



University of HUDDERSFIELD

University of Huddersfield Repository

Sheehan, Susan, Harris, Ann and Sanderson, Pete

Identifying key criteria of written and spoken English at C1: A qualitative study of key language points at C1

Original Citation

Sheehan, Susan, Harris, Ann and Sanderson, Pete (2017) Identifying key criteria of written and spoken English at C1: A qualitative study of key language points at C1. Project Report. British Council.

This version is available at <http://eprints.hud.ac.uk/id/eprint/31024/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

**IDENTIFYING KEY CRITERIA
OF WRITTEN AND SPOKEN ENGLISH AT C1**

A qualitative study of key language points at C1

AR-A/2017/1

Susan Sheehan, Pete Sanderson, Ann Harris

ABSTRACT

This project set out to identify criterial features of written and spoken English at C1 by examining test data to establish which of the language points included at C1 in the British Council – EAQUALS Core Inventory for General English (the Core Inventory) were produced by the test-takers. The test data were also examined to see if there were recurring language points which had not been included in the Core Inventory.

The Core Inventory was created to provide a practical inventory of language points that characterise the levels of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). It includes functions, grammar points, discourse markers, vocabulary and topics. This project aimed to establish if the features included in the Core Inventory were present in test data. The test had been created to satisfy the requirements of an external validation agency. The target level of the test was C1 and so was the test-takers level of language proficiency. This had been established through IELTS scores and scores on pre-sessional course tests. A total of 36 participants took the test which included tasks focused on grammar, lexis and an extended writing task. A test of spoken English was the second part of the test. Aptis test data were also included to increase the scale of the project.

The test data were manually coded using qualitative data analysis software. The occurrences of each feature were counted. The Core Inventory contains 35 language points at C1 and six topic areas. This project found that 15 language points could be described as criterial in written and spoken English. One of the C1 topic areas featured consistently in the test data. This would seem to suggest that limited data-based evidence has been found to support an existing theoretical framework. A significant finding of the project is, perhaps, the identification of an approach of establishing criterial language points through use of the Core Inventory. This approach could be replicated with larger data sets across the full range of the CEFR levels discussed in the Core Inventory.

Authors

Dr Susan Sheehan is the Senior Lecturer TESOL at the University of Huddersfield. Her main research interests are language testing and language assessment literacy. Dr Sheehan was one of the co-authors of the British Council – EAQUALS Core Inventory for General English. She project-managed the development of a new placement test for the network of British Council teaching centres.

Dr Pete Sanderson is the Dean of the School of Education and Professional Development. His PhD in Psychology from the University of Leeds was concerned with applying social judgement and cognitive process theories to the activity of assessing expository text, based in part on his experience as a Senior Examiner for Sociology A Level. He is particularly interested in computer-assisted qualitative assessment.

Dr Ann Harris has been with the School of Education and Professional Development since 1994. She is currently the Head of Department and Principal Lecturer in the Department of Education and Community Studies. Between 1984 and 1987, she was Senior Research Officer at the Secondary Examinations Council in London, researching and advising on a variety of national policy initiatives in relation to curriculum and assessment, and evaluating an innovative program of in-service teacher training across England and Wales.

CONTENTS

1. INTRODUCTION	4
2. CONTEXT OF THE STUDY	4
2.1 Background	4
2.2 Development of the Core Inventory	4
2.3 Investigating C1	5
2.4 Socio-cognitive approach	6
2.5 Criterial features	7
3. RESEARCH QUESTIONS	7
4. METHODOLOGY	8
4.1 Participants	8
4.2 Materials	8
4.3 Data collection methodology and procedures	10
4.4 Ethical procedures and consent	10
4.5 Data analysis	10
5. FINDINGS	11
5.1 Criterial features for written English	11
5.2 Criterial features for spoken English	13
5.3 Differences between participants' language points and the Core Inventory	13
5.3.1 Beyond C1: More sophisticated language points	14
6. LIMITATIONS	14
7. CONCLUSION	15
REFERENCES	16
British Council Assessment Research Awards and Grants	18

1. INTRODUCTION

This is a report of a project conducted in an attempt to establish the criterial features of written and spoken English at C1 level. The report begins with a review of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). The British Council – EAQUALS Core Inventory for General English (hereafter the Core Inventory) is then discussed. There follows a discussion of the approaches to establishing criterial features. The methods of data collection and analysis are described. Finally, conclusions are presented with recommendations for further research.

