

University of Groningen

Embedding the CEFR in the Academic Domain: Assessment of Language Tasks

Lowie, W.M.; Haines, K.; Jansma, P.

Published in:
Procedia: Social and Behavioral Sciences

DOI:
[10.1016/j.sbspro.2010.07.027](https://doi.org/10.1016/j.sbspro.2010.07.027)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Lowie, W. M., Haines, K., & Jansma, P. (2010). Embedding the CEFR in the Academic Domain: Assessment of Language Tasks. *Procedia: Social and Behavioral Sciences*, 3, 152 - 161.
<https://doi.org/10.1016/j.sbspro.2010.07.027>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Telling ELT Tales out of School

Embedding the CEFR in the academic domain:

Assessment of language tasks

W.M. Lowie^{a*}, K.B.J. Haines^a, P.N. Jansma^a

^a*University of Groningen, Faculty of Arts, Department of English, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, Netherlands*

Abstract

The CEFR is becoming established as the framework of choice for the description of the language proficiency of students at universities across Europe. To attain consistent assessment, familiarization and standardization are crucial prerequisites. In this paper we report a case study of embedding a standardization procedure in writing assessment activities at the University of Groningen. In spite of the difficulty to find consistently assessed ‘flat’ samples, which is inherent to the nature of language proficiency, the project shows the value and the feasibility of standardisation procedures within the CEFR.

© 2010 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: CEF; Common European Framework; academic domain; language tasks; assessment

1. Introduction

The Common European Framework of Reference (CEFR, Council of Europe, 2001) has contributed to the innovation of language teaching in many important ways. Focusing on providing language learners a tool for (self-) evaluation, it has promoted a major change in the perception of language proficiency from a teacher oriented product to a learner-oriented process. To evaluate the different aspects of language proficiency, the CEFR distinguishes 34 scales for listening, reading, speaking and writing, which cover both general competences and communicative language competences for a range of domains to define six levels of proficiency (Little, 2007). These levels were validated in both quantitative and qualitative studies (Alderson, 2002; Hasselgreen, 2003).

In spite of the validation of the proficiency levels for the different skills, the descriptors defining these levels have also been criticized. Hulstijn (2007) discusses the extent to which the descriptors accurately reflect the nature of language proficiency and wonders if quantitative (how much can the learner do) and qualitative (how well can the learner do this) characteristics can be objectively combined in one set of descriptors. He points to the fact that the CEFR is based on unidimensional scales of language proficiency (also see North & Schneider, 1998), while

* Corresponding author: W. M. Lowie. Tel.: +31-(0)50-363-5807.

E-mail address: w.m.lowie@rug.nl.

language proficiency is more likely to be multidimensional. Although the CEFR allows for learners to have a B1 speaking level and a C1 writing level, the qualitative and quantitative aspects within these levels may not be adequately distinguished. This distinction can be expected to play an important role in the assessment of proficiency levels in the academic context, where qualitative aspects of language proficiency are valued at least equal to quantitative aspects.

The current paper reports on a project aiming at the standardisation of the assessment of written production in the academic context. This project resulted from the innovation of the language teaching curriculum in agreement with the CEFR guidelines. The aim of the project was to produce a set of writing samples that are representative for the different components of the CEFR levels in the local context of our university, which could then be used for standardization of writing assessment. Although this project was not set up as an empirical evaluation of the CEFR, attempts to embed the CEFR into the day-to-day teaching practice serve as a relevant verification of its feasibility.

1.1. *The Embed Project*

Over the past five years, the CEFR has been embedded in the assessment procedures for English courses at the University of Groningen. Initially, student awareness of their level of proficiency in English has been promoted through the use of Dialang (www.dialang.org). More recently, teachers have been encouraged to use Cefrain (www.helsinki.fi/project/ceftrain) and Webcef (www.webcef.eu) alongside Dialang in order to familiarise themselves with the levels and criteria for assessment contained in the Framework. In the past few years, internal projects at the faculties of Arts, Law and Economics & Business have succeeded in increasing the profile of the CEFR and demonstrated the value of its structural use in a greater number of Major and Minor courses. The current paper describes one such project, *Embed*, in which teachers at the faculties of Law and Arts have cooperated in the production of structured online CEFR familiarisation and standardisation resources for teachers of English courses that train the development of academic writing skills.

