

Standard Setting for the ALCE Examination

The Advanced Level Certificate in English (ALCE) examination is divided into four parts: a Listening Section of 40 items; a Grammar, Vocabulary and Reading Section, comprising 40 grammar items, 40 vocabulary items and 20 reading items; a Writing Section with one task; and a Speaking Section.

A separate total score and cut-off score is provided for each of the four Sections, and these are then recalibrated to give equal weighting in the final scoring. The overall pass/fail grade is awarded through aggregate scoring, which allows candidates a narrow fail in one Section, provided their overall score is above the combined overall cut-off score.

The procedures for the standard setting of the ALCE examination follow those set out in the *Common European Framework of Reference for Languages*, the *Preliminary Pilot Version of the Manual on Relating Language Examinations to the Common European Framework of Reference for Languages*, and the *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages*.

The procedures used were as follows:

- the selection of the most appropriate and effective method of standard setting for each Section of the ALCE examination
- the selection of a large number of judges, based on their qualifications and experience of language teaching and testing
- the training of the judges on the appropriate criteria, as defined in the *CEFR* and the *Manual for the Advanced Level Certificate in English*
- the establishment by the judges of cut-off scores for each Section
- data analysis of the judges' cut-off scores

Since the Sections of the ALCE examination which test receptive skills, the Listening Section and Grammar, Vocabulary, Reading Section, are machine scored, the standard setting for these two Sections followed precisely the same procedures. However, as the Sections which test productive skills, the Speaking Section and the Writing Section, are scored by individual raters, the standard setting for each was performed separately.

The standard setting was carried out on the items used in the January 2006 ALCE examination, and occurred in successive meetings during January and February 2006. Standard setting was again carried out on the Writing Section in June 2006 and this is described below.

1. Listening Section & Grammar, Vocabulary, Reading Section

Selection of the most appropriate and effective method of standard setting

Standard setting for the Listening Section and Grammar, Vocabulary, Reading Section of the ALCE examination was carried out using the modified Angoff method (Taube, 1997). This method was chosen due to its appropriateness for a multiple-choice format and its efficiency of use.

In carrying out this method, carefully selected judges first took part in a thorough familiarization and training session before assessing each Section (see Training below). They were then asked to examine each item and assess the percentage of candidates minimally acceptable at the level that would be likely to choose the correct answer choice for this item (see Setting Cut-Off Scores

below). The judges' decisions were collated and a cut-off score for each Section set (see Data Analysis below).

Selection of judges

The number of judges suggested by the literature is 5 – 15 (Felianka Kaftandjieva, 2004, p. 23). However, for greater precision in setting the cut-off score estimates, 20 judges were used for the standard setting of the Listening and the Grammar, Vocabulary, Reading Sections.

The judges used were selected for their qualifications and experience of teaching English as a Foreign Language and of testing in this field. Ten of the judges had prior experience of standard setting procedures, of which nine also had experience of item writing for high stakes tests. All the judges were practicing teachers at the time of the standard setting and all had experience as Oral Examiner for a variety of tests at a variety of levels.

Each of the judges was asked to complete a Background Information Form (see Figure 1, below). The form records a summary of their educational and teaching experience, as well as any experience they may have had with the ALCE examination, for inclusion in the documentation of the standard setting. In addition, a Curriculum Vitae of each judge was also kept on file.

The completed Background Information Forms of each of the judges are documented in Appendix A.

BACKGROUND INFORMATION FORM

Please fill in the following information on your qualifications and experience:

PERSONAL DETAILS

NAME

NATIONALITY

EDUCATION

BA: SUBJECT

MA: SUBJECT

OTHER:

TEACHING EXPERIENCE

NUMBER OF YEARS OF TEACHING EXPERIENCE

LEVELS TAUGHT

EXPERIENCE OF THE ALCE EXAMINATION

HAVE YOU EVER TAUGHT STUDENTS FOR THE ALCE?

HAVE YOU EVER EXAMINED FOR THE ALCE?

DID YOU HAVE ANY STUDENTS FOR THE JANUARY 2006 TEST?

Figure 1: Background Information Form for Judges

Training of the Judges

In order to familiarize the judges with the CEFR and the CEFR scales, first, they were introduced to the rationale and context of the CEFR. The next part of the judges' training involved familiarization with the 'can do' statements of the CEFR.

Before each Section of the ALCE examination, the judges were required to examine the 'can do' statements for that Section, both the overall descriptors and those more specific, in order to rank them according to the six scales of the CEFR. The participants carried out this task individually and then compared their rankings in pairs and groups.

A detailed discussion of the rankings followed, with participants justifying their decisions and receiving feedback on the order as defined in the CEFR. All present were able to reach a consensus on the ranking without difficulty and developed their awareness of what defines each level as described by the CEFR. Participants were then directed to focus on the descriptors for the C1 level – the level which the ALCE examination is aimed at. This level was discussed in detail and contrasted with the C2 level above and the B2 level below in order to clarify exactly what C1 means in terms of the CEFR. The descriptors used for the standard setting are recorded in Appendix B.