2. CONTEXT OF THE STUDY

2.1 Background

The CEFR may be considered to have had a huge impact on the teaching and assessment of English (Hawkins and Buttery, 2010). Since 2001, the meta-language used to describe level seems to have shifted irreversibly towards that of the CEFR, such as 'A1', and away from the terms like 'beginner', 'advanced' and the plethora of terms incorporating the word 'intermediate'. One of the original authors of the CEFR ascribes its success to two main factors. The first is the positiveness of the 'Can Do' statements and the recognition of success at all levels (North, 2014). The second is the balance the CEFR achieves between the practical language learning aims and the socio-cultural aims (ibid, 2014). Other writers have attributed the success to more prosaic reasons. The three broad levels – A, B and C – felt familiar and similar to terms already used, such as 'beginner', 'intermediate' and 'advanced' (Hultjuins, 2007; Little, 2007). Furthermore, the claims of validity for the descriptors appealed to users and felt appealingly new when placed together with the reassuringly familiar levels.

Whilst the reasons for the success of the CEFR may be disputed, that it has generated debate cannot. Although the CEFR is much discussed, there are authors who suggest that the real impact of the CEFR is in fact limited. Little, who is often seen as an advocate of Council of Europe initiatives such as the European Language Portfolio, notes that: "On the whole, the CEFR has no more occasioned a revolution in curriculum development than it has prompted the radical redesign of established language tests" (ibid, 2007: 649).

2.2 Development of the Core Inventory

The stated aim of the Core Inventory was: "to make the CEFR accessible to teachers and adult learners of General English" (North, Ortega and Sheehan, 2010: 6). The authors of the Core Inventory stated that it was directed towards teachers rather than testers. North amplified this point when he stated: "One could say that the Core Inventory documents what is taught at these levels and that English Profile documents what is learned" (op. cit., 2014: 91).

When writing the Core Inventory, the C levels were the most problematic. Indeed, the lack of consensus around what should be taught at C levels led to C2 being excluded from the project. Limited consensus was found around C1. Research undertaken by Green (2012) shows that 80% of functions were present at B2. So, it would seem that new language is not a feature of language courses at C1 level. Thus, this would seem to suggest that the features which distinguish C1 from other levels are not newly taught language functions; rather that quality and sophistication is what marks out C1 from other levels. This would seem to make the challenge of identifying criterial features yet more challenging, as sophisticated and subtle features are being examined.

A validation of the Core Inventory was conducted under the aegis of the Aptis Research program. Jones (2015) found that nearly 50% of the language functions were correctly placed. Where functions were misplaced, Jones (2015) argues that they were included at a level above that which they should have been placed. That is to say, a language point at B1 would have been better placed at A2. North (2014: 91) compared the Core Inventory with English Profile Version 1.0. He found: “a high level of agreement in the classification of content by the two projects”.

Thus, it would seem that there is some limited work to support the classifications of the Core Inventory. However, there are two caveats which should be acknowledged. As North (2014) notes, similarities between the Core Inventory and the English Profile Project (EPP) may be expected as both projects have a European focus and thus similar types of learners are involved in both. Secondly, Jones (2015) chose to represent his findings that fewer than 50% of language points were correctly placed and thus took a more negative view of the Core Inventory. The current project is not a validation of the Core Inventory. Rather, it seeks to use it to establish criterial features of C1.

2.3 Investigating C1

C1 is a challenging CEFR level to investigate as there is a lack of consensus around what makes a person a C1 level language user (Weir, 2005b). At the same time, C1 is a level which is requested by various authorities. For example, for a teacher certification agency, C1 is the required level for someone wishing to teach English as a foreign language. The creation of a test to satisfy this requirement was the impetus for the current project. This topic will be returned to in the materials section of the report. Several education authorities in Europe have declared C1 as the minimum level required by academics whose own language is not English. One of the challenges of identifying the criterial features of C1 is the increasing level of sophistication of language use. This sophistication can manifest itself in the use of more elaborate cohesive devices and the use of fewer explicit connectors (North, 2014). Thus, the researcher is left with the challenge of trying to identify an absence. There is some evidence (Jones, 2015) that learners at lower levels are coached to include a number of cohesive devices. This may be particularly evident in classes that prepare students to take written examinations. Thus, learners no longer need to rely on these explicit connectors and the absence of them could be considered a criterial feature at C1.

C1, it has been claimed (Weir, 2005b; Alderson 2007), is an underspecified level. The writers of the CEFR have acknowledged this and, as part of the EPP project, further descriptors have been added at C level. Green (2012) provides a detailed account of an approach adopted to developing new descriptors at C level. He also provides a thorough explanation of the difficulties of establishing criteriality at C level. In part, he attributes this to the nature of the tasks C level learners need to perform and, in part, to the possibility that socio-linguistic and pragmatic competences may be more criterial than linguistic competences.

In some ways, this project shares the same aims as the EPP. It is seeking to identify criterial features, albeit of one CEFR level rather than all six levels. This project is radically different in the methodology adopted in the pursuit of those aims. The EPP also seeks to investigate the transfer effects of a person's own language. This goes far beyond the scope of the current project. The EPP, and indeed the use of corpus analysis to define levels of proficiency, have come in for criticism. North (op. cit.: 23) writes that corpus analyses: “still offer only a glorified snapshot of the aggregated proficiency of groups of learners at one point in time”. Whilst both the CEFR and the Core Inventory are, it could be argued, based on teacher judgement of levels of proficiency, there is a need to bring the learner into the process of establishing level. Undoubtedly, there are limits which North has identified but the language users should also be included.

