

English



YLIOPIILASTUTKINTOLAUTAKUNTA
STUDENTEXAMENSÄMÄNDEN

INDEX

Introduction	2
Reading Comprehension Items	7
Listening Comprehension Items	19





Introduction

There is only one national examination system in general education in Finland, the Matriculation Examination, which is taken at the end of the upper secondary school (age 18-19) by roughly half of the age group. In some parts of the country this proportion is considerably higher. The Matriculation Examination celebrated its 150th anniversary in 2002. The examination was initially administered orally at and by the University of Helsinki as its own entrance examination to a handful of applicants. When secondary education gradually spread to all parts of the country and the number of potential applicants increased considerably, it was impossible to arrange the examination in this fashion. Instead, upper secondary schools were charged to administer the written examination papers sent by the Matriculation Examination Board. The examination now takes place at all senior secondary schools (close to 500) twice a year, in the spring and in the autumn. The examination need not be taken at one sitting but parts of it can be sat at consecutive sessions. Those students who failed in one or more subjects can try to obtain a pass at a subsequent session and those who passed can also try to improve their marks. This option of distributing the examination and/or trying to get better marks is taken by a substantial majority.

The examination is constructed, organized and marked by the Matriculation Examination Board, which is appointed for a period of three years by the Ministry of Education and enjoys a very high degree of autonomy. Its 30-plus members consist of representatives from universities and from the National Board of Education. The cost of running the examination is covered entirely by candidates who have to pay a general registration fee plus a fee for each subject they take.

The purpose of the examination is to find out if the students taking it have reached sufficient 'intellectual maturity' and master the content that is included in the official curricula. Thus the examination is intended to be a measure of both general ability and of what students have learned.

Tests are constructed on the basis of the official curricula, and separate tests are administered for different curricula (representing programs of study which vary in terms of length of study and extent of content): for instance two tests of different difficulty are available in the (foreign and national) languages and in mathematics.

Test construction

The tests are constructed by regular members of the Board assisted by a group of experienced item writers. As the tests are not piloted because of test security, a considerable amount of effort is devoted to test review within the specific language groups. When this is completed the drafts are presented to the whole language section and, depending on the language, 5-15 members read and comment the drafts. This is quite thorough. Occasionally the draft is considered to have so many inadequacies that the whole package needs to be presented after thorough revision. In most cases a varying amount of editing is required.



This stimulus review benefits very much from the item analyses of previous tests which are available twice a year. Several members have a very long involvement in this process and have acquired extensive experience in estimating item difficulty. As a consequence, the best reading and listening tests with some 30 items reach reliabilities over .90 and even the poorer ones are usually close to or above .80 (without pilot testing).

There are training events arranged twice a year, usually one for test constructors and one for markers. In recent times, training has focussed largely on the implications of the Common European Framework of Reference for test construction and the possibility of relating the examination grades to the CEF levels.

There are always reserve tests in case of breach of test security. This is a caution practically never needed, as schools manage their role very well. Thus the reserve tests can be used as the next regular test.

Marking system

The students' answers are marked first by their own teachers using the official general guidelines/criteria. In the case of modern languages, examination-specific guidelines are provided by the language teachers' associations after consultation with the Board members. Teachers enter their mark on the test papers and on a separate list. All this material is sent to the Board, which in the case of English uses the services of 30-35 experienced raters to mark the papers centrally.

If the external rater differs from the teacher's mark by a certain margin, the paper is given to a second or even a third marker. The candidates can ask for remarking, in which case three new raters are brought in.

Over the past few years, about 30,000 students have sat the examination each year, including both the spring and autumn examinations.

Marker agreement is monitored at regular intervals.

Second language/foreign languages (FL) tests

In 1990 the Board started to implement a new phase in its L2/FL tests, which makes it possible to vary the test types within a set of an approved list. There are some differences between languages, but all languages use selected answer items (usually in L2), typically 5 short answers in listening and reading comprehension tests (students give brief answers in the target language or in the mother tongue), and the earlier short fill-in cloze test of grammar has been largely replaced with an integrated MC-cloze test of grammar, vocabulary and idioms (35-50 items). The latter task is generally based on a unified text, or several shorter passages but can also include separate items in a sentence context. A written composition of 150-200 words is written on a topic chosen from 3-4 rubrics. There is no oral test.

