One Framework to Unite Them All? Use of the CEFR in European University Entrance Policies

Bart Deygers, Beate Zeidler, Dina Vilcu & Cecilie Hamnes Carlsen

📅 Published online: 10 Apr 2017.

✎ Submit your article to this journal ⬏

🔍 View related articles ⬏

CrossMark ⬚ View Crossmark data ⬏

Routledge
Taylor & Francis Group

One Framework to Unite Them All? Use of the CEFR in European University Entrance Policies

Bart Deygers[a], Beate Zeidler[b], Dina Vilcu[c], and Cecilie Hamnes Carlsen[d]

[a]KU Leuven, Leuven, Belgium; [b]telc GmbH, Frankfurt am Main, Germany; [c]Universitatea Babes-Bolyai, Cluj-Napoca, Romania; [d]VOX, Bergen, Norway

ABSTRACT

Fifteen years after its publication, the Common European Framework of Reference for Languages is a commonly used document in language tests and policies across Europe. This article considers the CEFR's impact on university entrance language tests and policies that are used to regulate the entrance of international L2 students who wish to study in a national language of the host country. Using a qualitative approach, this study aims to (a) outline and compare the target language demands toward L2 students entering European universities and to (b) determine the impact of the CEFR on European university entrance policies, tests, and testers. This article offers an overview of the university entrance language requirements for foreign L2 students in 28 European countries or regions with an autonomous educational policy. It is based on structured interviews with 30 respondents involved with university entrance test development and knowledgeable about university entrance policies in their context. The results show that the CEFR is omnipresent in European university entrance language tests and that the B2 is the most commonly used level in that context. The data also show that normative CEFR use is very common and that in many contexts CEFR levels are misused for marketing purposes or to control university admission.

The CEFR: A 45-years-old debate

The goal of the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2001) is to promote the free movement of people and ideas by increasing the transparency across educational systems through the common use of the same proficiency levels (Van Ek, 1975). The first drafts of what was to become the CEFR appeared in the early 1970s with the development of the *Threshold* level, later called B1 (Van Ek, 1975). Through the years new levels were added (Van Ek & Trim, 1991b, 2001), existing levels were refined (Van Ek & Trim, 1991a), and some 30 years after the first drafts, the project culminated in the "blue book" we know today. The CEFR proposes six consecutive levels of language proficiency, ranging from A1 to C2. It focuses on what learners putatively can do with language and includes 53 illustrative scales, which list language-independent descriptions of each proficiency level for a given skill or ability. Arguably, these illustrative scales have become the most influential and the most heavily criticized aspect of the CEFR (Figueras, 2012; Little, 2007).

CEFR criticism is usually either political in nature or content-related (Figueras, 2012). The politically oriented criticism sees the CEFR as an instrument of power that encourages a simplistic, level-driven logic at the expense of a needs-based, user-driven policy (McNamara, 2007; Shohamy, 2011). According to this strand of criticism the CEFR levels are often used as a

---

CONTACT Bart Deygers ✉ bart.deygers@kuleuven.be ▣ KU Leuven, CTO, Blijde Inkomststraat 7, Leuven, 3000 Belgium.

normative standard that provides policy makers with an easy tool to assist in gatekeeping (Fulcher, 2004, 2012). In a defense against this criticism, North warns against confusing intended CEFR use with actual but inappropriate use (North, 2014a) and considers any normative use of the CEFR to fall under this category (Martyniuk, 2010; North, 2014a). Furthermore, so the argumentation goes, the CEFR discourages a simplistic level-based gatekeeping policy, because it stimulates discussion between decision makers and language experts (Porto, 2012). On the other hand, content-related criticism focuses on those aspects of the CEFR that are problematic in the context of language testing—the field where the CEFR's influence has been most keenly felt (Little, 2007). Some authors tackle the foundation of the CEFR, pointing out that it lacks empirical validation (Fulcher, 2012), ignores insights from second language acquisition theory (Alderson, 2007; Hulstijn, 2007; Little, 2007), does not pay equal attention to all skills (Alderson et al., 2006; Staehr, 2008; Weir, 2005), or does not cover the full range of levels in every scale (Alderson et al., 2006). It is now commonly recognized that the theoretical support for the descriptors of the receptive skills is not quite robust (Alderson, 2007; North, 2014a), but a growing body of recent research has been providing language-specific empirical validation for the CEFR (Carlsen, 2014; Salamoura & Saville, 2009). A second strand of criticism related to the content focuses on the deficiencies of the CEFR as a common reference point in language testing. This line of criticism maintains that (a) the levels are not equidistant because they contain overlaps and inconsistencies and (b) the descriptors are general, language-independent, containing impressionistic terminology (Alderson et al., 2006; Alderson, 2007; Fulcher, 2012; Papageorgiou, 2010).