The EPP has also been criticised on the grounds that the levels the EPP researchers are working to may not reflect the CEFR levels (O'Sullivan, 2011). The examinations from which the EPP data are taken were created before the CEFR. This may account for the discrepancy between the CEFR levels and the interpretation of the CEFR levels by test developers. Alderson (2007) notes that the EPP does not include any independent estimate of the learners' proficiency levels. The EPP has been criticised for the examination-based nature of its corpus (ibid, 2007). The language used by a learner in a test may not actually represent the language they would choose to use outside the examination hall. As has been noted, teachers may coach students to use particular connectors in a test, as use of these creates a positive impression on the examiner (Jones, 2015). Developing a methodology that could investigate this phenomenon is beyond the scope of the current project.

Having stated above that C1 is an under-defined level, attention will now be turned to research conducted to date. North (op. cit.: 50) states that: "Level C1 is characterised by access to a broad range of language that results in fluent, spontaneous communication...They (C1 speakers) are also very accurate, with a solid grasp of the main grammatical structures of the language". Research from EPP would seem to suggest that learners are more accurate at C1 than B2. (North, 2014; Green, 2012). Furthermore, vocabulary knowledge extends to cover more senses of already known headwords (Capel, 2010; North, 2014). Thus, it should be expected that C1 level learners are significantly more accurate than B2 level learners and they are also able to use more members of a word family. Such an understanding reflects the philosophy underpinning the development of the Word Family Framework (WFF). The WFF is a word list which has ascribed a CEFR level to each different part of the word family. So, for example, the verb 'to notice' is placed at A1 while 'noticeable' is placed at C1. West (2012) provides an account of the development of the WFF.

2.4 Socio-cognitive approach

The Core Inventory and the Aptis tests share an amount of common history. The Core Inventory team was influenced by studies of grammatical progression undertaken by the International Language Assessment team (O'Sullivan, 2015). The Core Inventory: "represents a significant attempt to add detail to the CEFR level descriptors" (ibid, 2015b: 14). The Aptis development team drew on the Core Inventory for the descriptions of lexical, grammatical, socio-linguistic, pragmatic and strategic competence (ibid, 2015b). Thus, the Core Inventory has an acknowledged role in the development of Aptis. The current project could, therefore, provide evidence to support this choice by the Aptis development team.

The Aptis tests were developed following the socio-cognitive validation model. This was first developed by Weir (2005) and subsequently further developed by O'Sullivan and Weir (2011). A description of the theoretical basis of the Aptis test system can be found in O'Sullivan (2015). The model is based on three elements: the test-taker, the test system and the scoring system (ibid: 2015). This project is relevant to the test system element as it is looking at the linguistic demands of the test task. Test developers need to develop an understanding of the expected output of test candidates.

This project is also relevant to the scoring system element as it seeks to establish the criterial features of C1. Scoring systems need to relate to the CEFR and, to do so, more evidence needs to be obtained as to what these levels actually mean.

2.5 Criterial features

The final section of this introduction is a discussion of criterial features. Since the publication of the CEFR, projects have been undertaken to specify language which can be determined as characteristic or criterial of any given CEFR level. This, in part, can be ascribed to the language neutral aspect of the CEFR. Although much of the initial work on the scales was conducted with teachers of English, the CEFR can be applied to any European language. Various institutions concerned with language teaching and learning have undertaken projects to identify which language points correspond to which CEFR level. An example of one is Profile Deutsch which provided language lists for learners of German. Green (2012) notes that the authors of the Profile Deutsch struggled to find consensus at C levels and did not initially publish vocabulary lists for these levels. This would suggest that identifying language features which are typical of C1 or beyond, which could be described as criterial, is tremendously challenging as learners at higher levels specialise in particular language uses and focus on more specialised vocabulary areas. For a full discussion of the controversies and challenges of defining higher levels of proficiency, see Harsch (2014).

The definition of a criterial feature adopted for this project is: “Criterial features describe changes from one level to the next and are important for both practitioners and theoreticians to know about.” (Green, 2012: xvii). In the case of the current project, this means features which separate C1 from B2 and C2.

This project is about language points and level and not the language learner. Learners have profiles rather than levels. This project does not seek to identify what a C1 level learner can do. It has also been argued that the criterial features at C1 may not be linguistic competences (ibid, 2012). The subtle and nuanced use of language at C levels may mean that the criterial features may be: “socio-linguistic and pragmatic” (ibid: 112). While this may present a challenge to the researcher, it should still be possible to identify criterial differences.