The *Embed Project* aims to “increase the transparency for teachers, testers, examination users and test takers about the content and quality of the examination or test and its relationship to the CEFR”, as expressed in the 2009 Council of Europe manual (hereafter referred to as *A Manual*) (2009: 27). In an international academic environment, it is common to have a heterogeneous team of English teachers drawn from a wide variety of backgrounds, drawing on a variety of assessment cultures. It is a great advantage under such circumstances that students’ writing skills can now be assessed using the CEFR descriptors for Range, Accuracy, Coherence provided in Table C4 of *A Manual* (2009: 187). Alongside these generic linguistic categories, we have also used the task-specific descriptors for essay and report writing (CEFR 2001: 62).

The advantage in using the CEFR is that it provides a single common structure upon which we can base our interpretations of the linguistic performance of students. This structure has the further advantage that it has been developed empirically by groups of experts in the language testing field, as described in the 2001 publication. This framework is supplemented by tools, such as Cefrain and Webcef, which promote good practice in assessment by encouraging teachers to go through rigorous procedures when assessing language proficiency. These tools represent a concerted effort by experts from a number of prominent European universities to apply the common framework to everyday assessment practice. Finally, a manual has been developed with the involvement and feedback of experts and practitioners throughout Europe (see *A Manual* 2009: ix), which guides towards good practice when applying the CEFR to tasks embedded in our curricula. It does this through task specification, familiarisation exercises, and standardisation sessions, the documented use of which constitute evidence of a claim to have based assessments on the CEFR. Our use of the 2009 Manual during the *Embed Project* has provided us with local evidence to support such a claim.

Despite the above advantages, we have come to recognise that limitations in the CEFR, when combined with the practical constraints of the university environment, will often produce a far from perfect result. Firstly, the CEFR levels are described in general terms, using phrasing that leaves much room for interpretation by individual assessors. Also, the levels are very broad and thus academic writing may vary widely in terms of its linguistic quality even within a certain overall band i.e. there is a considerable difference between what we may consider a ‘high C1’ and a ‘low C1’. Furthermore, although the CEFR manual consistently encourages us to benchmark local samples to external illustrative samples, only a limited number of such samples are actually available that are relevant both to the level of our students (generally B2/C1) and to the genre and domain of their academic work (for

instance, Business Studies, Law etc.). To the above limitations, we should add local constraints: these include complex teaching schedules that mean our teachers are not always available for training sessions and high turnover amongst teachers, particularly freelancers, which produces inconsistencies in the extent of familiarisation and standardisation within the teaching teams.

The above factors result in a relatively high threshold for any teacher applying the CEFR to the written work of students. Through the *Embed Project*, we are seeking to maximise the potential of the CEFR as a structure through which we can make language assessment in university courses more consistent. We aim to produce an easy-access information site or CEFR Toolkit for coordinators and teachers, based on the procedures for familiarisation and standardisation presented in the Manual, together with locally produced samples of academic work from a variety of domains. Firstly, course coordinators will have access to a dedicated structure for managing the standardisation of assessment during courses. This structure will be recognisable to their teachers as samples with immediate relevance to the local domain. Secondly, teachers will have immediate access to specific information on the CEFR as the use of meta-data techniques will allow them to search samples for local written evidence of language that represents specific levels for each of the criteria described in the CEFR (for instance, Accuracy at level B2). Thirdly, as a spin-off of this project, students will have access to indicative samples to help them understand the qualitative differences between the levels.