Having established the criteria for the level, the judges were then trained in the criteria for making a decision on actual test items. They were asked to examine each of the test items in terms of the percentage of candidates minimally acceptable at the level that would be likely to choose the correct answer choice.

Establishment of cut-off scores

Familiarization was carried out for Listening, Grammar, Vocabulary and Reading in turn. After each familiarization activity, the judges examined the items relating to that particular language area. For the Listening Section, a recording was played with a longer pause allowed after each item for the judges to record their decision.

Forms were given to the judges to record their decisions for each of the language areas and the same procedures followed for each (see Appendix C). They first recorded their individual decision and were then given the opportunity to discuss their findings in pairs and groups. The Key was given at this point, after which they once again recorded their decision based on the discussion. The judges were then given empirical statistics on how students performed on the items, followed by a plenary discussion, and then they recorded their final decision. In this way, each judge recorded three impression marks for each of the items (Figure 2, below).

Judge's Name:	
Judge's Code:	

GRAMMAR

	Individual	Discussion	Empirical
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

	Individual	Discussion	Empirical
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			

	Individual	Discussion	Empirical
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			

	Individual	Discussion	Empirical
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			

Figure 2: Example of Rating Form - Grammar

Data analysis of the cut-off scores

The judges' decisions were collated and merged to give the overall cut-off scores for each section.

The figures for each item of the Listening Section and Grammar, Vocabulary, Reading Section were entered into a database for each Section (see Figure 3: Listening, and Figure 4: Grammar, Vocabulary, Reading, below). The median, average and standard deviation for each rater were calculated, as well as the minimum and maximum score given. In this way, intra-rater and inter-rater correlations could be examined.

An average rating for each item was produced from the figures given and the average of these ratings was calculated to give the cut-off score. For the Listening Section, the cut-off score was judged to be 56 percent, rounded down from 56.2. For the Grammar, Vocabulary, Reading Section, the cut-off score was calculated at 57 percent, rounded down from 57.36.

Conclusion

The cut-off score obtained from the standard setting for both Sections was lower than that expected. It had been anticipated that the cut-off score should be approximately 65 percent for each.

The results of the standard setting imply that the ALCE examination may be pitched at a slightly higher level than intended, with the inclusion of many more difficult items than should be the case for an examination aimed at the lower end of C1 level. Some of the items certainly appear to be high C1 level, according to the judges' ratings.

These insights gained from these results will influence the choice of items for the next administration of the ALCE examination, with fewer items outside the desired range being included.

It is intended that a further standard setting will take place within the next 12 months in order to verify that the difficulty level of the examination is aimed at the lower reaches of C1 of the CEF.