3. RESEARCH QUESTIONS

The challenges of identifying criterial features of written and spoken English at C1 have been outlined above. The development of the Core Inventory has also been described. The current research project uses the Core Inventory as a tool through which the criterial features of C1 can be identified.

The specific research questions that were addressed are the following:

1. Which of the features identified as core in the Core Inventory feature in the candidate's responses to the written test?
2. Which of the features identified as core in the Core Inventory feature in the candidate's responses to the oral test?
3. Are there other features which recur in the candidates' responses but which are not included in the Core Inventory?
4. What are possible explanations for the differences?

4. METHODOLOGY

4.1 Participants

There were two groups of participants. The first group of participants was all MA TESOL students. They had all been judged to have C1 level English. This was established through two routes. Firstly, IELTS scores of 6.5 – 7 were taken to mean the candidate had C1 level. The use of an IELTS or other score to indicate a CEFR level could be considered to be controversial. An IELTS score of 6.5 has been found to be equivalent to C1 (Lim, Geranpayeh, Khalifa and Buckendahl, 2012). The authors acknowledge that such equivalences are controversial and other studies have reported an IELTS score of 7 to be equivalent to C1. Secondly, scores obtained in pre-sessional English course tests were also used. The participants were all aspiring teachers of English. Participants came from China, Libya, Vietnam, Kurdistan and Pakistan – the first three countries in the list providing the most participants. The participants were all aged between 25 and 35 and all aspired to careers as English teachers upon completion of their studies. The majority was aiming for a career in tertiary education.

There were 24 students in the MA TESOL cohort. One of them was from the United States and not invited to participate in the project as English was considered to be his own language. Another student was not asked to participate as his pre-sessional test scores were low and his level of English was judged to be below C1 by course tutors. Of the 22 students invited to participate, 15 chose to do so. This was a disappointing response rate. There are several possible reasons for this lack of interest in the project. Firstly, a decision was taken not to offer the externally validated teaching certificate to this group of students. This meant the students could not see any gain for themselves by participating in the project. Secondly, the taught MA program is short and the students may not have been immersed long enough in academic culture to appreciate the importance of participating in research projects. Students were given repeated opportunities to participate in the project. Those who attended were offered cakes. Course tutors encouraged students to participate in data collection sessions.



As the number of students was low, it was decided to repeat the data collection process with the following year's cohort of students. On this occasion, 21 out of a possible 23 students participated in the project. The two students who did not participate were away from the university for health reasons. It is not clear why the second cohort of students were keener to participate in the project.

In total, 36 MA TESOL students participated in the project.

The second group of participants was made up of Aptis test-takers. The small size of the first cohort highlighted the need for a larger data set. To increase the scale of the project, Aptis test data were provided by the test developers. The test data were supplied in two tranches. The first tranche contained test data from tests administered around the world. The second tranche contained test data from people who were aspiring Aptis examiners as, in order to be accepted as an examiner, it is necessary for a person to demonstrate a certain level of English.

4.2 Materials

Participants were asked to complete a written test and an oral interview. The test used in this project was developed to satisfy the requirements of an external validation agency. The test was designed to identify suitable candidates for a course to train people to become teachers of English as a foreign language (EFL). Candidates are expected to demonstrate a C1 level of language proficiency. The written test was composed of three parts. The first part was a multiple-choice grammar test with 16 items. The items were based on the exponents listed in the Core Inventory. The exponents included in the grammar test are listed below. The numbers follow the numbering sequence used in the Core Inventory.

- 93 If only
- 92 Mixed conditionals 
- 30 Expressing shades of opinion and certainty
- 34 Developing an argument systematically
- 25 Expressing certainty, probability doubt
- 27 Synthesizing, evaluating, glossing information
- 34 Developing an argument systematically
- 92 Mixed conditionals 
- 76 Future revision
- 188 Idiomatic expressions
- 181 Collocations
- 183 Approximating
- 159 Inversion negative adverbials
- 38 Responding to counterarguments
- 32 Expressing reaction, e.g. indifference
- 21 Expressing feelings and attitudes precisely

Multiple-choice items were developed based on those language points.

The second part was a test of lexis. A head word was supplied to which the participants had to supply the other members of the word family. The source for the lexical items was the Word Family Framework. The words included: administer, age, continue, extend, mobile, notice, part and simplicity.

The final part of the written test was to provide a commentary on two pieces of student writing. The participants had to analyse the strengths and weaknesses of the two samples of student writing and suggest possible remedies for the weak points. The external validation agency demanded that candidates for its course should be able to demonstrate language awareness in order to be able to teach English. The interview questions were also based on the external agency's requirements and focused on past language learning experience and reasons for wanting to be an English teacher. For the second cohort of students, the final section of the test was changed to an essay. This change was made as the first cohort responded to the task by using a very similar and narrow range of language. The language of advice, which is placed at A2, in the Core Inventory strongly featured in all the written work. It would seem that the nature of the task had strongly influenced the language used. So, it is not the case that advice could be considered a criterial feature of C1. Rather, it would seem to suggest that test developers should not under-estimate the effect of the task on the language used by test-takers.