This fierce criticism toward the CEFR may seem somewhat overstated in comparison with the document's intended use, which appears relatively modest. The authors have repeatedly pointed out that the CEFR descriptors were meant to be general and that the levels were never meant to make up an interval scale, so in part the content-related criticism points out a characteristic that was purposefully built in to the CEFR (Little, 2007; North, 2014a). In their defense, CEFR authors also state that despite the general nature of the CEFR descriptors, professionals in organizations, such as ALTE and EALTA, interpret the CEFR levels in a uniform fashion (North, 2014a). When it comes to the use of the CEFR, its authors insist that it was not meant as a normative document but as a malleable heuristic that stimulates reflection, facilitates discussion about language learning, and aids curriculum planning and language certification (North, 2007, 2014a, 2014b).

The actual use of the CEFR has not always been in tune with its intended use however (North, 2014a), and many language tests across Europe have been required to link their scoring system to the CEFR (Fulcher, 2004) or have been redeveloped with the CEFR descriptors in mind (Galaczi, Ffrench, Hubbard, & Green, 2011). In addition, because many countries and institutions in the EU now use CEFR levels to set legally binding citizenship requirements (Van Avermaet & Pulinx, 2013), curriculum goals (University of Cambridge, ESOL Examinations, 2011), and university entrance demands (Xi, Bridgeman, & Wendler, 2014), language testers in Europe have been required to follow suit.

It is clear then that 15 years after its publication, the CEFR has inspired a large body of literature, and has gained critics and champions. It has been praised for facilitating dialogue (North, 2014a) and denounced for causing validity chaos (Fulcher, 2012). It has been studied in the context of rating scale design (Harsch & Martin, 2012) and concurrent validation (Deygers, 2017), and it has been the subject of sociolinguistic debates (Roever & McNamara, 2006). It has been adopted by users in other contexts and on other continents (Negishi, Takada, & Tono, 2013), and has been accused of linguistic imperialism (Curtis, 2015). The CEFR has led to a lot of language testing research, but its actual impact on the lives of test takers has been largely unresearched. Still, because of its effect on language tests, the CEFR has potentially affected the lives of millions. The study presented in this article focuses on the CEFR's impact in one specific field: university entrance language testing for L2 students from abroad.

THE CEFR and L2 university entrance requirements: Research questions

There is little reliable information about the impact of the CEFR on L2 university entrance language policy across Europe, and the information that is available is fragmented or scattered across the websites of Europe's universities. While factual data about the use of the CEFR in university entrance policy are scarce, there are a few commonly held assumptions that this study was intended to verify.

Assumption 1: After the Bologna agreement (1999) and the publication of the CEFR, university entrance policies across Europe have become somewhat streamlined (EACEA, 2010).

Assumption 2: B2 is the most commonly used level in European university entrance tests (Xi *et al.*, 2014).

Assumption 3: The CEFR has had a great impact on university entrance language testing (Fulcher, 2012).

The questions that guide this explorative study are intended to verify these assumptions, which are largely unsupported by research. The first research aim of this article relates to assumptions 1 and 2:

RQ1: How uniform are the European university entrance language requirements for international L2 students?

This study was also designed to determine the influence of the CEFR on university entrance policies and tests and to map the respondents' opinions about the CEFR (assumption 3).

RQ2a: What is the impact of the CEFR on university admission policies and university admission language tests in Europe?

RQ2b: How do test developers perceive the CEFR and its impact?

To date, no studies have compared the language demands toward L2 students in Europe. No study has gauged the actual impact of the CEFR on university entrance language testing (Figueras, 2012), and no study has verified whether the B2 level is the most commonly used level for determining L2 students' university entrance.

Method

Respondents

The information required for this study is not easily available, and is only known by relatively few people who work in the context of university admission or university entrance language testing. Consequently, collecting data via randomized sampling was of little use, and respondents were chosen through purposeful selection (Freeman, 2000), which implies identifying knowledgeable and information-rich respondents (Reybold, Lammert, & Stribling, 2013). All respondents were professionally involved in language testing, and in many cases in the development of language tests for university entrance. The researchers contacted members of European language testing organizations that are full members or affiliates of the Association of Language Testers in Europe (ALTE). Thirty-nine organizations were asked to take part in the study, and in the end representatives of 30 organizations operating within 28 states or regions with autonomy over educational matters agreed to participate. Although test operationalization was not a selection criterion, all university entrance language tests included in this study are skill based, seven of which also feature a grammar and vocabulary section.

