



EXAMPLES OF SPEAKING PERFORMANCE AT CEFR LEVELS A2 TO C2

(Taken from Cambridge ESOL's Main Suite exams)

Project overview

April, 2009
University of Cambridge ESOL Examinations
Research and Validation Group



Contents

Contents	2
Foreword	3
Introduction.....	4
Background to the project.....	4
Brief description of Cambridge ESOL's Main Suite speaking tests.....	4
Procedure and Data collection.....	5
Instruments	6
Data Analysis.....	7
References	10
Appendix A: CEFR Assessment scales (Global and analytic).....	13
Appendix B: Example of a Rating form	15



Foreword

This documentation accompanies the selected examples of speaking tests at CEF levels A2 to C2. The selected speaking test performances were originally recorded for examiner training purposes, and are here collated for the use of the Council of Europe's Language Testing Division, Strasbourg. The sample material is not collated to exemplify the exams on this occasion, but to provide speaking exemplars of CEF levels. These speaking test selections are an additional resource (to the existing one on the Council's website) that Cambridge ESOL would like to share with other language testing and teaching professionals.

The persons shown on these recordings have given their consent to the use of these recordings for research and training purposes only. Permission is given for the use of this material for examiner and teacher training in non-commercial contexts.

No part of the selected recordings may be reproduced, stored, transmitted or sold without prior written permission. Written permission must also be sought for the use of this material in fee-paying training programmes.

Further information on the content and exams exemplified in these sample tests is available in the Exam Handbooks, reports, and past papers, which can be obtained via the Cambridge ESOL website, <http://www.cambridgeesol.org/>

or by contacting:

University of Cambridge
ESOL Examinations
1 Hills Road
Cambridge
CB1 2EU
United Kingdom

Tel. +44 (0) 1223 553355
Fax. +44 (0) 1223 460278
e-mail: ESOL.helpdesk@ucles.org.uk



Introduction

Background to the project

In line with the launch of an updated version of First Certificate of English (FCE) and Certificate in Advanced English (CAE) examinations in December 2008, Cambridge ESOL initiated a project with the aim to provide typical speaking test performances at levels A2 to C2 of the CEF which could be used as calibrated samples in CEF standardisation training and ultimately in aiding a common understanding of the CEF levels.

The samples used were taken from Cambridge ESOL General English Examinations, henceforward referred to as Main Suite. Main Suite is five-level suite of examinations ranging from A2 to C2, namely, Key English Test (KET), Preliminary English Test (PET), FCE, CAE, and Certificate of Proficiency in English (CPE).

Background to Cambridge ESOL's Main Suite speaking tests

The Cambridge approach to speaking is grounded in communicative competence models, including Bachman's (1990) Communicative Language Ability (built on the work of Canale & Swain, 1980 and Canale, 1983) and the work of other researchers working in the field of task-based learning and assessment (Skehan, 2001; Weir, 1990, 2005). As Taylor (2003) notes in her discussion of the Cambridge approach to speaking assessment, Cambridge ESOL tests have always reflected a view of speaking ability which involves multiple competencies (e.g., lexico-grammatical knowledge, phonological control, pragmatic awareness), to which has been added a more cognitive component which sees speaking ability as involving both a knowledge and a processing factor. The knowledge factor relates to a wide repertoire of lexis and grammar which allow flexible, appropriate, precise construction of utterances in real time. The processing factor involves a set of procedures for pronunciation, lexico-grammar and established phrasal 'chunks' of language which enable the candidate to conceive, formulate and articulate relevant responses with on-line planning reduced to acceptable amounts and timings (Levelt, 1989). In addition, spoken language production is seen as situated social practice which involves reciprocal interaction with others, as being purposeful and goal-oriented within a specific context.

The features of the Cambridge ESOL speaking exams reflect the underlying construct of speaking. One of the main features is the use of direct tests of speaking, which aims to ensure that speech elicited by the test engages the same processes as speaking in the world beyond the test and reflects a view that speaking has not just a cognitive, but a socio-cognitive dimension. Pairing of candidates where possible is a further feature of Cambridge ESOL tests which allows for a more varied sample of interaction, i.e. candidate-candidate as well as candidate-examiner. Similarly, the use of a multi-part test format allows for different patterns of spoken interaction, i.e. question and answer, uninterrupted long turn, discussion. The inclusion of a variety of task and response types is supported by numerous researchers who have made the case that multiple-task tests allow for a wider range of language to be elicited and so provide more evidence of the underlying abilities tested, i.e. the construct, and contribute to the exam's fairness (Bygate, 1988; Chalhoub-Deville, 2001; Fulcher, 1996; Shohamy 2000; Skehan, 2001).