For reasons of confidentiality and test security, detailed information about the Aptis tasks will not be provided. Further information about the Aptis tests and the task types can be obtained from <https://www.britishcouncil.org/exam/aptis>

4.3 Data collection methodology and procedures

The materials and the data collection procedures were piloted during the previous academic year. As a result of piloting, the rubrics were changed considerably and further examples were included for the lexis test. The pilot participants had reported that they found the task in this section of the test to be confusing.

The participants were invited via email and in-person to attend data collection sessions. The sessions were held at times when the participants were not time-tabled to attend lectures. The data collection sessions were held in classrooms in the institution. The purpose of the data collection session was explained to all the participants. Following the explanation, the participants were invited to sign the consent form. Details of the consent form will be given below in the ethics section.

The participants completed the written test first. They had one hour to complete the test. Most participants completed the test before the hour was up. The test was completed under examination conditions. The participants sat more than 1.5 metres away from each other. Mobile phones and other similar devices had to be turned off. Participants could not use dictionaries. The participants were interviewed individually. The interviews were digitally recorded.

4.4 Ethical procedures and consent

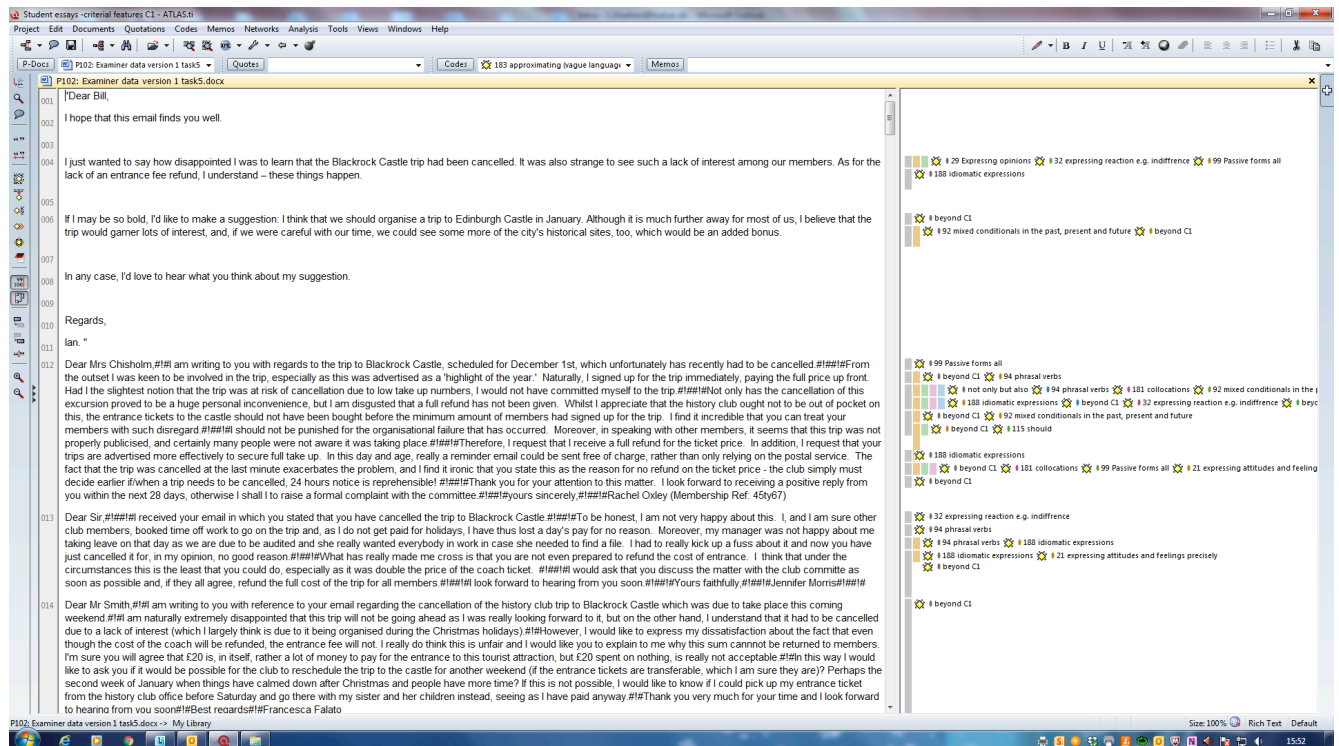
The planning and execution of the project followed the guidelines of the University of Huddersfield and were reviewed by the appropriate ethics committees. Participation was entirely voluntary and no type of pressure was exerted on students to participate in the project. The participants signed a consent form which covered a range of areas which are listed next. The participants had the right to withdraw from the project at any time. They could review project material at any time and will have access to the final report. Participants were not offered any kind of incentive or inducement to take part in the project. To ensure high ethical standards were being maintained, the guidelines of the British Educational Research Association (BERA, 2011) were followed. None of the 36 participants chose to withdraw their consent after having agreed to participate in the data collection process.