Marking and grading

Selected answer responses are read optically and the score is often weighted. Short answers to listening and reading comprehension tasks are marked using a scoring guide (0-1, 0-2, 0-3



or 0-4) with sample answers representing different score points. A more extensive scale is often used if the answers are in the target language. Also the scores for the open-ended tasks may be weighted. The essays (compositions) are rated using a rating scheme that covers 8 bands consisting of 4 analytical criteria (scale 0 – 99).

The scores from the different subtests are summed the maximum score being 299. Listening comprehension accounts for a maximum 90 points (testing time about 55 minutes). Reading comprehension and grammar/vocabulary account for a maximum of 110 points and the composition for the maximum of 99 points (6 hours allowed for the three written sub-tests).

Over the past 10-year period the average percentage correct for the advanced English listening comprehension tests was 76% (range 69.0 – 86.5; median 77%); reading comprehension 78% (range 71.2 – 86.3; median 79%).

Analyses

The selected answers are analysed using classical test theory. As the items are not pilot tested because of test security, occasionally an item may prove inadequate and be removed from the total score.

When the examination is over there are extensive analyses of the examination results in terms of various candidate background data (eg., mother tongue, type of curriculum studied).

Cut-off scores

The sum score is used for dividing the results into six passing grades and one failing grade. The grades are awarded on the basis of a modified normal distribution scale and taking into account the empirical difficulty of the test. The percentage of failed students is usually 3-5%.

Standard setting related to the submitted items

The results of standard setting for the autumn 2001 advanced English test are reported in Kaftandjieva, F. & Takala, S. (2002). Fourteen (14) experienced raters carried out a set of tasks for setting standards using the Council of Europe reference levels.

1. They sorted independently descriptors for different skills to 6 levels as follows: Listening Comprehension 20, Reading Comprehension 20, Writing 25, Grammar 18, and Vocabulary 18.
2. They rated independently all (110) the Matriculation examination test items (testing the above-mentioned skills) using the CEF 6-point skill-specific scales; the instruction told the raters *to indicate for each item at what proficiency level a person would already be able to answer the item correctly*

They also rated a sample of 30 written compositions randomly picked but covering the whole range of proficiency, with most compositions representing the middle range of the dimension



Reliability of ratings

The reliability of rating the items and the compositions was studied by using the alpha coefficient. The results indicated that the rating of the 30 compositions was the most reliable ($\alpha = .977$, 14 raters). The corresponding figures were .852 for grammar and vocabulary items (13 raters), .737 for listening comprehension (13 raters) and .723 for reading comprehension (13 raters). The mean agreement among raters varied from a fairly low correlation of .631 to a relatively high correlation of .825 (mean .768). The correlation between the raters' assessment and the actual scores was higher, and varied between .667 and .893 (mean .825). Rater 3 diverged most clearly from the rest and removing her from the data would raise the inter-rater agreement to .780.

Intra-judge consistency

The results of this study show that the correlation between the experts' rating of the items and their empirical difficulty is rather low (.26 in average), varying between .45 and .03, in other words, in this case we are facing again the familiar problem of intra-judge inconsistency.

The fact that the reliability of the ratings was comparatively high (between .723 for listening and .852 for grammar) does not contradict the above mentioned conclusion. The high reliability coefficient in this case simply means that there is internal consistency between raters in their relatively poor ability to assess the empirical difficulty of the items and to categorize them in a consecutive order.

Standard Setting

General description of the method

The standard setting method applied in this case study can be classified as a test-centered continuum method and can be regarded as a modification and an extension of the classical Angoff yes/no method.

After the aggregation of the individual ratings of the items for every skill, the number of items belonging to a certain level can be detected. Then, one possible way of establishing the cut-off scores is to follow the cumulative frequency distribution of the items. An illustrative example of the described procedure is presented in Table 1.

Table 1: Cut-off score establishment – an illustrative example

CEF levels	Number of items per level	Cumulative Frequency	Cut-off scores
A1	3	3	≤ 3
A2	4	7	[4 – 7]
B1	17	24	[8 – 24]
B2	15	39	[25 – 39]
C1	7	46	[40 – 46]
C2	4	50	[47 – 50]



Aggregation procedure

Different aggregation procedures were studied in order to choose the one which fits best the empirical difficulty of the items. As a result, the rounded average of the experts' ratings was chosen. Its correlation with the empirical difficulty of the items is 0.39, which obviously, despite the fact that it was the highest possible, is quite low and therefore unsatisfactory.