A further feature of the Cambridge ESOL speaking tests is the authenticity of test content and tasks, as well as authenticity of the candidate's interaction with that content (Bachman, 1990). A concern for authenticity in the Cambridge ESOL exams can be seen in the fact that particular attention is given



during the design stage to using tasks which reflect real-world usage, i.e. the target language-use domain, and are relevant to the contexts and purposes for use of the candidates (Bachman, 1990; Saville, 2003; Spolsky, 1995).

As well as informing speaking test format and task design, the underlying construct of spoken language ability also shapes the choice and definition of assessment criteria, which cover Grammar/Vocabulary, Discourse Management, Pronunciation, and Interactive Communication. The use of both analytical and global criteria enables a focus on overall discourse performance as well as on specific features such as lexical range, grammatical accuracy and phonological control.

Task specifications at all levels of the Speaking papers (e.g. in terms of purpose, audience, length, known assessment criteria, etc) are intended to reflect increasing demands on the candidate in terms of Levelt's (1989) four stages of speech processing. Tasks at the higher levels are more abstract and speculative than at lower levels and are intended to place greater demands on the candidates' cognitive resources. Scoring criteria are targeted at greater flexibility in the language used at the level of the utterance, in interaction with other candidates or the examiner and in longer stretches of speech.

Procedure and Data collection

Sample description

The project involved a marking exercise with 28 test takers distributed in 14 pairs and eight raters. The test-taker samples came from a pool of existing Cambridge ESOL speaking test performances which are high-quality test recordings used in rater training. In selecting the test takers to be used in the marking exercise, a variety of nationalities was targeted, not just European, and both male and female test takers were included.

The project consisted of two phases. Twenty test takers distributed in 10 pairs were used during phase 1. They were taken from an available pool of 25 speaking tests which are used for rater training purposes and are marked against a global and analytic Main Suite oral assessment scale. The selection of the 10 pairs was based on the Main Suite marks awarded, and typical performances were operationalised as performances at the 3/3.5 band range of the Main Suite scale, while borderline performances were located at the 1.5/2 range of the scale. Based on the typical/borderline criteria adopted, one typical pair and one borderline pair were selected per level, to further confirm raters' ability to distinguish between borderline and typical candidates.

Phase two of the project focused on performances at the C levels only where in phase 1 raters had a low level of agreement and the sample comprised four additional pairs of test takers (two at CAE/C1 and two at CPE/C2). During this phase of the project a typical performance at CAE/C1 or CPE/C2 was operationalised as being at bands 4/4.5 of the Main Suite scale and a borderline performance was located at bands 2.5/3. (See Findings for a more detailed discussion of the two project phases.)

Entire speaking test performances, rather than test parts, were used in the sample in order to allow for longer stretches of candidate output to be used by the raters when rating. The use of whole tests also added a time-dimension to the project, as full tests are more time consuming to watch and may introduce elements of fatigue. The raters had to spend a minimum of 8 minutes and a maximum of 19 minutes per single viewing. Such practical considerations limited the number of performances at each phase of the project to two per level.



Raters' Profile

The eight raters participating in the project were chosen because of their extensive experience as raters for Main Suite speaking tests, as well as other Cambridge ESOL exams. They had also participated in previous Cambridge ESOL marking trials and had been shown to be within the norm for harshness/leniency and consistency. The raters had many years of experience as speaking examiners ranging from 11 to over 25 years, and were based in several parts of Europe. In addition, they had experience spanning different exams, with different task types and assessment scales, which had enriched their experience as raters. In terms of familiarity with the CEFR, seven of the raters indicated that they were familiar/very familiar with the CEFR, while one rater reported a low-level of familiarity with the CEFR. As will be seen in the "Instruments" section, a CEFR familiarisation activity given prior to the marking exercise was used to ensure that all raters had an adequate level of familiarity with the CEFR.