For reasons of readability we use the term "context" throughout the text when referring to both nation-states and regions that have (quasi) autonomy over educational matters. Consequently, even though this study covers 26 nation-states, we report on 28 contexts. Because of the independent

educational policies in Wallonia, Flanders (both in Belgium) and the Basque region (Spain), these regions were considered distinct contexts. In other countries, such as Switzerland, there are different linguistic regions, but the policy and the entrance tests in these regions share the same vision and characteristics, and as such, they are counted as one context. In some cases there were two respondents per context because there was more than one full member of ALTE for that country, and more than one member wished to participate (R10 and R11, R29 and R30). These doubles allowed the researchers to compare the answers to the factual questions. In both cases the same information was given by both respondents. Table 1 lists the codes of the respondents and the countries surveyed in this study.

Data collection and analysis procedures

Because it was considered important to collect data that shed light on each individual context and allowed for meaningful comparison, structured interviews were used. Respondents were asked a fixed set of questions, but they were free to contextualize their answers in a way that questionnaires cannot accommodate (Schwartz, Knäuper, Oyersman, & Stich, 2008). The interview scenario consisted of three parts: (a) one concerning the university entrance policy, (b) one concerning the entrance tests, and (c) one concerning the interviewee's personal opinions about the CEFR and university entrance language tests. Respondents received the factual questions 3 days beforehand via e-mail and were encouraged to look up any information they did not have on hand. Questions relating to personal opinions were not distributed beforehand to avoid constructed or socially desirable answers. The scenario was trialed twice in interview conditions, and the interviews were conducted by four trained researchers via Skype and were recorded using the Audacity software.

All interviews were transcribed verbatim. The transcriptions were checked by two researchers who replayed the interview, corrected any inaccuracies, coded the transcripts independently, and compared codes and outcomes afterward. Coding was done both manually and using the qualitative software NVivo for Mac, because a combination of manual and computer-assisted coding is likely to yield the most reliable results (Welsh, 2002). Both researchers used a set of agreed-upon a priori codes that corresponded with the questions asked in the structured scenario. After the first round of coding, the exact inter-rater agreement was checked for the a priori coding categories (86.4%), based on a random sample representing 30% of the total transcribed text. Both coders also used a grounded theory approach (Glaser & Strauss, 1967; Miles & Huberman, 1994), which means that they also coded salient issues that emerged from the data. This double approach allowed the researchers to compare factual data across contexts and to spot opinions or views that surfaced during the interviews without being explicitly probed for. After coding independently, the researchers discussed

Table 1. Countries and regions surveyed.

R1	Austria	R17	Luxembourg
R2	Belgium (Flanders)	R18	Malta
R3	Belgium (Wallonia)	R19	Netherlands
R4	Bulgaria	R20	Norway
R5	Czech Republic	R21	Poland
R6	Denmark	R22	Portugal
R7	Estonia	R23	Romania
R8	Finland	R24	Slovenia
R9	France	R25	Spain (Basque)
R10	Germany	R26	Spain
R11	Germany	R27	Sweden
R12	Greece	R28	Switzerland
R13	Hungary	R29	United Kingdom
R14	Ireland	R30	United Kingdom
R15	Italy		
R16	Lithuania		

their coding categories over several meetings until consensus was reached. The complexity of the coded data was reduced by quantifying recurring practices and patterns (Ziegler & Kang, 2016). These quantifications are presented in the tables below.

All interviews were conducted in English, and the quotes included in this article are literal transcriptions that have been lightly edited for the sake of readability. Editing was restricted to correcting grammatical flaws and omitting word repetitions and filled pauses. Throughout the article "I" is used to refer to the interviewer, and respondents are referred to by their code (i.e., R1). Misinterpretation of what was said during the interview is always a possibility. For that reason, the transcriptions were sent back to the respondents so they could comment on any factual flaws. When this happened, the information was used in the data analysis, but the original transcripts were not altered. As a final step in the verification of the interpretation of the data, the respondents, as well as the members of ALTE's CEFR special interest group received a prefinal draft of this article, which they could amend.

Results

In 23 contexts, passing a language test is mandatory for university entrance for L2 students. In three of these countries, only one centralized test is accepted for first-year university entrance, but the 20 remaining contexts all have a system where multiple tests are used for the same high-stakes purpose. In 13 contexts both centralized tests and local tests—developed by the receiving university itself—are accepted (Table 2).

Respondents do not generally regard local tests positively and express concern over their quality, transparency, and comparability. Of the 14 respondents from contexts where both centralized and local tests can grant university access, 8 respondents wished to streamline the university entrance system.

R10: [The locally developed tests] don't apply piloting, pretesting and statistical analysis, so it's not astonishing that the results were very heterogeneous and the exams are of a different level of difficulty.

In 22 contexts there is a university entrance language requirement that is expressed in CEFR terms (Table 3). In five contexts (Finland, Luxemburg, Portugal and Spain, including the Basque

Table 2. Tests accepted for university entry (*N* = 28).

	Number of contexts
Multiple centralized and local tests	13
Multiple centralized tests	5
One centralized test	3
Multiple local tests	2
No test	5

Table 3. CEFR level required for university entrance (*N* = 28).

	Number of contexts
A2 and B2[†]	1
B1 or higher [†]	1
B2	9
B2+	1
B2 or C1 [†]	8
C1 or higher [†]	2
Requirement not CEFR-related	1
No requirement	5

[†] Depending on the program.

region) there are no specific language requirements, and in Sweden the requirements are not CEFR related.

The findings of the study confirm the assumption that B2 is the most commonly required level for university entrance across Europe: in nine contexts B2 is the only level, and in another ten it is one of the required levels. But even though B2 is the most commonly used CEFR level for university entrance, there is no agreement among the 30 individual respondents that B2 users have the linguistic resources required to function at the start of university (Table 4).

The respondents may doubt the B2 level as an adequate level for university entrance, but test developers are rarely the ones who make the decision on where to set the entrance level requirements. In 16 contexts (Table 5) the language level requirements for university entrance are partly or completely up to the university or faculty. In four of 16 instances the ministry decides on a general rule, but the university determines the actual level, and in the 12 remaining cases the level requirements are left to the university entirely. In seven contexts there is a national regulation stipulating the level requirements.

In most contexts where a language level is required for university entrance, that level was not determined on the basis of empirical data or needs analyses, the respondents report.

R4: There isn't any official study that said that B2 is the right level for students. It's just intuitive, from the practice in universities. Knowing that other countries require B2 they just decided to introduce B2.

R24: [The university] asked us about our opinion, and we said, in Europe it's mostly B2. And they said, okay, it should be B2.

R21: I don't think it's based on any empirical ground. It's just an administrative decision [...] Those in charge of making decisions start with the levels. Then the whole field is trying to adjust to the decisions rather than starting with an analysis of the needs [...] When the process has been reversed, it is very difficult to turn it around.

In the 23 contexts where L2 university entrance is regulated by language tests linked to the CEFR, language requirements rarely seem to be based on empirical data (Table 6). Only one respondent claimed that the requirements in his/her context were empirically founded.

Table 4. Is B2 enough to function linguistically at the start of university? (N = 30).

		Number of contexts
No		7
B2+ as a minimum	(4)	
C1 as a minimum	(3)	
Not quite		8
If additional language support is offered	(3)	
Depends on student needs	(5)	
Yes		10
B2 is enough	(3)	
B2 is the absolute minimum	(7)	
Don't know		5

Table 5. Who decides on language requirements (N = 28).

	Number of contexts
No requirement	5
Government (general rule) & university (specifics)	4
Government or ministry of education	7
University, faculty or department	12

Table 6. Empirical research to support required language level (*N* = 23).

		Number of contexts
No empirical foundation		19
Following other institutions	(6)	
Literature study	(6)	
Unknown	(6)	
CEFR	(3)	
Partly empirically founded		3
Needs analysis	(2)	
Expert counsel	(1)	
Empirically founded		1

This does not imply that the tests themselves are not based on empirical studies. Some tests are based on extensive research and needs analysis, but universities are free to decide on the required entrance level themselves. In many cases universities do not seem to base their requirements on needs analyses but on common practice or on what their competitors do.

R30: There is empirical research and there was research done when the test was first being designed and when the test was being trialed. [...] Obviously, the official advice is that [people from the accepting institution] should sit down and evaluate the courses, say for a course of that nature, we need a minimum standard of whatever. Now, some universities have done that kind of thing, but others are essentially responding to the market in terms of seeing what other people are asking for. [...] In the sense of actually doing that formal sit down, standard setting session, I think only a limited number have done that. [...] I would be happier if I believed that universities were really doing a standard setting, rather than simply responding to what their competitors and peers were doing.