The positive result was that there was a significant difference between the average empirical difficulty of the items allocated to the three levels concerned, which was not the case with the other aggregation procedures tried out.

To overcome the problem with the low correspondence between item difficulty and the assigned levels of language proficiency and to assure more adequate representation of the proficiency levels in the test, a new aggregated variable was constructed, based simply on the recoding of item difficulty index into 5 categories (A2, B1, B2, C1, and C2) with cut-off points corresponding to the ends of the confidence intervals of the means.

As a result, the levels assigned to the items corresponded to the empirical difficulty of the items, and the average difficulty of the items per level was within the confidence interval of the average difficulty of the items that belong to the same level according to the rounded mean aggregation procedure.

The results of this standard setting study are used as one basis in assigning CEF-levels to the presented items.

Reference

Kaftandjieva, F. & Takala, S. Relating the Finnish Matriculation Examination English Test Results to the CEF Scales. Helsinki Seminar, June 30- July 2, 2002

(available on request from Sauli Takala: sjtakala@hotmail.com)



Reading Comprehension Items

The empirical item statistics are based on about 14,000 matriculation examinations takers in 2001. All the items were rated in terms of their CEF levels by 14 very experienced and CEF-familiarized exam markers. Some of the items were also rated by a team of five experienced experts representing five different European countries.

Read the text and then answer the questions. Choose the best alternative for each item.

News in brief

Task 1

Laurie Garret, a medical reporter for Newsday in New York, says she worries so much about how readers will react to her stories, especially those about cancer treatments, that “I try to deliberately write in very neutral tones. I tend to avoid adjectives altogether. I almost dry it out, make my writing bland so that whatever emotional baggage the reader brings to the story in the first place is not going to be exaggerated by the tone with which I deliver the information.”

Item

How could Laurie’s writing be described?
 A She writes using a matter-of-fact approach
 B She stresses the positive outcome of diseases
 C She uses precise language at all costs

Content analysis

Text Characteristics:

- Text source: newspaper
- Authenticity: authentic
- Discourse type: narrative/reports
- Domain: occupational/public
- Topic: other/the world of work
- Nature of content: fairly extensive abstract content
- Text length 83 words
- Vocabulary: rather extended vocabulary
- Grammar: limited range of complex structures
- Text likely to be comprehensible by learner at CEF level: B2



Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: make explicit inferences about main idea/gist
- Item level estimated: C1
- Key: A
- Simple: 18-19-year old; senior secondary school; 10 years of English
- Per cent correct: 60.8 %
- Discrimination (item/total correlation): .43

Task level estimated: C1

Comment: The international experts slightly underestimated the level (difficulty) of the item as well as the task level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.



Task 2

Being disgraced is no longer what it was. It used to mean being cast into outer darkness, spending the rest of your life in virtual exile skulking in the shadows. The opposite is true today: disgrace guarantees you at least a spell in the limelight, massive and well-paid publicity and, for anyone with a gift for talking and writing, a successful media career.

Item

What is the paradox presented in the text?
 A The media is both attractive and upsetting
 B One becomes famous, but has to hide
 C Scandalous behaviour could be profitable

Content analysis

Text Characteristics:

- Text source: magazine
- Authenticity: authentic
- Discourse type: narrative
- Domain: public
- Topic: contemporary life
- Nature of content: fairly extensive abstract content
- Text length 63 words
- Vocabulary: extended vocabulary
- Grammar: limited range of complex structures
- Text likely to be comprehensible by learner at CEF level: B2/C1

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: make explicit inferences about main idea/gist
- Item level estimated: B2
- Key: C
- Sample: 18-19-year old; senior secondary school; 10 years of English
- Per cent correct: 70,5 %
- Discrimination (item/total correlation): .56

Task level estimated: B2

Comment: there was a very good agreement about the item level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.



Task 3

Charges that Britain helps America and others spy on its European allies have annoyed some across the Channel. So much for Britain's commitment to European solidarity. Its real union is claimed to be with America.

Item

What does the text imply about Britain?

- A It is committed to Europe
- B It is insulted by the spy rumours
- C It is disloyal to its neighbours

Content analysis

Text Characteristics:

- Text source: magazine
- Authenticity: authentic
- Discourse type: mainly expository
- Domain: public
- Topic: other/politics
- Nature of content: fairly extensive abstract content
- Text length 35 words
- Vocabulary: rather extended
- Grammar: wide range of complex structures
- