you wish to describe—which is far from being a given—there are three double binds. Firstly, as the CEFR 2001 and North (2000) explained, the descriptors need to be theoretically-based, but accessible to practitioners—and ideally learners—using categories that will be comprehensible to them. Secondly, for a common framework, descriptors need to be context-relevant, yet context-free because they must be relatable to a very wide range of contexts; a paradox. Finally, you need a lot of words to say what you would like to say in a descriptor, but experience both in 1994-1995 and in 2015-2016 showed that teachers do not accept descriptors longer than about 20 words or 250 characters. Therefore, one needs a principled development and validation methodology that mobilises large numbers of people to scrutinise the draft descriptors. It is easy to criticise the compromises and formulations in the end result, but it is not easy to produce something better—and cover all the levels people expect. It is also easy to say that certain descriptors are not relevant to one's students, or might be relevant but do not reflect the local context. But one should remember that they are not intended to necessarily apply unadapted to that context and those students; they are generic, illustrative examples that may need tweaking or replacing, or which may inspire a totally different approach.

Finally, any instrument like the CEFR needs to be used responsibly. Any educational implementation needs to be accompanied by long term teacher education programmes if it is to be successful. It is important to emphasise that the CEFR is a heuristic, not a standard. It is a reference tool for reflection,

not a panacea to be 'applied.' The CEFR descriptors are a source for curriculum design, not a collection of rating scales:

> The aim of the descriptors is to provide input for curriculum development. The descriptors are presented in levels for ease of use. Descriptors for the same level from several scales tend to be exploited in adapted form on checklists of descriptors for curriculum or module aims and for self-assessment. (CoE 2018: 40, 2020: 41).

As regards immigration and citizenship, one should remember that Governments do not actually need the CEFR to set linguistic standards for these purposes; the English-speaking world, for example, did fine with IELTS and other tests beforehand. In addition, even though the CEFR as an educational resource is not intended to be used for gatekeeping, should it be appropriated for this purpose, at least it brings transparency, the recognition of low levels of proficiency (now including Pre-A1) and the recommendation to define appropriate profiles, which, taken together could offer a possible basis for the enlightenment of and negotiation with policy makers by language professionals.

The CEFR is certainly not perfect, but it is open-ended, as shown by the recent update with the CEFR/CV. The CEFR is still not used to its full potential. It anticipated and facilitates the actional turn, the pluri/multilingual turn and the linking of language learning to democratic citizenship and social justice. The CEFR/CV builds on and extends this foundation, hoping to set a trend, as happened 20 years ago. The provision of descriptors for aspects of mediation and for plurilingual and pluricultural competence provides concrete tools for that purpose. This represents a serious attempt to broaden the scope of language education—as the CEFR 2001 helped to do with its 'Can Do' descriptors. The theoretical underpinning of the development is given by Piccardo and North (2019). The aim is the furthering of plurilingualism and interculturality in inclusive, quality education for all.

9 References

Alderson, J. Charles. 1991. Bands and scores. In J. Charles Alderson & Brian North (eds.), *Language testing in the 1990s*, 71-86*. London: Macmillan/Modern English Publications: British Council.

Alderson, J. Charles, Neus Figueras, Henk Kuijper, Günther Nold, Sauli Takala & Claire Tardieu. 2006. Analysing tests of reading and listening in relation to the Common European Framework of Reference: the experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly* 3(1). 3-30.

Alderson, J. Charles & Alexander, H. Urquhart. 1984. *Reading in a foreign language.* Harlow: Longman.

Aziz, Mohd Sallehhudin Abd & Nurul Farehah Mohd Uri. 2017. CEFR in Malaysia: Current issues and challenges in the implementation of the framework. *3L: The Southeast Asian Journal of English Language Studies* 24(3). 168-183 http://doi.org/10.17576/3L-2018-2403-13 (accessed 22.05.20).

Bachman, Lyle F. & Adrian S. Palmer. 1996. *Language testing in practice*. Oxford: Oxford University Press.

Backus, Ad, Durk Gorter, Karlfried Knapp, Rosita Schjerve-Rindler, Jos Swanenberg, Jan D. ten Thije & Eva Vetter. 2013. Inclusive multilingualism: Concept, modes and implications. *European Journal of Applied Linguistics* 1(2). 179-215. DOI*: https://*doi*.org/*10.1515*/*eujal-2013-0010 (accessed 22.05.20).

Barnes, Douglas & Frankie Todd. 1977. *Communication and learning in small groups*. London: Routledge and Kegan Paul.