In many contexts CEFR requirements depend not on needs analyses but on financial considerations. Some universities require a CEFR level that is lower than what is required in other universities to attract students (R30 above, and R13, R15 below). In at least eight contexts surveyed it is common practice for universities or ministries to use CEFR levels to control the flow of incoming students for financial reasons, to attract a certain type of students, or to control access to the labor market. Level requirements would rise and fall, not because the language needed to participate in academic life changes but because certain faculties or policy demands require more or fewer students.

R15: In some universities in the south part where the students are very few, they provide very easy entrance tests [...] while in the north part we have more difficult tests.

R13: Universities can decide to lower the level [...] if they desperately need students.

R4: The situation for international students is different depending on which country they come from. If they are from the European Union they are not obliged to have B2.

R6: The C1 level [...] that's the normal thing. But we now have a special case: medical doctors and nurses in health care that were educated in their homeland. [...] They informed us from the ministry, that they needed people within the health system and they found out that the C1 level, it was too difficult to pass.

R9: We have an institute [of higher education] working with oil and there we don't ask any level, because we need people working with oil.

Of the 30 respondents involved in this study, 27 worked on CEFR-related language tests. Most of these respondents stated that the CEFR was used to set or define the level (20), and/or to draw up rating scales (4) and/or to design task specifications (4). Two respondents state that their tests were fully based on the CEFR, and for most respondents, the CEFR is part of their daily practice:

R2: It's always in the background.

R15: Every day I consider the CEFR.

R28: Everything we do is based on the levels.

R30: We refer to it all the time.

Twenty-nine respondents have positive opinions about the CEFR. They refer to it as a standard you can trust, a useful tool, and an instrument that has changed the way they think and talk. The CEFR's most frequently mentioned strengths are providing a common language (7), facilitating comparison between tests (6), and offering a common standard (7). Frequently mentioned weaknesses are vagueness (16), normative use (5), and being outdated (5). When respondents were asked what they would like to change about the CEFR, most suggestions dealt with additions, such as including special purpose language use, digital communication, insights from recent CEFR studies, and language-specific descriptors (Table 7). No comments were made about any other aspect of the CEFR than the levels and the illustrative scales.

The respondents mention three main effects of the CEFR on language testing in Europe. First, they feel that the CEFR has brought standardization where there was disharmony. Second, the CEFR has promoted skill-based language testing. Third, it has led language tests to adopt a new level structure, which is now so well established that test developers may experience pressure to align to it. Seventeen respondents have experienced an external pressure to align to the CEFR, either politically or economically.

R2: It was in fact the demand [of the funding body] to take into account the CEFR. So because of this, we took the B2 level as a starting point for developing our exam.

R10: The CEFR levels are not precise enough, but if you want to sell an examination or a textbook without indicating the levels, then it will be very difficult on the market, so this is why most of the institutions use the CEFR levels when they want to sell something.

R18: To pass our B2 exam you need to be quite good, but there are other exams which aren't as complete or as strict in their marking and then you get a student who is at a lower level but still gets a B2 certificate and would not manage to pass our exam. So, market-wise, it's creating a bit of a problem here. [...] The product has to be marketable, you know?

Table 7. Most frequently suggested changes to CEFR.

Add to CEFR	25	Special purpose descriptors
		Insights from recent empirical studies
		Digital communication
		Language-specific descriptors
Change in CEFR	18	Revise level descriptors
		Language-specific descriptors
		Work on text/translation quality
Research	5	Validate CEFR scales

There are established procedures for aligning tests with the CEFR, but not all test developers may have linked their tests to the CEFR using a robust procedure. Respondents of 11 contexts spontaneously mentioned examples of unfounded CEFR links in university entrance tests used in their context. In some cases there was no procedure at all.

R12: We had tests before, I have all the presidential decrees here, and now the announcement is that it is CEFR compatible in all levels [...] So these four levels became six now, according to the CEFR framework.

I: Where does [normative CEFR use] come from you think?

R30: Cause it's easy I suspect. "Ooh it's there, it's in the book. Yes! B2 is very good, ooh that's fine, yes!" Boom, job done. I think there's a lot of that.

R18: It's not the CEFR descriptors which are creating a problem, really. It's those who are interpreting. [...] Everybody else is issuing certificates at a B2 level, but is there any guarantee that they're actually at that level?