Design

A fully-crossed design was employed where all the raters marked all the test takers on all the assessment criteria. The decision to select 8 raters was based on advice given by Cizek & Bunch (2007: 242), and by the Council of Europe (2004). In addition, the number of observations recorded (8 raters giving 6 marks to 28 candidates) was in agreement with the sample size required by FACETS and allowed for measurements to be produced with a relatively small standard error of measurement.

Instruments

The raters were sent the following materials:

- Two scales from the CEF Manual: a global scale (COE, 2001: 24, referred to as Table 5.4 in appendix A), and an analytic scale (COE, 2001: 28-29, referred to as Table 5.5 in Appendix A) comprising five criteria: Range, Accuracy, Fluency, Interaction, Coherence (see Appendix A);
- A DVD with 10 Main Suite speaking tests (20 candidates total);
- A CEF familiarisation task (see Appendix B);
- A rating form for recording the level awarded to each candidate and related comments (see Appendix B);
- A feedback questionnaire.

The CEF scales used were slightly adapted from the original, and levels A1+ and C1+ were added. It was felt that the raters needed to have a full-range of the scale available, with the possibility to award borderline levels at all available levels, including A1+ and C1+, which are not in the original CEF scales. Taking into account the borderline levels, the scale used in the project had 12 steps.



The raters were sent detailed instructions about the marking, which are given below:

Please go through the following **steps**:

1. Read through the CEF scales to get a feel for the detail of description for the global and analytic categories (Range, Accuracy, Fluency, Interaction, Coherence).
2. Highlight key elements of the descriptors that indicate differences in performance at each level.
3. Do a self-assessment exercise in order to become more familiar with the scales prior to rating. Think of a foreign language you speak. If you do not speak a foreign language, think of a specific language learner who you have taught in the past or a language learner you are familiar with. Assess that learner using the global assessment scales first. Then give an assessment for each of the categories in the analytic scales. Record your ratings on the form given.
4. Start rating the candidates on the DVD. Assess each performance in the order given on the DVD.
5. To make an assessment, start with the global assessment scale in order to decide approximately what level you think the speaker is. Assign a global rating during your first 2-3 minutes of the test. Then change to the analytic scales and assess the candidates on all five criteria (Range, Accuracy, Fluency, Interaction, Coherence). As you are watching, note features of candidate output to help you arrive at your final rating and refer to the scales throughout the test.
6. At the end of each performance, enter your marks for each criterion on the rating form. Add comments to explain your choice of marks, linking your comments to the wording of the band descriptors, and giving examples of relevant candidate output where possible. You may need to watch the performance again to cite examples but your assessments should not be changed. Please limit the number of viewings of each performance to a maximum of two.
7. NOTE: Even if you can recognize the tasks/test, and therefore level, from the materials used, it is important not to assign a CEF level automatically, based on your prior knowledge of the test. Use the descriptors in the CEF scales, so that you provide an independent rating, and support your choice of level by referring to the CEF.
8. Complete the feedback questionnaire.

Data Analysis

The marks awarded by the raters and the responses to the feedback questionnaire were compiled in an Excel spreadsheet. The marks were then exported into SPSS to allow for the calculation of descriptive statistics and frequencies. In addition, a Multi-Facet Rasch analysis (MFRA) was carried out using the programme FACETS. Candidate, rater, and criterion were treated as facets in an overall model. FACETS provided indicators of the consistency of the rater judgements and their relative harshness/leniency, as well as fair average scores for all candidates.

Findings

Ascertaining the consistency and severity of the raters was an important first step in the analysis, as it gave scoring validity evidence to the marks they had awarded. The FACETS output generated indices of rater harshness/leniency and consistency. As seen in Table 1, the results indicated a very small difference in rater severity (spanning 0.37 to -0.56 logits), which was well within an acceptable severity range and no cases of unacceptable fit (all outfit mean squares were within the 0.5 to 1.5 range), indicating high levels of examiner consistency. These results signalled a high level of homogeneity in the marking of the test, and provided scoring validity evidence (Weir, 2005) to the ratings awarded.