Bartning, Inge, Maisa Martin & Ineke Vedder (eds.). 2010. *Communicative proficiency and linguistic Development: Intersections between SLA and language testing research*, *Eurosla Monographs Series* 1. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

Bärenfänger, Olaf, Claudia Harsch, Bernd Tesch & Karin Vogt. 2018. Reform, Remake, Retousche? Diskussionspapier der Deutschen Gesellschaft für Fremdsprachenforschung zum *Companion to the CEFR 2017.*

Breen, Michael P. & Christopher N. Candlin. 1980. The essentials of a communicative curriculum in language teaching. *Applied Linguistics* 1(2). 89-112.

Brown, Gillian, Anne Anderson, Richard Shilock & George Yule. 1984. *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Brumfit, Christopher J. 1984. *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Bryant, Antony. 2014. The grounded theory method. In Patricia Leavy (ed.), *The Oxford handbook of qualitative research*, 116–136. Oxford: Oxford University Press.

Burton, Deidre. 1980. *Dialogue and discourse; A sociolinguistic approach to modern drama dialogue and naturally occurring conversation*. London: Routledge and Kegan Paul.

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

Carlsen, Cecilie. 2010. Discourse connectives across CEFR-levels: A corpus based study. In Inge Bartning, Maisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic Development: Intersections between SLA and language testing research*, *Eurosla Monographs Series* 1, 191-210. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

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

Champney, Horace. 1941. The measurement of parent behavior. *Child Development* 12(2). 131-166.

Conteh, Jean & Gabriela Meyer (eds.). 2014. *The multilingual turn in languages education: Opportunities and challenges*. Bristol: Multilingual Matters.

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

Council of Europe. 2007 *The Common European Framework of Reference for Languages (CEFR) and the development of language policies: challenges and responsibilities.* Intergovernmental Language Policy Forum, Strasbourg, 6-8 February 2007, Report, Strasbourg: Council of Europe. https://www.coe.int/en/web/common-european-framework-reference-languages/documents (accessed 22.05.20).

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg, France: Council of Europe. https://www.coe.int/en/web/common-european-framework-reference-languages/documents (accessed 22.05.20).

Council of Europe & Council of Ministers. 2008. Recommendation CM/Rec(2008)7 to member states on the use of the Council of Europe's "Common European Framework of Reference for Languages" (CEFR) and the promotion of plurilingualism. https://www.coe.int/en/web/common-european-framework-reference-languages/extracts-recommendation-2008-7 (accessed 22.05.20).

Creswell, John W. & Vicki L. Plano Clark. 2018. *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage.

Cummins, Jim. 1980. The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly* 14. 175-187.

De Carlo, Maddalena & Sandra Garbarino. forthcoming. Intercomprehension: strengths and opportunities of a pluralistic approach. In Enrica Piccardo, Aline Germain-Rutherford & Geoff Lawrence (eds.) *Routledge handbook of plurilingual language education*. London and New York: Routledge.

Dendrinos, Bessie & Voula Gotsouilia. 2015. Setting standards for multilingual frameworks in foreign language education. In Bernard Spolsky, Ofra Inbar & Michal Tannenbaum (eds.), *Challenges for language education and policy. Making space for people*. 23-50. London and New York: Routledge.

Deygers, Bart. 2019. The CEFR Companion Volume: Between research-based policy and policy-based research. *Applied Linguistics* doi:10.1093/applin/amz024 (accessed 22.05.20).

Deygers, Bart, Beate Zeidler, Dina Vilcu & Cecilie Hamnes Carlsen. 2018 One framework to unite them all? Use of the CEFR in European university entrance policies. *Language Assessment Quarterly* 15, 1, 3-15, DOI: 10.1080/15434303.2016.1261350 (accessed 22.05.20).

Díez-Belmar, Maria Belén. 2018. Fine-tuning descriptors for CEFR B1 level: insights from learner corpora. *ELT Journal* 17(2). 199-209.

Eichelmann, Teresa. 2015. Der gemeinsame europäische Referenzrahmen für Sprachen und die Hochschulspezifik am Beispiel UNIcert [The Common Europe Framework of Reference and the Higher Education Context, the example UNICert]. Unpublished PhD thesis, University of Dresden.

ELTDU. 1976. *Stages of attainment scale*. Oxford: English Language Development Unit (ELTDU), Oxford University Press.

European Commission. 2012. *First European survey on language competences, Executive summary.* http://www.surveylang.org/media/ExecutivesummaryoftheESLC_210612.pdf (accessed 22.05.20).