Figure 3: RESULTS OF LISTENING STANDARD SETTING

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Med	AV	SD	Min	Max
1	70	90	65	70	85	70	80	75	80	70	80	78	70	70	80	85	80	65	70	75	75	75,4	7,00	65	90
2	95	75	30	70	40	50	70	50	80	78	70	77	65	60	80	50	80	36	40	60	67,5	62,8	17,78	30	95
3	40	65	55	60	75	60	70	60	60	60	60	55	60	60	75	50	70	25	50	55	60	58,25	11,50	25	75
4	75	50	65	55	15	70	50	60	60	60	60	62	65	60	70	50	85	20	40	55	60	56,35	16,56	15	85
5	30	40	50	50	75	40	50	40	60	40	50	60	45	50	70	60	75	40	20	55	50	50	14,14	20	75
6	20	65	70	55	40	50	70	60	60	30	50	60	45	50	75	65	20	50	29	55	52,5	50,95	16,18	20	75
7	90	70	70	50	70	60	70	45	55	50	50	40	50	30	50	50	40	40	80	60	50	56	15,10	30	90
8	75	80	69	60	80	85	90	80	70	10	90	75	80	80	90	90	85	65	70	75	80	74,95	17,56	10	90
9	75	80	65	60	75	70	90	65	75	65	80	65	70	70	80	70	90	57	70	70	70	72,1	8,75	57	90
10	50	60	65	55	70	60	80	60	65	60	65	52	60	60	65	50	60	50	70	60	60	60,85	7,52	50	80
11	90	85	80	80	75	80	70	90	80	70	85	78	80	75	75	85	75	70	90	79	80	79,6	6,39	70	90
12	70	52	50	60	65	65	60	75	50	65	65	56	65	60	65	60	50	50	80	55	60	60,9	8,47	50	80
13	76	57	70	70	51	60	70	70	60	60	60	62	60	60	65	75	50	40	90	55	60	63,05	10,86	40	90
14	25	60	60	70	60	65	70	70	55	60	60	60	60	65	65	75	50	50	80	59	60	60,95	11,32	25	80
15	90	80	75	70	75	85	80	80	80	70	85	85	80	80	85	90	75	60	80	79	80	79,2	7,12	60	90
16	50	65	60	65	60	45	70	75	50	70	50	60	65	70	80	40	50	50	70	70	62,5	60,75	11,04	40	80
17	85	67	70	70	70	80	75	75	80	70	70	65	70	70	75	80	75	50	70	75	70	72,1	7,24	50	85
18	75	50	60	60	70	65	60	75	60	55	55	60	55	60	70	70	40	40	60	45	60	59,25	10,17	40	75
19	50	80	80	70	60	75	80	95	85	70	70	80	80	80	85	80	50	55	80	85	80	74,5	12,24	50	95
20	30	60	65	65	70	50	70	80	65	60	55	65	60	65	70	70	50	50	80	60	65	62	11,40	30	80
21	40	40	40	60	60	60	50	55	45	40	60	29	40	30	50	60	65	60	40	40	47,5	48,2	11,12	29	65
22	40	50	30	40	40	70	70	55	45	31	50	31	40	30	60	40	25	60	30	55	40	44,6	13,65	25	70
23	55	40	40	40	50	50	50	60	50	25	50	30	30	30	30	40	40	40	40	49	40	41,95	9,61	25	60
24	50	50	40	50	60	60	70	60	60	40	55	50	50	50	50	40	25	40	40	75	50	50,75	11,62	25	75
25	40	50	40	50	40	40	60	50	50	37	50	40	50	40	40	50	25	40	60	55	45	45,35	8,54	25	60
26	90	75	65	60	80	85	70	80	80	70	80	70	80	70	75	60	75	50	80	65	75	73	9,65	50	90
27	75	57	65	50	50	50	60	65	65	55	55	55	55	50	50	50	50	40	40	50	52,5	54,35	8,49	40	75
28	40	45	50	50	40	40	50	45	40	40	50	32	30	30	30	50	50	35	40	45	40	41,6	7,30	30	50
29	40	50	50	40	60	45	60	40	60	60	50	55	55	60	40	60	60	40	30	43	50	49,9	9,53	30	60
30	30	35	50	40	50	35	60	50	45	40	50	35	40	40	35	60	20	45	20	38	40	40,9	10,85	20	60
31	40	30	30	50	60	60	45	55	40	40	50	40	40	40	50	40	60	35	40	40	40	44,25	9,22	30	60
32	70	40	50	50	60	45	50	45	40	40	50	40	40	35	45	45	60	40	40	30	45	45,75	9,36	30	70
33	70	50	50	50	60	50	60	55	55	50	60	55	45	50	55	50	55	30	60	40	52,5	52,5	8,35	30	70
34	30	45	20	40	50	30	50	40	30	40	40	55	30	30	65	60	55	35	70	35	40	42,5	13,43	20	70
35	45	40	30	40	40	20	40	45	30	30	40	30	30	40	50	30	40	30	28	33	36,5	35,55	7,42	20	50
36	80	55	60	50	50	60	70	60	60	55	60	65	50	50	60	45	65	50	60	47	60	57,6	8,52	45	80
37	50	47	50	40	50	50	50	45	50	40	60	50	45	45	45	40	65	45	40	30	46	46,85	7,47	30	65
38	40	50	60	50	60	60	70	70	50	50	50	55	55	45	50	50	65	50	70	45	50	54,75	8,81	40	70
39	30	45	40	50	50	25	50	40	25	40	40	35	40	35	40	40	55	33	60	35	40	40,4	9,24	25	60
40	39	50	40	50	50	45	65	45	40	50	65	45	46	40	45	50	40	45	60	37	45	47,35	8,08	37	65
	56,38	56,88	54,35	55,38	58,53	56,63	64,38	61,00	57,25	51,15	59,38	54,80	54,40	52,88	60,88	57,63	56,13	45,15	56,68	54,23	55,94	56,20		10,00	95,00

Figure 4: RESULTS OF GRAMMAR, VOCABULARY, READING STANDARD SETTING

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Med	Av	SD	Min	Max
41	50	67	65	70	65	65	60	63	50	60	60	65	70	70	70	70	75	50	50	60	65	62,75	7,72	50	75
42	30	59	60	70	65	60	60	60	50	50	60	60	70	60	70	60	50	60	70	52	60	58,80	9,35	30	70
43	60	75	80	70	65	75	70	80	70	60	75	70	70	75	70	80	25	55	70	72	70	68,35	12,19	25	80
44	70	55	70	60	65	60	60	10	70	50	50	60	60	60	70	70	75	45	50	55	60	58,25	14,07	10	75
45	70	50	60	70	65	55	65	60	50	60	60	50	60	50	50	70	75	45	60	45	60	58,50	8,90	45	75
46	60	70	60	50	60	50	55	60	50	60	60	65	65	60	60	60	50	45	60	55	60	57,75	6,17	45	70
47	70	60	65	50	60	60	55	60	45	50	50	65	50	55	60	65	25	40	60	60	60	55,25	10,32	25	70
48	45	60	70	70	70	70	65	75	65	60	70	60	60	60	70	70	75	45	60	70	67,5	64,50	8,41	45	75
49	30	72	75	70	70	70	75	80	75	70	60	65	65	70	80	75	75	43	70	69	70	67,95	11,98	30	80
50	30	75	70	70	65	70	75	80	70	60	70	80	75	80	70	60	50	40	70	55	70	65,75	13,31	30	80
51	50	80	75	70	60	75	80	88	70	55	70	70	75	70	80	80	75	56	40	65	70	69,20	11,83	40	88
52	45	50	60	50	60	65	50	75	60	70	60	50	45	50	60	55	75	35	70	62	60	57,35	10,61	35	75
53	70	70	60	50	60	65	65	65	55	60	60	65	65	60	65	65	75	30	60	57	62,5	61,10	9,20	30	75
54	65	66	70	70	60	60	65	66	60	60	70	65	65	60	70	50	50	35	50	50	62,5	60,35	9,16	35	70
55	30	50	55	50	50	60	50	46	35	50	50	55	60	50	65	50	75	38	60	44	50	51,15	10,25	30	75
56	45	70	50	50	60	65	60	65	60	55	55	50	60	60	60	60	50	42	60	49	60	56,30	7,24	42	70
57	75	70	60	70	65	50	45	60	50	60	60	50	50	50	60	50	75	45	40	40	55	56,25	10,87	40	75
58	60	67	65	50	65	60	60	70	78	70	55	75	65	70	60	60	50	45	70	65	65	63,00	8,49	45	78
59	50	55	70	70	65	60	60	70	65	60	60	70	65	70	65	65	75	50	60	60	65	63,25	6,74	50	75
60	40	60	70	70	65	60	50	51	60	60	65	60	50	60	60	65	50	35	50	55	60	56,80	9,18	35	70
61	30	61	50	70	60	40	40	55	45	50	60	40	35	40	50	50	25	50	45	37	47,5	46,65	11,22	25	70
62	55	50	50	50	60	50	50	55	60	50	55	45	55	40	60	60	50	48	50	57	50	52,50	5,39	40	60
63	60	55	60	70	60	60	55	75	60	60	75	55	50	50	60	60	75	50	60	55	60	60,25	7,86	50	75
64	65	60	70	70	70	70	70	80	68	70	55	72	60	70	70	50	50	55	60	48	69	64,15	8,83	48	80
65	80	77	75	50	70	70	70	85	70	65	70	72	70	70	70	70	75	70	70	70	70	70,95	6,63	50	85
66	66	65	65	50	65	60	63	65	80	70	65	60	60	65	70	75	50	45	70	62	65	63,55	8,26	45	80
67	65	80	80	50	60	60	75	75	70	65	80	80	65	70	70	65	50	65	80	71	70	68,80	9,31	50	80
68	35	50	57	50	65	50	55	60	65	65	70	65	60	60	75	70	75	55	60	52	60	59,70	9,72	35	75
69	50	50	40	70	60	50	39	60	50	50	50	50	60	50	70	70	50	50	50	58	50	53,85	8,85	39	70
70	80	85	90	70	60	50	80	90	80	80	80	80	70	80	85	80	75	70	80	62	80	76,35	10,09	50	90
71	60	60	60	50	70	80	55	70	60	70	60	60	55	55	80	50	75	45	60	68	60	62,15	9,72	45	80
72	75	65	70	50	70	60	70	75	70	70	60	74	60	65	40	55	75	60	75	69	69,5	65,40	9,37	40	75
73	45	40	50	50	50	50	40	55	45	50	45	40	70	40	70	60	50	50	40	40	50	49,00	9,12	40	70
74	35	60	60	50	60	50	60	75	60	60	60	60	40	60	65	60	50	47	60	42	60	55,70	9,45	35	75
75	40	67	60	70	65	60	60	70	70	60	75	60	50	60	70	65	75	50	75	65	65	63,35	9,10	40	75
76	50	55	60	70	60	65	60	65	65	60	65	50	50	50	70	60	50	53	50	35	60	57,15	8,70	35	70
77	50	67	65	50	70	65	70	75	65	70	70	70	40	70	30	65	50	60	70	37	65	60,45	12,98	30	75
78	20	20	50	70	50	40	20	70	25	30	50	25	40	20	30	65	25	50	20	30	30	37,50	17,28	20	70
79	60	30	50	60	50	40	25	70	40	30	60	40	50	30	60	65	25	60	30	40	45	45,75	14,53	25	70
80	75	40	50	60	50	50	40	70	50	50	50	40	50	40	60	75	75	70	50	50	50	54,75	12,19	40	75
81	35	50	60	70	65	50	50	55	50	75	60	50	45	55	60	55	50	50	50	45	50	54,00	9,12	35	75
82	90	60	50	70	65	50	60	80	70	60	65	60	50	60	70	80	50	60	60	60	60	63,50	10,77	50	90
83	50	70	65	70	70	60	70	75	75	80	75	70	65	70	75	70	75	60	70	65	70	69,00	6,81	50	80