Linking a test score with an external measure, such as the CEFR, requires a standard that is somehow fixed or uniformly interpretable. Even if all tests link to the CEFR in the most thorough ways, its levels could only serve as a common currency when the same levels roughly mean the same thing to different users. Only two respondents unequivocally assumed that the B2 level is operationalized uniformly throughout Europe.

Discussion

How uniform are the European university entrance language requirements for international L2 students?

There is no uniform policy within Europe when it comes to language requirements for foreign L2 students who wish to study in a national language. In eight contexts surveyed there is a national regulation that specifies the official university entrance language requirement for L2 students. In the other cases there is either no language requirement whatsoever or each university sets its own language level requirements. This diversification is reflected in the testing policy. In 20 contexts multiple language tests are accredited for granting university access, and in most of these cases centralized tests are accepted alongside local tests that have been developed by the accepting institutions themselves. Many respondents doubted the quality and comparability of these local tests, and if there is one common wish the respondents share for the L2 university entrance policy in their context, it is increased standardization.

Throughout Europe universities have a lot of autonomy in setting the entrance requirements. Usually, they are free to determine the language proficiency level they require for admission, although in some contexts the government may set some general requirements. In seven contexts, universities have no autonomy in deciding the required language level and are obliged to follow governmental decrees. Irrespective of which body is responsible for determining the entrance level however, the reasons for choosing a certain language requirement appear to be quite unfounded. In only one of the 23 contexts where university entrance is conditional on CEFR-linked language requirements, the required level is based on an empirical study. In 22 others, those requirements are not or only partly based on an analysis of the target language use context. Most often the language level requirements are determined by what other countries or universities do, or by the text of the CEFR itself. Moreover, in about one-third of the contexts surveyed it is not uncommon for institutions to lower the linguistic entrance requirements to attract more students or to manage

the access of students with less desirable profiles. Because many institutions can decide on the entrance level they require, they are both policy makers and stakeholders, which invites a situation where economic considerations may overrule actual student needs.

What is the impact of the CEFR on university admission policies and tests in Europe?

How do test developers perceive the CEFR and its impact?

This study shows that the CEFR has fundamentally impacted university entrance language testing in Europe, and the most influential aspects of the CEFR are the six levels and the illustrative scales. In just one of the contexts surveyed the university entrance language requirements are not CEFR related. This study confirms that the level most commonly used for university admission in Europe is B2, even though the respondents were not convinced that B2 is operationalized in the same way throughout Europe. Half of the respondents did not feel comfortable considering B2 the default starting requirement for university. All in all, however, the respondents of this study are rather positive about the CEFR, because they feel it offers a common—albeit sometimes vague—standard that has improved test score comparability.

Respondents generally feel positive toward the CEFR and often use it in a prescriptive way. Test developers rarely design university admission policies though, and when decision makers are also stakeholders, misuse is not uncommon. The results of this study show that sometimes CEFR levels are interpreted very rigidly in a process mirroring the reification Fulcher (2004) warns against: B2 is used because it is B2, not because it is the level that best suits the user needs. Sometimes the entrance level requirements are based on what other universities demand, and in some cases CEFR levels are used to manage the flow of L2 students by adjusting the demands for specific groups of students.

The data in this study offer little support to the claim that the CEFR stimulates discussion between decision makers and language experts (North, 2014a). No respondent mentioned the CEFR as a catalyst for conversations between policy makers and test developers, and only one respondent claims that the university entrance level in their context was based on a needs analysis. In the short term this kind of CEFR misuse is unfair—because it does not offer every student equal opportunities across contexts—or irresponsible—because it ignores user needs or target language use demands in favor of norm-driven labeling. In the long run it could prove potentially destructive for the CEFR, because the levels might lose credibility. The CEFR has always been an open source hermeneutic, but in many contexts it now serves as a self-administered seal of quality. It can give university admission officers a semiobjective tool to control university entrance, and it may allow test developers to claim a link to a certain level without having to offer any kind of proof for this.

In recent publications it has been pointed out that the CEFR was not intended as a prescriptive standard and that any normative application qualifies as unintended use (e.g., North, 2014a, 2014b). The current study shows that the use of CEFR levels as standards qualifies as the de facto norm in university entrance language testing, even though this is not what the authors of the CEFR intended. Testers need standards, cutoff points and norms, and if the market or the policy requires the use of a certain framework, normative use is not illogical. Unintended use is different from misuse however. When the CEFR is used to give a pretense of objectivity to an otherwise arbitrary decision, it is misused, because it serves a purpose inimical to valid, fair, and just language testing. Discussion between test developers and policy makers could perhaps serve to limit the misuse of CEFR levels, but as was stated above, the data in this study do not indicate a vibrant debate between test developers and the people who decide on admission criteria.