Færch, Claus & Gabriele Kasper. 1983. Plans and strategies in foreign language communication. In Claus Færch & Gabriele Kasper (eds.), *Strategies in interlanguage communication*, 20-60. Harlow: Longman.

Foley, Joseph A. 2019. Issues on the initial impact of CEFR in Thailand and the region. *Indonesian Journal of Applied Linguistics* 9. 359-370. doi: 10.17509/ijal.v9i2.20233 (accessed 22.05.20).

Forsberg, Fanny & Inge Bartning. 2010. Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French. In Inge Bartning, Maisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic Development: Intersections between SLA and language testing research*, Eurosla Monographs Series 1, 133-158. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

Franz, Jens & Adisa Teo. 2018. 'A2 is Normal' – Thai secondary school English teachers' encounters with the CEFR. *RELC Journal* 49(3). 1-17. DOI: 10.1177/0033688217738816 (accessed 22.05.20).

Frost, Dan & Jean O'Donnell. 2015. Réussite : Être ou ne pas être B2 : Telle est la question. (le projet ELLO Étude longitudinale sur la langue orale) [Success B2 or not B2 – That is the question. (The ELLO project Étude longitudinale sur la langue orale)]. *Cahiers de l'Apliut XXXIV* (2). https://journals.openedition.org/apliut/5195 (accessed 22.05.20).

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

Fulcher, Glenn. 2008. Testing times ahead? *Liaison Magazine* 1. 20-24 July 2008. UK Subject Centre for Languages, Linguistics and Area Studies, University of Southampton.

Fulcher, Glenn. 2010. *Practical language testing*. London: Hodder Education.

Graves, Katherine. 2008. The language curriculum: A social contextual perspective. *Language Teaching* 41(2). 147-181.

Halliday, Michael A.K. 1989. *Spoken and written language*. Oxford: Oxford University Press.

Harrison, Julia, & Fiona Barke. 2015. *English profile in practice*. Cambridge: Cambridge University Press.

Harsch, Claudia. 2019. What it means to be at a CEFR level. Or why my Mojito is not your Mojito – on the significance of sharing Mojito recipes. In Ari Huhta, Gudrun Erickson & Neus Figueras (eds.), *Developments in language education: A memorial volume in honour of Sauli Takala*, 76-93. Jyväskylä: University of Jyväskylä Centre for Applied Language Studies/EALTA (European Association for Language Teaching and Assessment). http://www.ealta.eu.org/documents/resources/Developments%20in%20Language%20Education%20A%20Memorial%20Volume%20in%20Honour%20of%20Sauli%20Takala.pdf (accessed 22.05.20).

Hulstijn, Jan. 2007. The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language. *Modern Language Journal* 91(4). 663-666.

Hulstijn, Jan H., J. Charles Alderson & Rob Schoonen. 2010. Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In Inge Bartning, Maisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic Development:*

Intersections between SLA and language testing research, *Eurosla Monographs Series* 1, 11-20. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

Idris, Mardiana. & Abdul Halim Abdul Raof. 2017. The CEFR rating scale functioning: An empirical study on self- and peer Assessments. *Sains Humanika* 9(4). 2. 11-17. https://sainshumanika.utm.my/index.php/sainshumanika/article/view/1355 (accessed 22.05.20).

Kramsch, Claire. 1986. From language proficiency to interactional competence. *The Modern Language Journal* 70(4). 366-372.

Krumm, Hans-Jürgen. 2007. Profiles instead of levels. The CEFR and its (ab)use in the context of migration. *The Modern Language Journal* 91**(**4). 667-669.

Lado, Robert. 1961. *Language testing*. London: Longman.

Larsen-Freeman, Diane. 2011. A complexity theory approach to second language development/acquisition. In Dwight Atkinson (ed.), *Alternative approaches to second language acquisition*, 48-72. New York: Routledge.

Little, David. (ed.) 2018. The CEFR Companion Volume with New Descriptors: Uses and implications for language testing and assessment. 6th EALTA CEFR SIG Meeting, Trinity College Dublin, 27 January 2018. https://www.researchgate.net/publication/330741857_The_CEFR_Companion_Volume_with_New_Descriptors_Uses_and_Implications_for_Language_Testing_and_Assessment_VIth_EALTA_CEFR_SIG (accessed 22.05.20).

Lüdi, Georges. 1999. Ein Gesamtsprachenkonzept für die Schweizer Schulen: Leitgedanken [A comprehensive concept for language education in Swiss schools]. *Babylonia* 4/98. 11-13.

Martin, Maisa, Sanne Mustonen, Nina Reiman & Marja Seilonen. 2010. On becoming an independent user. In Inge Bartning, Maiisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic Development: Intersections between SLA and language testing research*, *Eurosla Monographs Series* 1. 57-80. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

May, Stephen. 2014. *The multilingual turn: Applications for SLA, TESOL and bilingual education*. New York: Routledge.

McNamara, Tim. 2011 Managing learning: Authority and language assessment, *Language Teaching* 44(4). 500-515. October 2011.

McNamara, Tim, Janne Morton, Neomy Storch & Celia Thompson. 2018. Students' accounts of their first-year undergraduate academic writing experience: Implications for the use of the CEFR. *Language Assessment Quarterly* 15(1). 16-28. DOI: 10.1080/15434303.2017.1405420 (accessed 22.05.20).

Milton, James. 2010. The development of vocabulary breadth across the CEFR levels. In Inge Bartning, Maiisa Martin & Ineke Vedder (eds.), *Communicative proficiency and linguistic Development: Intersections between SLA and language testing research*, *Eurosla Monographs Series* 1, 211-232*. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

North, Brian. 1995. The development of a common framework scale of descriptors of language proficiency based on a theory of measurement. *System* 23(4). 445-465.

North, Brian. 1997a. Perspectives on language proficiency and aspects of competence. *Language Teaching.* 30. 92-100.

North, Brian. 1997b. The development of a common framework scale of descriptors of language proficiency based on a theory of measurement. Paper given at the Language Testing Research Colloquium, Tampere 31 July-3August 1996. In Ari Huhta, Viljo Kohonen, Liisa Kurki-Suonio & Sari Luoma (eds.), *Current developments and alternatives in language assessment,* 423-449. Jyväskylä, University of Jyväskyla.

North, Brian. 2000. *The development of a common framework scale of language proficiency*. New York: Peter Lang.

North, Brian. 2007. The CEFR illustrative descriptor scales. *Modern Language Journal* 91. 655-658.

North, Brian. 2008. The CEFR levels and descriptor scales. In Linda Taylor, L. and Cyril J. Weir, (eds), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference, May 2005.* 21-66. Cambridge. Studies in Language Testing Series 27, Cambridge University Press.

North, Brian. 2014. *The CEFR in practice*. Cambridge: Cambridge University Press.

North, Brian & Elizabeta Jaroscz. 2013. Implementing the CEFR in teacher-based assessment: approaches and challenges. In Evelina Galaczi & Cyril J. Weir (eds.), *Exploring language frameworks: Proceedings of the ALTE Krakow Conference, July 2011*. 118-134. Cambridge. Studies in Language Testing Series 36, Cambridge University Press.

North, Brian & Enrica Piccardo. 2016. Developing illustrative descriptors of aspects of mediation for the Common European Framework of Reference (CEFR). Strasbourg, France: Council of Europe. https://www.coe.int/en/web/common-european-framework-reference-languages/documents (accessed 22.05.20).

North, Brian & Günther Schneider. 1998. Scaling descriptors for language proficiency scales. *Language Testing* 15(2). 217-262.

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

Piccardo, Enrica, Aline Germain-Rutherford, & Richard Clément (eds.). 2011. Adopter ou adapter le Cadre européen commun de référence est-il seulement européen? [Adopt or adapt the Common European Framework of Reference; is it only European?] *Synergies Europe*, 6. http://ressources-cla.univ-fcomte.fr/gerflint/Europe6/Europe6.html (accessed 22.05.20).

Piccardo, Enrica & Brian North. 2019. *The Action-oriented approach: A dynamic vision of language education.* Bristol: Multilingual Matters.

Piccardo, Enrica, Brian North & Eleonora Maldina. 2019 Promoting innovation and reform in language education through a quality assurance template for CEFR implementation. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*. *Special Issue* 22(1). 103-128.

Porto, Melina. 2012. Academic perspectives from Argentina. In Michael Byram & Lynne Parmenter (eds.), *The Common European Framework of Reference: The globalisation of language policy,* 129-138. Bristol: Multilingual Matters.

Richards, Jack C. 2013. Curriculum approaches in language teaching: forward, central and backwards design. *RELC Journal* 44(1). 5-33.

Richer, Jean-Jacques. 2017. Quand le monde du travail peut venir en renfort de la didactique des langue [When the world of work can come to reinforce language pedagogy]. *Revue TDFLE* 70 *La pensée CECR. [TDFLE Review* 70*: The CEFR Concept].* 1-34. http://revue-tdfle.fr/revue-70-6/36-quand-le-monde-du-travail-peut-venir-en-renfort-de-la-didactique-des-langues- (accessed 22.05.20).