4.5 Data analysis

There were two main approaches to data analysis. The tests of grammar and lexis were marked using a key of correct answers. The number of participants meant that statistical analysis was not useful. The extended writing task and interview data were analysed using ATLAS.ti software. The codes used were based on the Core Inventory. The codes were then applied to data. So, for example, 'Critiquing and reviewing, 33' was used to code all examples of participants using such language. In cases where the Core Inventory did not have the appropriate language point, a code was created by the research team.

Once the data had been coded, it was possible to identify which aspects could be defined as criterial. Absolute totals for occurrences were generated. For a language point to be considered criterial, it needed to have a minimum of 15 occurrences. This follows the practice established by the writers of the Core Inventory. The same approach to data analysis was used on both the Aptis test data and the data collected with the MA TESOL students. Incorrect or complete samples of language were excluded from the project. This follows an approach adopted by Jones (2015) and, to some extent, shares some of the same difficulties in using a similar coding process. It should be stated that the methodology for the current project was developed independently.

Below is a screenshot which shows how the codes were applied to the data.



5. FINDINGS

Given the size of the data set, the findings presented here must be taken with a note of caution, and doubts about the generalisability of these findings can be raised with some justification. Nevertheless, some interesting insights into criterial features can be presented. A group of language points was found to be criterial for written English. A second group was found to be criterial for spoken English. The findings are organised by research question.

5.1 Criterial features for written English

This section will discuss findings related to the three written components of the test: the grammar test, the lexis test and the extended writing task, and relates to the first research question. In terms of the grammar test, most of participants were able to answer most of the questions correctly. The highest score was 15 and the lowest score was 9.

Two test items were answered incorrectly by most of the participants. The language point and the exponent are listed below. The numbers are taken from the Core Inventory. These were:

- 30 Expressing certainty, probability, doubt – *I rather doubt that he'll come.*
- 181 Collocations – *The suspense is palpable.*

The test item on collocations was a feature which most participants were unable to answer correctly. So, knowledge of collocations can be considered to be criterial but, within the category, there may be collocations which are particularly difficult for language learners. The test may have included an obscure one. The source of the test items was the list of exponents for language content included in Appendix E of the Core Inventory. Jones (2015) has suggested that language used in these exponents is sometimes obscure and a little old-fashioned. The exponents, therefore, may be in need of revision.

As most of the students answered the two items listed above incorrectly, this may suggest that these features could be placed at a level which is above C1. This might be C2 or could even be beyond C2.

The lexis test revealed some interesting patterns but, given the small scale of the data, it is not possible to go beyond these general patterns. C1 words were correctly identified 125 times. This compares to 82 correctly identified words at B2 level and 31 words at C2. The word *mobile* had the most correctly identified C1 word family members and *age* had the fewest. The highest average scores were found at C1 level with the exception of A1. This finding seems to support Capel's (2012) assertion that knowledge of word families is a defining characteristic of learners at C levels.

The language points identified in the Core Inventory as core that were found with sufficient consistency for them to be considered as criterial are listed below. Across all the different data sets, 1,623 quotations were identified. Each language point includes the number of times it was coded in brackets. This indicates the level of frequency at which the language point featured across the data sets. Each point is illustrated with an example taken from one of the extended writing tasks.

- 21 Expressing attitudes and feelings precisely (108) – *it surely could have been worded better.*
- 29 Expressing opinions (270) – *Intrusion into individual may be the top concern of most citizens.*
- 32 Expressing reaction, e.g. indifference (73) – *I'm not that bothered about the trip being called off.*
- 33 Critiquing and reviewing (48) – *The writing is organised by using some cohesive devices though it still contains mistakes.*
- 34 Developing an argument systematically (47) – *This sure is a beneficial thing to a city since it allows monitoring of public safety. Furthermore, when it comes to legal evidence, those tapes could provide evidence in a criminal case.*
- 54 Markers to structure and signpost formal speech and writing (89) – *The appeal is two-fold.*
- 94 Phrasal verbs (88) – *to put a lot of effort into*
- 99 Passive forms, all (78)– *The route map should be improved.*
- 181 Collocations (41) – *One gets the gist of the main points.*
- 203 Technical and legal language (99) – *Ability to use different tenses, e.g. Present Tense, Future Simple, Past Simple.*

The possible problem with the topic area of technical language being criterial is that it will be different for different groups of students and indeed for individuals. The technical and legal language used related to teaching and learning English. This was true of both Aptis test data sets and the test data gathered from the MA TESOL students. This would suggest that teachers of English, both those in practice and those in training are also frequent takers of tests of English. A C1 level test aimed at engineers or financiers would necessarily feature different technical language. There appears to be little communality between the different language points included in the list. They may simply be a reflection of the tasks given to the participants. A different set of tasks could, potentially, generate a different list of criterial features.