Table 1 FACETS output: Rater severity and consistency

Rater	Measure (logit)	Standard Error	Outfit MnSq
1	.37	.09	.62
2	-.24	.10	.80
3	.35	.09	1.32
4	-.19	.10	.70
5	.31	.09	1.10
6	-.20	.10	.78
7	-.56	.10	0.95
8	.16	.09	1.17

Phase 1 results

The results indicated very strong rater agreement in terms of typical and borderline performances at levels A2 to B2. As noted earlier, the internal team's operationalisation during sample selection had considered a performance at band 3/3.5 as typical of a given level and a performance at band 1.5/2 as borderline. This operationalisation had worked very well at levels A2 – B2 and the selection of performances which the internal group had felt to be typical/borderline (as based on marks awarded against the Main Suite scale) was confirmed by the high agreement among the raters in assigning CEF levels across all assessment criteria to those performances.

At levels C1 and C2 there was a lower level of agreement among raters regarding the level of the performances; in addition, the marking produced mostly candidates with differing proficiency profiles and so no pair emerged as comprising two typical candidates across all assessment criteria at the respective level. The raters' marks for each performance also resulted in a CEF level which was consistently lower than what was predicted by the Main Suite mark. It is not possible to be certain why the discrepancy between Main Suite and CEF levels occurred. It is likely that it is simply more difficult to mark higher-level candidates whose output is more complex. This possibility is supported by the frequency of awarded marks in the present marking exercise. With all C2 candidates, the level of agreement between the raters was lower than it was with the lower-proficiency candidates.

We can also hypothesize that the CEF C levels and the corresponding Main Suite CAE/CPE levels have developed more independently than the lower levels. While it is the case that the CEF and the Cambridge levels are the result of a policy of convergence (Brian North, personal communication), the historical and conceptual relationship between the CEF and Cambridge ESOL scales indicates that the work on the Waystage, Threshold and Vantage levels seems to have progressed very much hand-in-hand between the Council of Europe and Cambridge ESOL (Taylor & Jones, 2006), and so a "tight" relationship there is to be expected. This does not seem to have been the case with the higher levels. It can be hypothesized, therefore, that the two scales may have developed somewhat independently at the higher levels, and so the alignment between Main Suite and CEF levels at the C levels is different from the alignment at the lower levels. Milanovic (2009) also draws attention to the under-specification of the C levels within the CEFR scales.

The lower level of agreement among raters regarding candidates at C1 and C2, and the difficulty of finding a pair of candidates typical of these two levels across all criteria introduced the need for a subsequent marking exercise which focused on the top two levels only. The Phase 1 result led to a change in the group's working operationalisation of a typical and borderline performance as measured against the Main Suite scale as far as the C levels are concerned. As such, performances in the 4/4.5 band range were selected for the subsequent phase 2 of the study.

Phase 2 results

The results from this phase produced a typical pair of test takers at C1 across all CEF assessment criteria, with very high rater agreement. The pairs used at C2 had more varied performances and no pair emerged as having two typical C2 performances across all assessment criteria. This result is not altogether surprising given that the performances used in the present exercise came from the rater training pool where both typical and borderline cases should feature to allow for raters to develop familiarity with a range of test taker abilities. The C2 pair which was selected, therefore, included one typical candidate at that level across all criteria, while the second test taker in the pair showed borderline performance at the C1/C1+ level.

The selection of the final sample

Taking the statistical evidence into account the following five pairs of candidates emerged as the best illustrations for levels A2 to C2 (see table 2 below). Two of the candidates, Rino and Ben, had performances which did not consistently reflect one single CEFR level in certain criteria. In these cases, there was still acceptably high rater agreement as to the awarded adjacent CEFR level. Such performances are not surprising since oral ability develops on a continuum whereas assessment scales work in clear cut categories.