By embedding the CEFR in the curriculum, the *Embed Project* aims to make the CEFR meaningful in the local context. Teachers have been involved in the project from an early stage through the development of local samples or in familiarisation and standardisation training sessions. These teachers have consequently developed a sense of the value and the limitations of the CEFR and are in a better position both to apply it and to offer support to less experienced colleagues. Furthermore, this Toolkit will have longevity as it represents a blueprint from which coordinating teachers can develop CEFR resources in other local domains. For instance, whereas the sample materials in the resource bank initially relate to language used in the field of Law, coordinators may extend the procedures described below to produce similar resource banks in other domains. We also believe that our approach can provide a framework for similar developments in other institutions and indeed for other languages.

2. Procedures

In the past, we attempted to make the CEFR descriptors more relevant to our local academic domain by revising and extending the descriptors. This process aimed to make the descriptors more meaningful to both teachers and students by making the link to tasks carried out in the academic environment more explicit. We drew on the recommendation in the 2003 manual that if “the rating grid (Table 5.5) seems inappropriate for samples (e.g. vocational context, narrowly defined task)” then we could possibly “revise the grid, consulting the CEF scales to do so” (2003: 88).

Ultimately, however, despite ‘trying earnestly’ in the words of one teacher to produce a version of the CEFR descriptors that was more relevant to our practice, we concluded that the road ahead could not lie in adapting the descriptors themselves. The result of such an approach was an adapted set of descriptors that were more meaningful to their creators, but arguably no more meaningful to other teachers and equally open to interpretation, which in turn affected the validity of the assessment tool. In fact, we had effectively reduced our claim to be relating our assessment to the CEFR, and we consequently referred to these revised descriptors as ‘adapted from the CEFR’. Eventually, we decided to return to the drawing board and seek another approach.

The publication of the 2009 Manual provided us with a new impetus. We have found *A Manual* (2009) significantly clearer than the 2003 version, particularly in its explanation of the value of Task Specification. We value the inclusion in *A Manual* of the CEFR Content Analysis Grids for Writing and Speaking Tasks (Appendix B2). Not only has this appendix drawn our attention to the need to tightly specify the tasks at the core of our assessment procedures, but it has also given us a relatively user-friendly mechanism through which we could reach specifications of tasks in direct relation to the levels described in the CEFR. This has proved invaluable as a starting point in our collection of suitable local samples.

At the core of the current project lies the collection of local samples of essays produced by students in the faculties of Arts and Law. This project has focussed on levels B2, C1 and C2 for writing, as these are the levels of the majority of students taking our courses. Dutch students generally enter university with at least a B2 level of English (Fasoglio et al., 2007) while non-Dutch students are required to have an IELTS score of 6.5 with no

individual element below 6.0. In practice, this often means that they have a 6.0 (approximately B2) for writing and compensate with higher scores for the receptive language skills. The level required for students to pass English in the Law programme is C1, which means the qualitative distinction between B2 and C1 is particularly relevant in our context.

We will describe below the procedures we have followed in producing samples for Law, making explicit references to the CEFR Manual to demonstrate our adherence to procedures described there as further evidence of our claim to be relating our curricula to the CEFR. Our procedures essentially represent a local response to the following question at the end of chapter 5 of *A Manual*: “how to ensure that such “local” material for benchmarking (and future training) purposes is of good quality” (2009: 54).

2.1. Task Specification

Our first step in the process of developing local samples was task specification. We used the Content Analysis Grid for Writing Tasks (section B2 of Appendix B in *A Manual*) to produce a description of the task, which “offered us ‘the possibility to work at the more detailed level of individual test tasks, classifying them by standard criteria’” (*A Manual* 2009: 28). We were later able to supply this information in an easily accessible summarised form to our panel so that they would understand the context and conditions under which the writing task had been performed (see appendix 1).