5.2 Criterial features for spoken English

The next section reports findings related to the second research question which focused on features which could be considered criterial in only spoken English. These were:

- 54 Markers to structure and signpost informal speech (132) – *The reason I want to become a teacher is because I love children. In addition, it is because I want to teach them a more sophisticated language.*
- 66 Narrative tenses for experience (99) – *The teacher would make us complete many exercises.*
- 181 Collocations (109) – *They had good intentions and they helped us to overcome problems.*
- 21 Expressing feelings and attitudes precisely (108) – *What really annoys me is when a colleague refuses to help me out.*
- 203 Technical and legal language – *Peer observation is a great way to improve myself and for me to help to develop my colleagues.*

It is disappointing to note that only five language points were found with sufficient frequency for them to be considered criterial in spoken English. Language users at this level of proficiency tend to focus on English for specific purposes, such as legal English where written ability is emphasised. There are three language points which were found to be criterial for both written and spoken language. These were:

- 21 Expressing feelings and attitudes precisely
- 181 Collocations
- 203 Technical and legal language.

As discussed above, both data sets contained language relating to the teaching and learning of English. All the examples of technical and legal language relate to English language teaching. This may be attributed to a tendency to test both aspiring and practicing teachers of English in order to gauge if their levels of proficiency are appropriate for a teacher of English. Indeed, such a test was the original inspiration for this project.

5.3 Differences between participants' language points and the Core Inventory

The third research question asked if there were language points which recurred in the data but were not included in the Core Inventory. Much of the participant output in both written and spoken forms was included in the Core Inventory. The responses included language points from across all CEFR levels. A common feature found in the extended writing tasks was what could be termed as 'sophisticated advice'. Advice is placed at A2 in the Core Inventory. The exponents given for it are: "You should ask the teacher, you could try the Internet". Examples of this language point taken from the commentaries include: *the road map should be improved* and *I'd recommend this student to*. The frequency of this feature may be in response to the nature of the given task. As Green (2010: 9) notes: "The tasks that learners are asked to perform have a substantial impact on the nature and quality of the language that they will be able to produce". It would seem that, in this case, the nature of the extended writing task influenced the language produced.

The final research question sought to identify reasons for the differences between the Core Inventory and the language used by the participants. The methodology of the Core Inventory may have failed to recognise different levels of sophistication in the use of a particular feature. Having been consigned to A2, 'advice' was in effect forgotten as other language points came under discussion. This may reflect the close relationship to structure of course books and the Core Inventory – 'should' is taught at A2 so 'advice' is a feature of A2. The C1 part of the Core Inventory, it could be argued, focuses too much on argument and academic English, and not enough on broader aspects of the language.

5.3.1 Beyond C1: More sophisticated language points

The inclusion of Aptis test data led to the development of a new code which was not taken from the Core Inventory. The code was named 'beyond C1'. This code was used on 58 occasions. It was applied to language points which were more sophisticated or complex than those used in other parts of the data.

This code was generally applied to the test data drawn from those who had taken the Aptis test in order to become an examiner for Aptis. One person wrote: *While I understand that the trip was cancelled due to lack of interest, I do not feel it is fair to penalise those members who did sign up by refusing to refund the cost of their entrance ticket.* This is an example of the more sophisticated use of language labelled as beyond C1.

Other language points coded this way were more colloquial. For example, another participant writing about their disappointment at a decision wrote simply: *That sucks.* Such a term was not used in other data sets. This would seem to suggest the difference between C1 and C2 or even what could be termed as 'beyond C2' is the range of language at the disposal of the writer or speaker. The language can be sophisticated or the range of vocabulary used is wider, and this allows greater precision in expression.

6. LIMITATIONS

Two major limitations have been identified. Firstly, the task types may have influenced the language used by the participants. Secondly, all the coding was undertaken by one researcher. Thus, the number of idiosyncratic codes may be higher than desirable.

7. CONCLUSION

Using test data it has been possible to investigate patterns of occurrence with respect to the C1 level. It has been found that 15 language points listed in the Core Inventory could be described as criterial. This would seem to suggest that limited data-based evidence has been found to support an existing theoretical framework. Perhaps the most significant finding of this project is the creation of an approach to establishing criterial features through the use of the Core Inventory. This approach could be replicated with larger data sets across the full range of the CEFR levels discussed in the Core Inventory.