Rocca, Lorenzo, Cecilie Hamnes Carlsen & Bart Deygers. 2019. Linguistic Integration of adult migrants: requirements and learning opportunities. Report on the 2018 Council of Europe and ALTE survey on language and knowledge of society policies for migrants. https://rm.coe.int/linguistic-integration-of-adult-migrants-requirements-and-learning-opp/16809b93cb (accessed 22.05.20).

Quetz, Jürgen & Henning Rossa (eds). 2019. The Common European Framework of Reference, Illustrative Descriptors, Extended Version 2017. [Special issue] *Zeitschrift für Fremdsprachenforschung* 30(2).

Savski, Kristof. 2019. Putting the plurilingual/pluricultural back into CEFR: Reflecting on policy reform in Thailand and Malaysia. *The Journal of Asia TEFL* 16(2). 644-652.

Shackleton, Caroline. 2018. Developing CEFR-related language proficiency tests: A focus on the role of piloting. *CercleS* 8(2). 333-352. https://doi.org/10.1515/cercles-2018-0019 (accessed 22.05.20).

Schneider, Günther & Brian North. 2000. *Fremdsprachen können: was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit.*

[Knowing a foreign language: What does that mean? Scales for the description, assessment and self-assessment of foreign language communicative proficiency]. Chur/Zürich: Nationales Forschungsprogramm 33: Wirksamkeit unserer Bildungssysteme [National Research Programme 33: Effectiveness of Our Education System]/Verlag Rüegger.

Smith, Patricia, Cain & J.M. Kendall. 1963. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology* 47(2). 149-155.

Stern, Hans-Heinrich (David). 1983. *Fundamental concepts of language teaching*. New York: Oxford University Press.

Swales, John M. 1990. *The genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Szabo, Tunde & Tim Goodier. 2018. Collated representative samples of descriptors of language competence developed for young learners. Resource for educators. https://www.coe.int/en/web/common-european-framework-reference-languages/bank-of-supplementary-descriptors (accessed 22.05.20).

Tarone, Elaine. 1983. Some thoughts on the notion of "communication strategy". In Claus Færch & Gabriele Kasper (eds.), *Strategies in interlanguage communication,* 63-68. Harlow: Longman. Originally published in *TESOL Quarterly* 1981.

Thewissen, Jennifer. 2013. Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal* 97(S1). 77-101.

Tono, Yuko. 2013. Automatic extraction of L2 criterial lexicogrammatical features across pseudo-longitudinal learner corpora: Using edit distance and variability-based neighbour clustering. *Eurosla Monographs Series* 2. 149-176. http://www.eurosla.org/publications/eurosla-monographs-discontinued-from-2017 (accessed 22.05.20).

Tracy, Rosemarie. 2017. Language testing in the context of migration. In Beacco, Jean-Claude, Krumm, Hans-Jürgen, Little, David, Thalgott, Philia, Council of Europe (eds.), *The Linguistic Integration of Adult Migrants/L'intégration linguistique des migrants adultes. Some lessons from research/Les enseignements de la recherche*, 45-56. Berlin/Boston: de Gruyter.

van Lier, Leo. 2010. The ecology of language learning: Practice to theory, theory to practice. *Procedia Social and Behavioral Sciences* 3. 2-6.

Vogt, Karin. 2011. *Fremdsprachliche Kompetenzprofile [Foreign language competence profiles]*. Tübingen: Narr Verlag.

Wiemann, John M. & Philip Backlund. 1980. Current theory and research in communicative competence. *Review of Educational Research* 50(1). 185-199.

Wisniewski, Kathrin. 2017. Empirical learner language and the levels of the Common European Framework of Reference. *Language Learning* 67. 232-253.

10 Biography

Brian North is a researcher and consultant to the Council of Europe. After developing the CEFR levels and descriptors, he co-authored the CEFR itself, the prototype European Language Portfolio, the Manual for relating assessments to the CEFR, and Eaquals' core inventories for English and French. Recent projects include the coordination of the CEFR/CV plus co-investigation in a study of CEFR use in Canada and Switzerland, alignment of the Canadian Language Benchmarks to the CEFR and the ECML project CEFR QualiMatrix. Recent (co-)publications include *The CEFR in Practice* (CUP 2014), *Language Course Planning* (OUP 2018) and *The Action-oriented Approach* (Multilingual Matters 2019).