84	50	48	50	70	60	50	50	55	65	60	55	55	50	50	60	70	50	55	60	52	55	55,75	6,77	48	70
85	60	35	50	50	60	50	40	50	40	50	60	40	50	40	40	65	75	52	40	40	50	49,35	10,44	35	75
86	45	57	50	70	60	50	65	60	50	60	65	50	55	55	50	65	50	25	50	45	52,5	53,85	9,87	25	70
87	50	50	50	70	50	50	45	55	50	50	60	45	55	45	60	65	50	35	50	45	50	51,50	7,80	35	70
88	65	85	80	70	75	75	78	75	80	80	70	75	75	75	50	65	75	45	80	72	75	72,25	9,85	45	85
89	25	30	50	50	50	40	40	55	40	50	50	40	50	35	30	70	25	40	50	59	45	43,95	11,58	25	70
90	25	50	60	70	65	60	60	75	70	70	60	60	60	60	50	70	50	45	60	50	60	58,50	11,37	25	75
91	55	35	60	70	50	50	40	55	50	40	60	40	30	30	40	50	50	48	40	40	49	46,65	10,40	30	70
92	65	70	70	70	70	60	60	60	70	70	60	60	50	60	70	60	75	55	50	50	60	62,75	7,69	50	75
93	45	55	50	70	50	50	40	60	50	50	50	45	60	50	50	60	50	50	40	42	50	50,85	7,34	40	70
94	65	55	50	50	55	50	50	65	60	50	60	55	55	50	55	70	50	55	50	45	55	54,75	6,38	45	70
95	25	45	40	50	50	50	40	55	40	40	60	40	50	40	45	60	50	60	40	39	45	45,95	8,86	25	60
96	85	40	50	60	60	50	40	60	50	50	56	40	45	40	50	60	75	55	40	40	50	52,30	12,20	40	85
97	35	60	61	70	70	65	70	80	68	70	80	70	70	70	60	50	50	61	70	59	69	64,45	10,57	35	80
98	50	50	60	60	60	50	50	60	50	50	70	45	55	45	50	65	90	40	40	40	50	54,00	11,88	40	90
99	30	45	60	60	60	55	50	60	50	55	56	50	50	45	50	60	75	55	40	40	52,5	52,30	9,68	30	75
100	50	60	65	60	60	55	55	75	60	60	65	60	60	60	55	65	50	60	60	50	60	59,25	5,91	50	75
101	35	65	60	60	60	50	60	70	50	70	65	65	65	65	70	70	75	55	60	55	62,5	61,25	9,16	35	75
102	35	55	50	70	60	40	40	50	50	40	50	40	40	40	40	50	25	50	40	45	42,5	45,50	9,72	25	70
103	80	50	50	60	60	60	60	80	70	60	65	60	60	60	65	60	75	48	60	60	60	62,15	8,66	48	80
104	90	50	50	60	70	55	50	75	50	55	65	50	55	50	50	70	25	65	50	58	55	57,15	13,21	25	90
105	50	55	50	70	70	50	60	80	70	50	70	55	65	50	70	65	75	55	70	61	63	62,05	9,64	50	80
106	55	50	60	60	70	65	60	70	70	60	75	67	70	70	65	40	50	45	70	55	62,5	61,35	9,63	40	75
107	30	55	60	60	60	55	60	70	60	60	65	60	55	55	60	45	50	48	60	60	60	56,40	8,38	30	70
108	75	70	70	60	70	65	70	85	75	70	75	70	75	70	65	60	50	65	70	65	70	68,75	7,23	50	85
109	30	40	50	60	50	40	40	65	50	50	60	40	30	35	30	50	25	50	40	35	40	43,50	11,13	25	65
110	90	60	60	60	60	60	60	80	70	60	75	60	55	60	55	60	50	66	60	52	60	62,65	9,55	50	90
111	25	30	45	60	55	40	30	60	45	40	60	30	30	30	30	65	25	60	30	35	37,5	41,25	13,85	25	65
112	70	77	70	60	65	70	87	77	80	70	80	80	70	70	80	65	75	55	70	60	70	71,55	8,08	55	87
113	50	37	50	60	60	40	30	50	30	30	30	40	30	30	40	60	25	45	30	60	40	41,35	12,19	25	60
114	50	50	60	60	60	50	50	70	50	50	50	50	50	45	50	60	50	50	50	55	50	53,00	5,94	45	70
115	60	50	60	70	70	50	50	70	50	50	50	50	50	45	50	60	50	48	50	55	50	54,40	7,84	45	70
116	25	40	50	60	60	40	40	40	35	40	40	50	50	40	40	40	25	45	40	45	40	42,25	8,96	25	60
117	45	60	65	70	60	50	50	70	50	65	75	60	60	60	40	50	50	45	55	50	57,5	56,50	9,47	40	75
118	50	50	50	60	50	45	40	60	50	50	60	50	55	50	60	40	25	45	50	48	50	49,40	8,26	25	60
119	79	67	70	60	65	55	70	80	65	70	65	65	60	60	70	45	75	45	50	45	65	63,05	10,63	45	80
120	85	40	50	60	60	50	45	70	50	40	60	45	55	40	40	60	75	48	40	50	50	53,15	12,63	40	85
121	75	55	80	95	75	75	70	80	70	50	80	65	60	55	60	90	90	50	60	35	70	68,50	15,48	35	95
122	95	97	90	95	80	95	80	80	90	80	80	77	70	70	80	72	90	60	70	75	80	81,30	10,33	60	97
123	75	65	60	65	75	70	80	70	70	75	60	60	60	65	70	85	75	65	60	70	70	68,75	7,23	60	85
124	30	50	65	50	50	60	80	45	60	50	40	45	40	45	60	50	50	30	40	35	50	48,75	12,13	30	80
125	30	50	70	45	70	80	60	50	60	60	40	40	40	40	50	35	75	30	40	60	50	51,25	14,95	30	80
126	30	35	50	30	40	50	50	60	50	40	40	40	50	40	25	30	25	40	30	35	40	39,50	9,72	25	60
127	70	50	60	50	50	50	70	50	50	50	55	55	45	50	52	40	50	50	60	40	50	52,35	7,84	40	70
128	30	40	50	45	70	60	80	40	50	40	50	50	45	45	48	35	25	60	50	55	49	48,40	12,78	25	80
129	65	57	60	45	60	40	80	60	60	60	55	60	50	50	60	50	50	40	70	60	60	56,60	9,60	40	80
130	20	47	45	50	60	50	70	50	35	55	40	45	35	40	41	45	50	40	25	35	45	43,90	11,35	20	70
131	70	50	50	50	60	50	70	50	50	50	45	46	45	45	46	40	50	40	30	40	50	48,85	9,40	30	70