Table 2 Selected performances

Candidate	Overall level	Range	Accuracy	Fluency	Interaction	Coherence
Mansour	A2	A2	A2	A2	A2	A2
Arvids	A2	A2	A2	A2	A2	A2
Veronica	B1	B1	B1	B1	B1	B1
Melisa	B1	B1	B1	B1	B1	B1
Rino	B2	B2	B1+/B2	B2	B2	B2/B2+
Gabriela	B2	B2	B2	B2	B2	B2
Christian	C1	C1	C1	C1	C1	C1
Laurent	C1	C1	C1	C1	C1	C1
Ben	C1/C1+	C1	C1	C1/C1+	C1+	C1
Aliser	C2	C2	C2	C2	C2	C2

Caveat/Disclaimer

In compiling this selection of speaking tests, we have made our best effort to select typical performances. However, we would like to draw the reader/viewer's attention to the fact that educational contexts/traditions/teaching and assessment practices vary from one country to another and this may have an effect on perceptions of typical levels of performances. Our experience in benchmarking projects has indicated that in certain educational contexts aspects of fluency are more favoured than aspects of accuracy and vice versa.



UNIVERSITY *of* CAMBRIDGE
ESOL Examinations

English for Speakers of Other Languages



References

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bygate, M. (1988). *Speaking*. Oxford: Oxford University Press.
- Canale, M. (1983). On some dimensions of language proficiency. In J. W. Oller, Jr. (ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.
- Canale, M, and Swain, M (1980) Theoretical bases of communicative approaches to second language teaching and testing, *Applied Linguistics* 1, 1-47.
- Cizek, G. J., & Bunch, M. (2007). *Standard setting: A practitioner's guide*. Newbury Park, CA: Sage.
- Chalhoub-Deville, M. (2001). Task-based assessments: Characteristics and validity evidence. In M. Bygate, P., Skehan, & M. Swain (Eds.), *Researching pedagogic tasks*. (pp. 167-185). London: Longman.
- Council of Europe. (2001). *Common European Framework of References for Languages. Learning, Teaching, Assessment*. Cambridge. Cambridge University Press.
- Council of Europe. (2004). *Reference supplement to the preliminary pilot version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Language Policy Division.
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13(2), 23-51.
- Levitt. W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Milanovic, M. (2009). Cambridge ESOL and the CEFR. *Cambridge ESOL Research Notes*, 37 (August), 2-5.
- Saville, N. (2003). The Process of test development and revision within UCLES EFL. In C. Weir and M. Milanovic (Eds.). *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*. Cambridge: CUP.
- Shohamy, E. (2000). Assessment. In M. Celce-Murcia & E. Olshtain (Eds.), *Discourse and Context in Language Teaching* (pp. 201-215). Cambridge: Cambridge University Press.
- Skehan, P. (2001). Tasks and language performance assessment. In Bygate, M., Skehan, P. & Swain, M. (Eds.), *Researching pedagogic tasks*. (pp. 167-185). London: Longman.
- Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.
- Taylor, L. (2003). The Cambridge approach to speaking assessment. *Cambridge ESOL Research Notes*, 13, 2-4.
- Taylor, L. & Jones, N. (2006). Cambridge ESOL exams and the Common European Framework of Reference (CEFR). *Cambridge ESOL Research Notes*, 24 (May), 2-5.
- Van Ek, J. A. & Trim, J. L. M. (1998a). *Threshold 1990*. Cambridge: Cambridge University Press.
- Van Ek, J. A. & Trim, J. L. M. (1998b). *Waystage 1990*. Cambridge: Cambridge University Press.



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

Weir, C. (1990). *Communicative Language Testing*. New York: Prentice Hall.

Weir, C. J (2005) *Language Testing and Validation: An Evidence-Based Approach*. Oxford: Palgrave.



Appendix A: CEFR Assessment scales (Global and analytic)