As discussed above, C1 is considered to be an underspecified level. The Core Inventory authors were unable to find consensus around the language points to be included at this level. The current project found that nearly one quarter of the language points included at C1 occurred with sufficient frequency to suggest that they could be considered criterial for the level. This may indicate that C1 level users share more in common than has previously been thought to be the case.

Contrary to what had been expected, the test-takers used markers to structure their written and spoken English. Thus, it seems that these devices remain a feature of the language used by sophisticated users. It was not, therefore, necessary to try to detect an absence. Or if it were there, it was too sophisticated for this researcher.

As stated above, the authors of the Core Inventory aimed at providing a resource for teachers and not testers. There is evidence from this project and perhaps that of the project conducted by Jones (2015) that the Core Inventory describes when learners have mastered a language point, rather than when learners should be taught that CEFR level. It may be argued, therefore, that using the Core Inventory as one of the sources of the Aptis test was an appropriate use of this resource.

The outcomes this project aimed to achieve included:

- an enhanced understanding of the key criteria of performance in speaking and writing at C1
- a publishable list of the criteria
- development of an approach to identifying key features which can be replicated by other institutions or with other CEFR levels.

The first outcome has been achieved to a limited extent as some language points have been indicated to be criterial. The second aim has been partially achieved as the 15 language points identified could be published as a list. At the outset of the project, it was hoped that the list would be longer. The third aim is the one which is closest to being realised. The approach outlined in this research project could be extended to cover other CEFR levels and could be replicated by other institutions.

REFERENCES

- Alderson, J. (2007). The CEFR and the need for more research, *The Modern Language Journal*, vol 91, pp 659–663.
- BERA. (2011). *Ethical Guidelines for Educational Research* available from: <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1>
- Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 Vocabulary, *English Profile Journal*, vol 3, pp 1–14.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Gad, L., Geranpayeh, A., Khalifa, H. and Buckendahl, C. (date?). Standard Setting to an International Reference Framework: Implications for Theory and Practice, *International Journal of Testing*, vol 13, pp 32–49.
- Green, A. (2010). Requirements for Reference Level Descriptions for English, *English Profile Journal*, vol 1, pp 1–19.
- Green, A. (2012). *Language Functions Revisited – Theoretical and empirical bases for language construct definition across the ability range*. Cambridge: Cambridge University Press.
- Harsch, C. (2014). General Language Proficiency Revisited: Current and Future Issues, *Language Assessment Quarterly*, vol 11, pp 152–169.
- Hawkins, J. and Buttery, P. (date?). Serial Features in Learner Corpora: Theory and Illustrations, *English Profile Journal*, vol 1 pp 1–23.
- Hulstijn, J. (2007). The Shaky Ground Beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency, *The Modern Language Journal*, vol 91, pp 663–667.
- Jones, G. (2015). *A Validation Study of the British Council – EAQUALS Core Inventory for General English*, available from: <https://www.britishcouncil.org/exam/aptis/research/publications/validating-core-inventory-general-english>
- Little, D. (2007). The CEFR: Perspectives on the Making of Supranational Language Education Policy, *The Modern Language Journal*, vol 91, pp 645–655.
- North, B., Ortega, A. and Sheehan, S. (2010). *British Council – EAQUALS Core Inventory for General English*. London: British Council.
- North, B. (2014). *The CEFR in Practice*. Cambridge: Cambridge University Press.
- O’Sullivan, B. (2011). Language Testing. In J. Simpson (Ed.), *The Routledge Handbook of Applied Linguistics*, pp 259–273. London: Taylor and Francis Group.
- O’Sullivan, B. and Weir, C. (2011). Test Development and Validation. In B. O’Sullivan (Ed.), *Language Testing: Theories and Practices*, pp 13–33. Basingstoke: Palgrave Macmillan.
- O’Sullivan, B. (2015a). *Aptis Test Development Approach*. London: British Council.
- O’Sullivan, B. (2015b). *Linking the Aptis Reporting Scales to the CEFR*. London: British Council.
- Weir, C. (2005a). *Language Testing and Validation An Evidence-Based Approach*, Basingstoke: Palgrave Macmillan.
- Weir, C. (2005b). Limitations of the Common European Framework for developing comparable examinations and tests, *Language Testing*, vol 22, pp 282–300.
- West, R. (2012). *Word Family Framework* London: British Council.

British Council Assessment Research Awards and Grants

If you're involved or work in research into assessment, then the British Council Assessment Research Awards and Grants might interest you.

These awards recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

IDENTIFYING KEY CRITERIA OF WRITTEN AND SPOKEN ENGLISH AT C1

A qualitative study of key language points at C1

AR-A/2017/1

**Susan Sheehan
Pete Sanderson
Ann Harris**

**ARAGs RESEARCH REPORTS
ONLINE**

ISSN 2057-5203

© **British Council 2017**

The British Council is the United Kingdom's international organisation for cultural relations and educational opportunities.