132	30	55	50	50	50	65	70	70	50	70	60	54	50	50	51	50	50	40	50	55	50	53,50	9,86	30	70
133	30	55	50	60	50	60	80	50	60	70	55	50	45	50	53	45	75	50	30	55	51,5	53,65	12,34	30	80
134	70	70	50	60	75	65	80	85	65	50	70	65	60	65	65	60	75	60	60	65	65	65,75	8,78	50	85
135	75	60	57	55	70	80	60	70	60	50	60	55	50	60	60	65	65	48	60	60	60	61,00	8,14	48	80
136	70	70	60	55	60	90	70	75	70	70	50	55	65	50	50	60	65	55	60	65	62,5	63,25	9,90	50	90
137	60	50	50	45	60	80	70	60	45	40	30	40	40	40	45	50	50	55	50	50	50	50,50	11,46	30	80
138	65	70	56	55	70	80	80	75	60	60	60	60	60	50	60	60	75	45	60	50	60	62,55	9,76	45	80
139	50	60	50	50	50	70	70	60	50	50	55	50	55	50	50	70	65	55	50	55	52,5	55,75	7,48	50	70
140	50	65	60	55	50	70	70	60	60	54	50	58	55	50	60	75	75	50	60	60	60	59,35	8,13	50	75

53,32 54,04 57,35 60,35 60,83 56,94 57,83 65,58 56,99 56,31 59,33 53,56 52,99 51,25 54,94 59,26 56,11 50,93 52,22 51,08 55,60 57,36 50,93 65,58
56,78

2. Writing Section: Benchmarking & Standard Setting

The Writing Section of the ALCE comprises one writing task, chosen from two options. The first option is an essay arguing a point of view, while the second is a report. For both options, points which should be developed in writing the task are provided. Candidates have thirty minutes to produce their essay or report.

Each report or essay is initially graded by two raters, independently of one another. These ratings are collated and where there is a substantial difference between the two ratings for a candidate's writing, it is then graded by a third rater. Each rater is given a Rater Code, allowing for intra-rater and inter-rater reliability analysis.

Raters are provided with the criteria developed for the Writing Section of the ALCE, developed to reflect the C1 level of the CEFR (see Figure 5, below). They are trained using graded samples of candidates' writing and the CEFR scales for written production.