Conclusion

Fifteen years after its first publication, the CEFR has fundamentally altered language testing in Europe, and its six levels have been widely accepted by all stakeholders, from policy makers to candidates. The CEFR has had a huge impact on university entrance policies and tests across Europe,

but its popularity also has a downside: there is substantial pressure on test developers to align with CEFR levels. In many contexts this pressure has led to CEFR misuse, such as claiming to test at a certain level without substantiating this claim, or using the CEFR to manage the flow of students. In almost all contexts surveyed the CEFR is used normatively by test developers, admission officers, or both.

If the CEFR is to fulfill its purpose of achieving transparency in language teaching and promoting international educational mobility, two important changes should be considered. First, there seems to be little merit in perennially rejecting normative use of the CEFR as unintended, when normative use is so widespread. The CEFR is owned by its users, not by its authors, and research would benefit from acknowledging normative CEFR use and differentiating between unintended—normative—use and misuse. Second, CEFR labels are very easily tacked onto a test or written into a policy, whether with or, as often happens, without any research-based arguments. This may over time endanger the CEFR's credibility and usefulness (Hulstijn, 2015). Decision making would benefit from a realistic assessment of the CEFR's limitations in this respect, recognizing it as a general theoretical framework that needs to be supplemented by language-specific and context-specific descriptors. This does not imply that linking a test to the CEFR is futile. On the contrary: CEFR linking provides a reference point, but without the assumption of exact equivalence between all tests at the same level. Furthermore, linking may provide valuable insights into those characteristics of the test that are language-specific and context-specific. Results of such a linking should not be regarded as the one truth but as one of many valuable aspects of validation.

Limitations and further research

This exploratory study was intended to verify a number of commonly held assumptions about the CEFR in Europe and to offer a state of affairs regarding the influence of the CEFR on university entrance policies throughout Europe. The results presented here were accurate when the manuscript was submitted (January 2016), but policies may change from year to year.

The respondents of this study are for the most part developers of centralized tests, and their perspective may differ substantially from that of admission officers, policy makers, test takers, or developers of small-scale local tests. The data cover many countries in Europe, but given the number of respondents, generalizations should be treated with some caution. Further research could flesh out trends discussed here, determine whether they would be the same for other stakeholders in the university entrance policy, and look at specific contexts or take some preliminary conclusions a step further.

References

Alderson, J. (2007). The CEFR and the need for more research. *The Modern Language Journal*, *91*(4), 659–663. doi:10.1111/modl.2007.91.issue-4

Alderson, J., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (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: An International Journal*, *3*(1), 3–30. doi:10.1207/s15434311laq0301_2

Carlsen, C. H. (2014, April 10-11). *How valid is the CEFR as a construct for language tests?* Presented at the ALTE 5th International Conference, Paris, France.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Strasbourg, France: Council of Europe.

Curtis, A. (2015, March 21). Discussant at the ILTA-AAAL joint event: Revisiting the Interfaces between SLA and Language Assessment Research, Toronto, Canada.

Deygers, B. (2017). University entrance language tests in Flanders: Examining assumed equivalence. In J. Davis, J. Norris, M. Malone, T. McKay, & Y. A. Son (Eds.), *Useful assessment and evaluation in language education.* Washington, DC: Georgetown University Press.

EACEA. (2010). *Focus on higher education in Europe.* Brussels, Belgium: Education, Audiovisual and Culture Executive Agency.

Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66(4), 477–485. doi:10.1093/elt/ccs037

Freeman, M. (2000). Knocking on doors: On constructing culture. *Qualitative Inquiry*, 6, 359–369. doi:10.1177/107780040000600305

Fulcher, G. (2004). Deluded by artifices? The common European framework and harmonization. *Language Assessment Quarterly*, 1(4), 253–266. doi:10.1207/s15434311laq0104_4

Fulcher, G. (2012). Scoring performance tests. In G. Fulcher, & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 378–392). London, UK: Routledge.

Galaczi, E., Ffrench, A., Hubbard, C., & Green, A. (2011). Developing assessment scales for large-scale speaking tests: A multiple-method approach. *Assessment in Education: Principles, Policy & Practice*, 18(3), 217–237. doi:10.1080/0969594X.2011.574605

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine de Gruyter.

Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing*, 17(4), 228–250. doi:10.1016/j.asw.2012.06.003

Hulstijn, J. (2007). The Shaky ground Beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663–667. doi:10.1111/modl.2007.91.issue-4

Hulstijn, J. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam, The Netherlands: John Benjamins.