2.2. Screening

Our next step was to screen an initial bundle of 120 essay samples from the Law course. *A Manual* recommends that the first illustrative examples used in the initial stage of standardisations “should show a performance with a relatively ‘flat profile’ across the categories of Table B4 (namely a writer who is, for example, B1, on all three categories Range, Coherence and Accuracy, and equally good at description and argument)” (2009: 45).

We believed that it would be challenging to identify such flat samples, as our experience with learners is that their language profiles will be spiked in some way due to the naturally uneven progression of different criteria in the development of language. Moreover, the descriptors do not distinguish qualitative and quantitative characteristics of language proficiency. Therefore, our aim was to locate authentic samples with a flat profile, and we initially screened the Law essays in order to produce a number of samples at each of the relevant levels (B2 and C1) that had originally been assessed by course teachers as flat in profile, producing a manageable bundle of seven texts for our local panel.

2.3. The Local Panel

Once we had identified a manageable bundle of samples, the next step was to set up a panel of experts. We set up a local panel consisting of five English teachers who have been involved in using the CEFR to assess the academic writing of students on a variety of courses since approximately 2005. Amongst these panelists are two Cambridge ESOL examiners, who bring with them external experience of familiarisation and standardisation procedures.

In order to check the interpretations of our panel, we also employed a local consultant, who is a former teacher with many years of experience in assessing students on the programmes from which the sample tasks were taken. This consultant is also a Cambridge examiner of many years standing and has been involved in the feedback process to the Language Policy division of the Council of Europe on the 2003 Manual. The role of this local consultant has been to act as a sounding board during the development of our procedures and to give a second opinion on the levels assigned by the panel to specific samples.

The role of the local panelists was to independently assess the level of each of the seven samples that had emerged from the initial screening. For this purpose, we developed an *Embed Project* Blackboard site which would also later be accessible to external assessors. The panelists were asked to assign a level to each sample for Range, Coherence, Accuracy and Argument. They were also asked to assign an overall level for each text and encouraged to add comments to justify their judgments, giving us the opportunity to evaluate the reasons for different interpretations.

In our project team, we next analyzed the interpretations of the local panelists and consequently discarded the three samples for which the assessments proved least consistent. The assessment by the local panel therefore resulted in four Law essay samples (samples 2, 3, 5 and 6), each with a relatively flat profile. Rather than weighing the different components against one another, the decision was reached by consensus. In Appendix 2, we have provided an excerpt from one of our samples (sample 2) so to give a fuller impression of the type of academic text under consideration. The panelists were in broad agreement regarding CEFR levels for each of the standard criteria (Range, Coherence, Accuracy and Argument) and the comments of the local consultant were broadly supportive of our local interpretations.

2.4. Trial Standardisation

Once the local panel had produced its four samples, we trialed the use of these samples in a standardisation session for the teachers in the English for Law course. This session formed part of the regular training of teachers which is central to our implementation of the CEFR in various faculties at the University of Groningen. It is usual for local samples for such standardisations to be produced by a course coordinator in cooperation with the in-house CEFR trainer, and with reference to the external samples we use as benchmarks. However, the immediate impact of this project was that we were using samples that had been produced through a panel, as recommended in *A Manual*.

In order to further evaluate the use of these samples in our standardisation procedures, we trialed the samples in a standardisation session with five teachers of academic writing for students of Law. Following procedures described in *A Manual* (2009), we first established an external benchmark by asking the teachers to follow the CEFR familiarisation procedures provided on the Cefrain site. We also asked them to assess writing samples 2 and 7 in the training section of that site, which we considered closest to the levels and type of writing being produced by our own students. Teachers were then asked to assess each of the four *Embed* samples and for each sample they were asked to compare their assessments in pairs, defending their interpretation by relating features in the text to the CEFR descriptors. We then opened up any significant areas of difficulty or interest produced in the pairs to the whole group.

For two of the samples, we next provided the group of teachers with raw commentaries that had been produced using the comments of the local assessment panel. The teachers were given the opportunity to compare the comments of the local panelists with their own comments in the paired and group discussions. On the one hand, this served the purpose of standardisation by helping to resolve individual deviations from the general interpretations, while on the other hand it allowed us to fine-tune the commentaries.