The standard setting for the ALCE Writing Section took place in June 2006 and was carried out by 14 judges.

Method of benchmarking & standard setting

Judges were trained in the CEFR writing scales and the ALCE scoring criteria, and then examined the written work of candidates from the June examination. After individual and collective discussion, a grade was assigned to each piece of written work to provide a standard for raters.

Selection of judges

The judges were selected for their knowledge and experience of teaching English at this level and their knowledge of testing procedures. Many of the judges also participated in other sections of the standard setting procedures.

Training of the judges

The training of the judges was carried out in a similar manner to that of the Listening and GVR sections. Having examined, sorted and discussed the relevant descriptors from the CEFR, the judges were given the rating descriptors for the ALCE (Fig. 5). These were discussed and key features distinguishing the scoring criteria were emphasized.

Figure 5: ALCE WRITING SECTION: DESCRIPTORS OF SALIENT FEATURES

	Task Completion	Topic Development	Range of Language	Accuracy	Cohesion
FOCUS	<p>Relevance to the task set</p> <p>Appropriacy of style & format for the genre</p> <p>Number of prompts developed</p>	<p>Text organization:</p> <ul style="list-style-type: none"> • introduction with thesis statement • topic sentences & elaboration • conclusion 	<p>Complexity of:</p> <ul style="list-style-type: none"> • grammar • syntax • vocabulary 	<p>Accuracy of:</p> <ul style="list-style-type: none"> • grammar • syntax • vocabulary 	<p>Discourse features:</p> <ul style="list-style-type: none"> • connections between paragraphs & between sentences • reference words
<p>3</p> <p>HIGH PASS</p> <p><i>Effective Writing</i></p>	<p>Completes the task fully and relevantly, covering all aspects of the topic effectively.</p> <p>Consistently uses an appropriate style and format.</p> <p>Uses effectively at least the minimum number of prompts for the chosen task.</p>	<p><i>Fully</i> develops the topic</p> <p>Provides an effective introduction and conclusion</p> <p><i>Effectively</i> describes, supports, and expands at length, using well-chosen examples and/or detailed reasons.</p>	<p>Shows great flexibility in formulating ideas by using a wide range of linguistic forms effectively.</p> <p>Uses a broad range of syntactical structures, and displays a wide lexical range.</p>	<p>Maintains a fairly high degree of grammatical and syntactical accuracy, even in complex structures.</p> <p>Consistently chooses vocabulary appropriate for the topic and style.</p> <p>Minor errors may occur.</p>	<p>Consistently uses a wide range of organizational patterns and cohesive devices in an effective manner.</p>
<p>2</p> <p>PASS</p> <p><i>Satisfactory Writing</i></p>	<p>Completes the task satisfactorily and mostly relevantly, covering some aspects of the topic more effectively than others.</p> <p>Flexibility in style is somewhat limited, but style and format are mostly appropriate for the genre.</p> <p>Uses satisfactorily at least the minimum number of prompts for the chosen task.</p>	<p>Develops the topic satisfactorily.</p> <p>Provides a relevant introduction.</p> <p>Provides an appropriate conclusion</p> <p>Describes, supports, and expands at some length, using relevant examples and/or reasons.</p>	<p>Displays a good command of a satisfactory range of language to accomplish the task.</p> <p>Uses some variety of syntactical structures and displays sufficient lexical range to accomplish the task satisfactorily.</p>	<p>Maintains a fairly high degree of grammatical and syntactical accuracy.</p> <p>Mostly chooses vocabulary appropriate for the topic and style.</p> <p>Some errors may occur in more complex structures, but do not impede communication.</p>	<p>Uses a range of organizational patterns and linking devices in order to achieve the smooth flow of the text</p>
<p>1*</p> <p>FAIL</p> <p><i>Inappropriate Writing</i></p>	<p>Has not completed the task satisfactorily because of irrelevancy, ineffective use of the prompts or through misunderstanding the task.</p> <p>Style and format are mostly not appropriate for the genre.</p> <p>Does not use the minimum number of prompts for the chosen task.</p>	<p>Does not satisfactorily develop the topic</p> <p>Does not provide an appropriate introduction and conclusion</p> <p>May mention the prompts, but does not provide sufficient subsidiary points, detail, reasons or relevant examples.</p>	<p>Uses a limited range of language in attempting to accomplish the task.</p> <p>Uses a very limited variety of complex syntactical forms and displays a lack of lexical range.</p>	<p>Errors may occur in both simple and complex structures, which may sometimes impede communication.</p> <p>Some vocabulary may be inappropriate for the topic and style.</p>	<p>Uses a limited range of cohesive devices to link sentences and paragraphs.</p> <p>Inappropriate linking between sentences and paragraphs may hinder the smooth flow of the text.</p>

Establishment of benchmarking scores

The judges examined 15 essays and 15 reports from the June 2006 administration. These were marked by the judges and then the marks were discussed until a consensus was reached for each piece of writing.