Little, D. (2007). The common European framework of reference for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, 91(4), 645–655. doi:10.1111/modl.2007.91.issue-4

Martyniuk, W. (2010). *Studies in language testing 33: Aligning tests with the CEFR. Reflections on using the council of Europe's draft manual*. Cambridge, UK: Cambridge University Press.

McNamara, T. (2007, June 15–17). *The CEFR in Europe and beyond*. Paper presented at the 4th Annual EALTA Conference, Sitges, Spain.

Miles, M., & Huberman, A. (1994). *Qualitative data analysis*. Baverly Hills, CA: Sage.

Negishi, M., Takada, T., & Tono, Y. (2013). A progress report on the development of the CEFR-J. In E. Galaczi, & C. Weir (Eds.), *Exploring language frameworks. Proceedings from the ALTE Kraków conference, July 2011* (Vol. 36, pp. 137–165). Cambridge, UK: Cambridge University Press.

North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, 91(4), 656–659. doi:10.1111/modl.2007.91.issue-4

North, B. (2014a). *The CEFR in practice. English profile studies* (Vol. 4). Cambridge, UK: Cambridge University Press.

North, B. (2014b). Putting the common European framework of reference to good use. *Language Teaching*, 47(02), 228–249. doi:10.1017/S0261444811000206

Papageorgiou, S. (2010). Investigating the decision-making process of standard setting participants. *Language Testing*, 27(2), 261–282. doi:10.1177/0265532209349472

Porto, M. (2012). Academic perspectives from Argentina. In M. Byram, & L. Parmenter (Eds.), *The common European framework of reference: The globalisation of language education policy* (pp. 129–138). Bristol, UK: Multilingual Matters.

Reybold, L., Lammert, J., & Stribling, S. (2013). Participant selection as a conscious research method: Thinking forward and the deliberation of "Emergent" findings. *Qualitative Research*, 13(6), 699–716. doi:10.1177/1468794112465634

Roever, C., & McNamara, T. (2006). Language testing: The social dimension. *International Journal of Applied Linguistics*, 16(2), 242–258. doi:10.1111/ijal.2006.16.issue-2

Salamoura, A., & Saville, N. (2009). Criterial features across the CEFR levels: Evidence from the English profile programme. *Research Notes*, 37, 34–40. Cambridge: Cambridge ESOL.

Schwartz, N., Knäuper, B., Oyersman, D., & Stich, C. (2008). The psychology of asking questions. In E. De Leeuw, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (1st ed., pp. 18–35). New York, NY: Routledge.

Shohamy, E. (2011, July 7–9). *An engagement with the CEFR. A critical view and time*. Paper presented at the ALTE 4th International Conference, Krakow, Poland.

Stæhr, L. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139–152. doi:10.1080/09571730802389975

University of Cambridge, ESOL Examinations. (2011). *Using the CEFR: Principles of good practice*. Cambridge, UK: University of Cambridge, ESOL Examinations.

Van Avermaet, P., & Pulinx, R. (2013). Language testing for immigration to Europe. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 376–389). Malden, MA: John Wiley & Sons, Inc.

Van Ek, J. (1975). *Systems development in adult language learning: The threshold level in a European-unit/credit system for modern language learning by adults*. Strasbourg, France: Council of Europe.

Van Ek, J., & Trim, J. (1991a). *Threshold 1990*. Strasbourg, France: Council of Europe.

Van Ek, J., & Trim, J. (1991b). *Waystage 1990*. Strasbourg, France: Council of Europe.

Van Ek, J., & Trim, J. (2001). *Vantage*. Cambridge, UK: Cambridge University Press.

Weir, C. (2005). Limitations of the common European framework for developing comparable examinations and tests. *Language Testing*, *22*(3), 281–300. doi:10.1191/0265532205lt309oa

Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *Forum: Qualitative Social Research, 3*(2). Retrieved July 30, 2015, from http://www.qualitative-research.net/index.php/fqs/article/view/865

Xi, X., Bridgeman, B., & Wendler, C. (2014). Tests of English for academic purposes in university admissions. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 318–333). Malden, MA: John Wiley & Sons, Inc.

Ziegler, N., & Kang, L. (2016). Mixed methods designs. In A. J. Moeller, J. W. Creswell, & N. Saville (Eds.), *Second language assessment and mixed methods research* (pp. 51–84). Cambridge, UK: Cambridge English.