2.5. The External Panel

Since *A Manual* recommends “that it is a good idea to include panelists external to the institution producing the test in question” (2009: 38), our final step was to set up an external panel of experts, in order to check whether the judgments of our panel were broadly in line with interpretations elsewhere in Europe. This consisted of assessors at universities from our network of partner universities in Italy (Padua), Finland (Helsinki) and Belgium (Leuven). Our criterion for selection of these external assessors was that they should be experienced in the use of the CEFR in academic writing in contexts directly comparable to ours. One of the assessors in Helsinki and the assessor in Leuven deliver English for Academic Purposes (EAP) courses to students of Law. The other assessors are experienced in the application of the CEFR descriptors to academic writing in other areas of the university curriculum. These experts assessed the four samples which had produced the greatest consensus amongst our panelists, using the same procedure embedded in Blackboard. The findings of all the panelists are discussed and compared in more detail in section 3 below.

3. Results

After making adaptations on the basis of the trial standardisation, we made the four samples available to the external panelists in Helsinki, Padua and Leuven through our Blackboard system. The tables below show for each of the four remaining Law essay samples the extent to which the interpretations of the external consultants in Finland and Italy compared with the overall interpretations of our local panel.

3.1. Law Sample 2

The local panel independently assessed Sample 2 (Table 1) as a good representation of level C1. Although some panelists noted elements of B2 for Coherence, this was not the majority view, and the local consultant also stated that she felt B2 was too low an assessment for Coherence. Interestingly, the external panelists vary between C1 and C2 for all criteria, with one exception for Accuracy. In fact, the majority Overall interpretation for the externals was C2, with an emphasis in their comments on the relevance and variety of usage. However, *Helsinki 2*, whose assessment largely coincides with that of the internal panel, notes the inconsistency of the text. Indeed, this inconsistency in the text may go some way towards explaining the varying assessments. We need to consider what consequences inconsistency has for the use and presentation of a sample. For instance, we may wish to present certain extracts to teachers but not others.

Table 1: CEFR interpretations for sample 2

Sample 2	R	C	A	Argument	Overall
Local assessors	C1	B2/C1	C1	C1	C1
Local consultant	C1	C1	C1	C1	C1
External Helsinki 1	C2	C2	C1	C2	C2
External Helsinki 2	C1	C1	B2	C1	C1
External Padua	C1	C2	C2	C2	C2
External Leuven	C2	C2	C2	C2	C2

3.2. Law Sample 3

The local panel assessed Sample 3 (Table 2) as a good representation of level C2 for all aspects of Language, but the level for Argument was consistently lower. The point was made during the Law trial standardisation that it may be necessary to use different samples for the Argument column, which effectively relates to task completion. It seems quite reasonable to conclude that students may produce a general level of language at C2 but prove unable to organise their ideas into a structure of the same level. In this example, the local consultant has pointed out that the text may be too short to make a full assessment on its structural organisation, and indeed both the internal and external panelists comment on the lack of examples (*Helsinki 2*: “the writer states that the world has changed, but gives no proof of this”).

Table 2: CEFR interpretations for sample 3

Sample 3	R	C	A	Argument	Overall
Local assessors	C2	C2	C2	C1	C2
Local consultant	C2	C2	C2	too short?	C2
External Helsinki 1	C2	C1	C2	C1	C1
External Helsinki 2	C2	C2	C2	C1	C2
External Padua	C2	C1	C1	C1	C1
External Leuven	C1	B2/C1	C1	B2.2	B2

3.3. Law Sample 5

The local panel assessed Sample 5 (Table 3) as a good representation of level C1. This was generally confirmed by the interpretation of the external assessors in relation to Range. However, there was significant variety in the

assessments for the other criteria, ranging from B2 to C2 for Coherence and Accuracy. This shows how a sample may fail to function, leading to widely different interpretations as, for instance, one assessor refers to the text as ‘perfectly supported by a sufficient number of discourse markers’ while another describes its ‘limited number of connectors’. This text might function as a sample for Range, but it would hardly serve this function for Coherence.