Conclusion

The benchmarked essays and reports (Appendix ...) formed the basis for rater training for the Writing Section of the June ALCE examination. During the benchmarking procedure, it was agreed by all judges that the descriptors for a Pass in the Writing Section of the ALCE examination correspond to the level set out by the descriptors from the CEFR at C1 level.

3. Speaking Section: Benchmarking and Standard Setting

The Speaking Test of the ALCE examination is rated by one rater who acts concurrently as interlocutor and rater and examines one candidate at a time. The Test lasts 11 – 13 minutes and comprises four tasks:

1. Warm-up; expression of personal information
2. Discussion based on visual/written prompt; using language effectively to speculate & give opinion
3. Wider discussion relating to previous topic; expressing a reasoned opinion in extended speech with effective support
4. Arguing a case convincingly and supporting this point of view effectively

All raters are given a rater code and also record which Form is used for each candidate. This allows for inter-rater and intra-rater analysis to check inter-rater and intra-rater reliability. 6 – 8 Forms are used at each administration of the ALCE examination.

Method of benchmarking & standard setting

Judges were trained in the CEF speaking scales and the ALCE scoring criteria, and then watched videos of 9 candidates being examined on materials from the June examination. After individual and collective discussion, a grade was assigned to each candidate to provide a standard for raters.

Selection of judges

The judges were selected for their knowledge and experience of teaching English at this level and their knowledge of testing procedures. Many of the judges also participated in other sections of the standard setting procedures.

Training of the Judges

The training of the judges was carried out in a similar manner to that of the Listening, GVR and Writing sections. Having examined, sorted and discussed the relevant descriptors from the CEFR, the judges were given the rating descriptors for the ALCE (Fig. 6). These were discussed and key features distinguishing the scoring criteria were emphasized.

Establishment of benchmarking scores

The judges examined videos of 9 candidates being examined on materials from the June 2006 administration. These were marked by the judges and then the marks were discussed until a consensus was reached for each candidate.

Conclusion

The benchmarked videos formed the basis for rater training for the Speaking Section of the June ALCE examination. During the benchmarking procedure, it was agreed by all judges that the descriptors for a Pass in the Speaking Section of the ALCE examination correspond to the level set out by the descriptors from the CEFR at C1 level.

Figure 6: ALCE SPEAKING SECTION: DESCRIPTORS OF SALIENT FEATURES

	Range of Language	Fluency	Accuracy	Interaction	Coherence
FOCUS	Complexity of: <ul style="list-style-type: none"> • grammar • syntax • vocabulary 	Length of utterances Hesitations Flow of speech	Accuracy of: <ul style="list-style-type: none"> • grammar • syntax • vocabulary 	Contribution Understanding	Linking of ideas
3 HIGH PASS <i>Effective Speaker</i>	Shows great flexibility in formulating ideas by using a wide range of linguistic forms effectively. Uses a broad range of syntactical structures, and displays a wide lexical range.	Expresses him/herself effectively on complex topics at length. Hesitations are rare.	Maintains a high degree of grammatical and syntactical accuracy, even in complex structures. Consistently chooses vocabulary appropriate for the topic and style. Minor errors may occur.	Contributes to the development of complex topics easily and effectively. Understands the interlocutor perfectly.	Consistently uses a wide range of organizational patterns and cohesive devices in an effective manner.
2 PASS <i>Satisfactorily Speaker</i>	Displays a good command of a satisfactory range of language to accomplish the tasks. Uses some variety of syntactical structures and displays sufficient lexical range to accomplish the tasks satisfactorily.	Expresses him/herself with ease using both short and longer utterances spontaneously. Only a conceptually challenging topic can hinder a natural, smooth flow of language. Some hesitations may occur.	Maintains a fairly high degree of grammatical and syntactical accuracy. Mostly chooses vocabulary appropriate for the topic and style. Some errors may occur in more complex structures, but do not impede communication.	Contributes to the development of complex topics satisfactorily. Understands the interlocutor, though may seek occasional clarification.	Uses a satisfactory range of organizational patterns and cohesive devices to link ideas.
1 FAIL <i>Limited Speaker</i>	Uses a limited range of language in attempting to accomplish the tasks. Uses a very limited variety of complex syntactical forms and displays a lack of lexical range.	Produces both short and longer utterances almost effortlessly on familiar topics, but encounters difficulty when dealing with complex topics. Hesitations are frequent enough to disrupt the flow of speech noticeably.	Errors may occur in both simple and complex structures, which may sometimes impede communication. Some vocabulary may be inappropriate for the topic and style.	Contributes to the development of familiar topics, but has difficulty developing complex topics. Can follow speech on familiar topics, but has difficulty following the interlocutor on complex topics.	Uses a limited range of cohesive devices to link ideas.