Table 5.4: GLOBAL ORAL ASSESSMENT SCALE

C2	<i>Conveys finer shades of meaning precisely and naturally.</i> Can express him/herself spontaneously and very fluently, interacting with ease and skill, and differentiating finer shades of meaning precisely. Can produce clear, smoothly-flowing, well-structured descriptions.
C1+	
C1	<i>Shows fluent, spontaneous expression in clear, well-structured speech.</i> Can express him/herself fluently and spontaneously, almost effortlessly, with a smooth flow of language. Can give clear, detailed descriptions of complex subjects. High degree of accuracy; errors are rare.
B2+	
B2	<i>Expresses points of view without noticeable strain.</i> Can interact on a wide range of topics and produce stretches of language with a fairly even tempo. Can give clear, detailed descriptions on a wide range of subjects related to his/her field of interest. Does not make errors which cause misunderstanding.
B1+	
B1	<i>Relates comprehensibly the main points he/she wants to make.</i> Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair may be very evident. Can link discrete, simple elements into a connected sequence to give straightforward descriptions on a variety of familiar subjects within his/her field of interest. Reasonably accurate use of main repertoire associated with more predictable situations.
A2+	
A2	<i>Relates basic information on, e.g. work, family, free time etc.</i> Can communicate in a simple and direct exchange of information on familiar matters. Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident. Can describe in simple terms family, living conditions, educational background, present or most recent job. Uses some simple structures correctly, but may systematically make basic mistakes.
A1+	
A1	<i>Makes simple statements on personal details and very familiar topics.</i> Can make him/herself understood in a simple way, asking and answering questions about personal details, provided the other person talks slowly and clearly and is prepared to help. Can manage very short, isolated, mainly pre-packaged utterances. Much pausing to search for expressions, to articulate less familiar words.
Below A1	Does not reach the standard for A1.
	<ul style="list-style-type: none"> • <i>Use this scale in the first 2-3 minutes of a speaking sample to decide approximately what level you think the speaker is.</i> • <i>Then change to Table 5.5 (CEF Table 3) and assess the performance in more detail in relation to the descriptors for that level.</i>



Table 5.5: ORAL ASSESSMENT CRITERIA GRID (CEF Table 3)

	RANGE	ACCURACY	FLUENCY	INTERACTION	COHERENCE
C2	Shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity. Also has a good command of idiomatic expressions and colloquialisms.	Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).	Can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it.	Can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly. Can interweave his/her contribution into the joint discourse with fully natural turntaking, referencing, allusion making etc.	Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.
C1+					
C1	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say.	Consistently maintains a high degree of grammatical accuracy; errors are rare, difficult to spot and generally corrected when they do occur.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.	Can select a suitable phrase from a readily available range of discourse functions to preface his remarks in order to get or to keep the floor and to relate his/her own contributions skilfully to those of other speakers.	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2+					
B2	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on most general topics, without much conspicuous searching for words, using some complex sentence forms to do so.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstanding, and can correct most of his/her mistakes.	Can produce stretches of language with a fairly even tempo; although he/she can be hesitant as he or she searches for patterns and expressions, there are few noticeably long pauses.	Can initiate discourse, take his/her turn when appropriate and end conversation when he/she needs to, though he/she may not always do this elegantly. Can help the discussion along on familiar ground confirming comprehension, inviting others in, etc.	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some "jumpiness" in a long contribution.
B1+					
B1	Has enough language to get by, with sufficient vocabulary to express him/herself with some hesitation and circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Uses reasonably accurately a repertoire of frequently used "routines" and patterns associated with more predictable situations.	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.	Can initiate, maintain and close simple face-to-face conversation on topics that are familiar or of personal interest. Can repeat back part of what someone has said to confirm mutual understanding.	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2+					
A2	Uses basic sentence patterns with memorised phrases, groups of a few words and formulae in order to communicate limited information in simple everyday situations.	Uses some simple structures correctly, but still systematically makes basic mistakes.	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident.	Can ask and answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her own accord.	Can link groups of words with simple connections like "and", "but" and "because".
A1+					
A1	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorised repertoire.	Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words, and to repair communication.	Can ask and answer questions about personal details. Can interact in a simple way but communication is totally dependent on repetition, rephrasing and repair.	Can link words or groups of words with very basic linear connectors like "and" or "then".
Below A1					



Appendix B: Example of a Rating form

SELF-ASSESSMENT TASK

Learner's name	Initial impression (CEFR Table 5.4)	Detailed analysis (CEFR Table 5.5)					Comments
	CEFR level	Range CEFR level	Accuracy CEFR level	Fluency CEFR level	Interaction CEFR level	Coherence CEFR level	

RATING TASK

Learner's name	Initial impression (CEFR Table 5.4)	Detailed analysis (CEFR Table 5.5)					Comments
	CEFR level	Range CEFR level	Accuracy CEFR level	Fluency CEFR level	Interaction CEFR level	Coherence CEFR level	
RINO							
GABRIELA							