Table 3: CEFR interpretations for sample 5

Sample 5	R	C	A	Argument	Overall
Local assessors	C1	C1	C1	C1	C1
Local consultant	C1	C1	C1	C1	C1
External Helsinki 1	C1	B2	B2	B2.2	B2
External Helsinki 2	C1	C2	C2	C1	C2
External Padua	C1	C1	C1	C2	C1
External Leuven	C2	C2	C2	C2	C2

3.4. Law Sample 6

The local panel assessed Sample 6 (Table 4) as a good representation of level B2. There was a large degree of consistency in the assessments made by the external panelists, with general agreement in all columns. All assessors agreed on the B2 level for Range, which is higher in general than for the other criteria. We wondered whether the descriptors are clearer or whether range is inherently less contentious. Only the assessor from *Padua* thought sample 6 was below B2 for Coherence, explaining that although there are many connectors, they are often used incorrectly. Despite these details, the interpretation of this sample showed the most consensus of all samples.

Table 4: CEFR interpretations for sample 6

Sample 6	R	C	A	Argument	Overall
Local assessors	B2	B2	B2	B2.2	B2
Local consultant	B2	B2	B2	B2.2	B2
External Helsinki 1	B2	B2	B1	B2.2	B2
External Helsinki 2	B2	B2	B1	B2.2	B2
External Padua	B2	B1	B1	B2.2	B2
External Leuven	B2	B2	B1	B2.1	B1

3.5. Discussion

The feedback provided by the external panel has helped us towards the solution to some specific difficulties that have become apparent during this project. Most notably, one finding of the Law standardisation was that it would be problematic to find a single sample that could be used to exemplify a certain level for all criteria. In particular, it is difficult to find a single authentic sample that will exemplify both the essentially linguistic (Range, Accuracy, Coherence) and the essentially structural (Description, Argument) features of a text. The feedback of the external panelists has caused us to take this conclusion even further, realizing that we may need to use a specific sample, or even extracts from that sample, to represent Range at a certain level of the CEFR, while a different sample might be needed to represent Accuracy at the same level.

The divergence in interpretations for some samples, even between assessors from the same institution, are inherent to assessment processes that involve individuals with differing perspectives and partly results from the unidimensional nature of the descriptors. An assessor makes literally hundreds of small judgments when assessing a single student essay, including both qualitative and quantitative aspects of language proficiency. This pilot only serves to confirm the extent of interpretation involved in making these judgments on the proficiency of student

writing in specific contexts, given the generic nature of the CEFR descriptors. This highlights the value of standardisation in the promotion of consistency amongst such assessors. As this project aims to produce workable samples for teachers, we have to work with, but not ignore, such inherent differences in interpretations.

This leads us to the conclusion that at the local level, where resources for the development of samples will inevitably be limited, it may be counter-productive to think in terms of specific samples that represent specific levels of the CEFR. It will even prove difficult, in relation to a complex academic task such as an argumentative essay, to find a sample that is a consistent representation of one criterion, such as Range, at any a given level. Indeed, through our rigorous procedure in this Law pilot, we have only been able to identify one criterion in one sample for which all interpretations; Range for Sample 6 is consistently assigned B2. We have noticed that the greatest consensus occurs for Range, so we can only imagine how much time and energy would need to be expended before ideal samples could be found for all the other criteria.

The solution therefore seems to lie in a slightly less idealistic approach: rather than offering teachers entire samples, we may decide to offer them excerpts from several samples drawn from a database which may expand over time. This would take into account the inherent inconsistency of learners, which is not only demonstrated across the range of criteria but is also seen for a single criterion within a given text. Writers do not readily produce authentic texts that match the confines of the CEFR descriptors; they may be producing Range at B2 for several paragraphs, showing a lack of ‘expressiveness and idiomaticity’, and then produce a paragraph at C1 level in which they do express themselves in an ‘appropriate style’ (see descriptors in Table C4, *A Manual* 2009: 187). This makes life complicated even for experienced assessors. For Coherence in Sample 3, the *Leuven* assessor writes: “Actually, my appreciation lies between B2 and C1”, noting the tension between the use of a variety of cohesive devices and the overly long sentences that result.

For less experienced assessors, perhaps confronted with the CEFR for the first time, the samples used in a standardisation procedure should help to establish the discreet differences between levels rather than the inevitable overlaps between them. Our next step is therefore to break our samples down into extracts that clearly represent the CEFR descriptors. We stress, however, that all such excerpts should be shown in the original text so that the teacher/assessor can see the relevance of the excerpt to the wider context. We also appreciate that this approach will not serve the purpose of standardisation for Argument, for which a sample will have to consist of an entire text.

4. Conclusion

Our aim from the start of the project was to produce a re-usable, extendable resource that coordinators of language courses could in future copy and then adapt in order to apply it to their specific domains, or indeed redesign for use with other languages. At the start of this project, we produced a Blackboard site containing reference and training materials and tools for assessment. In setting up the site and developing the procedures for assessment of student work, we followed the guidelines provided by *A Manual* (2009) as closely as possible. Our procedures have therefore strengthened our claim to be basing our assessment on the CEFR.

These initial results are based on a pilot during which we have used Law essay samples to trial our procedures and to experiment in the presentation of such local samples to our teachers. This procedure has confirmed the value of our approach in making a transparent link between local, domain-specific samples and the descriptors presented in the CEFR. We believe that by using local samples as evidence of students’ linguistic performance, and by describing these samples in CEFR terms, we have made the CEFR more applicable to the daily practice of our teachers. This potentially makes the link between the CEFR and course assessment more transparent to all parties, including the students themselves. At the same time, by providing the samples to the teachers as a web-based self-access resource, we have significantly lowered the threshold for their use, and hence we have increased the likelihood that the CEFR can be applied consistently within the English language components of academic programmes. The next step for this project is therefore to extend this procedure to essays produced by students in the Major and Minor programmes of the Faculty of Arts.

The question raised at the outset of this paper was if it would be possible to create a set of local, domain-specific writing samples that consistently represent the different levels of proficiency of our students. The stepwise procedure in the *Embed project* has shown that it has been difficult to find ‘flat’ samples, in which all components of the products would consistently be assigned the same CEFR level. Even after a careful preselection, followed by a further selection by experienced local assessors, the assessment of the remaining writing samples were not fully

consistent for all aspects of writing. Given the multidimensional nature of language proficiency, this is not very surprising. However, this does not mean that we should give up standardisation within the CEFR. The project has demonstrated that general agreement can be reached within the assessment of the different components of writing. The most important conclusion is that we should acknowledge the different dimensions of language proficiency. For standardisation, this means that we should focus on the assessment of the specific aspects of proficiency, giving explicit feedback on these and illustrate the rationale of the ‘expert’ choices with textual examples. From the learner’s point of view, this makes perfect sense. Different assessment levels for different aspects of language proficiency can help specify the learner’s strong and weak points.

To summarise, the *Embed Project* has developed and piloted a panel-based qualitative procedure for the development of task-specific local samples which can be embedded in the local curriculum through standardisation procedures such as those described in *A Manual* (2009). We are now extending our use of these procedures to other Major and Minor programmes in the faculties of Law and Arts at the University of Groningen. We also believe that the *Embed* procedure may be extended for use with other tasks and in other settings. Furthermore, if the project can be extended to include other universities in our network, it has the potential to function as the foundation of a significant database to support the consistent application of the CEFR within an institution or even across institutions with similar domains. Our involvement of assessors from partner universities in the validation of our samples has been one of the strengths of this project, and we believe it is worth reiterating that the aim of the CEFR is “to facilitate reflection, communication and networking in language education” (*A Manual* 2009: 3).

Acknowledgment

We are greatly indebted to the team members of the Embed Project at the University of Groningen: Sake Jager (project leader), Estelle Meima, Nicole Schmidt. Without their effort this paper would not have been possible.

References

- Alderson, J. C. (Ed.). (2002). *Case studies in the use of the Common European Framework*. Strasbourg, France: Council of Europe.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for languages: learning, teaching, assessment (CEFR): A manual*. Strasbourg: Language Policy Division.
- Fasoglio, D., Meijer, D. & Trimbos, B. (2007). *Handreiking schoolexamens moderne vreemde talen havo/vwo. Duits, Engels, Frans*. Enschede: SLO.
- Hasselgreen, A. (2003). *Bergen “Can do” Project*. Strasbourg, France: Council of Europe.
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language Proficiency1. *The Modern Language Journal*, 91(4), 663-667.
- Little, D. (2007). The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91, 645–655.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 217–262.

Appendix 1: summary produced using the CEFR Grid for Writing Tasks, *A Manual* (2009: 161-164)

Target level for task completion	C1
Task Purpose	To measure student ability to write an argumentative essay based on current world issues.
Background to task	Second semester writing exam Students had to write an argumentative essay at B2.2 level under exam conditions to pass semester one. They have been writing practice essays in class and receiving peer and teacher feedback on them.
Structure of task	A single writing task consisting of one 500- word essay, based on one of two preselected questions.

Specific information	The candidates are asked to write an essay to demonstrate their ability to formulate an argumentative thesis, support that thesis with appropriate arguments, refute an opposing argument and round off with an effective conclusion. The students will be assessed on their use of academic and legal language, use of formal register, their ability to link their ideas together in clear, smooth flowing text, and to maintain a high degree of grammatical correctness. The students must achieve an average of a C1 on accuracy, range, coherence and argument and specifically at least a C1 on accuracy in order to pass the course.
Number of tasks	One
Time for task (component)	75 minutes
Integration of skills	Integration of information from multiple sources (preparatory coursework)
Channel	Handwritten
Number of words expected	500
Genre	Academic essay
Imagined audience	Academics within the field of international law
Rhetorical function	Arguing, persuading and describing
Register	Formal academic
Domain	educational/academic
Assessment criteria	Range; Coherence; Accuracy; Argument.

Appendix 2: Law essay sample 2 (excerpt)

Should there be pressure for the abolition of power of veto held by the permanent members of the Security Council?

(... initial content omitted...)

One of the most prominent counter arguments concerning the abolition of the power of veto held by the permanent members of the Security Council is that the reason veto power was included in this particular UN body is still applicable today. Neither the USA or Russia would have joined the UN without this particular privilege. The UN would have become a redundant institution without the support of these two nations. It would have met the same fate as the League of Nations. This concern is still of relevance today. While the US and Russia may no longer be considered the super powers that they were after WWII they are still very influential nations. China is a nation that is becoming more important every year. It is very likely that these nations would leave the UN were they forced to give up this privilege. Their allies would follow their example. The UN would probably cease to exist. This is an unlikely scenario however for the simple fact that the power of veto cannot be abolished without express approval of the P5.

The P5 feel very secure in their positions of power. The US is a good example of the fact that these members already feel that they stand above UN authority. For example, the USA ignored all decisions that the UN made concerning relations between the US and Nicaragua. Abolishing the power of veto would change little in the grand scheme of things. It would, however, allow the UN to be restructured and perhaps enable it to become more effective in our modern day society as a result.

To conclude, there should be pressure for the abolition of the power of veto held by permanent members of the Security Council. This archaic privilege bedevils the success of the Security Council, making it undemocratic, lacking legitimacy and often sadly ineffective. It holds no place in our modern day world.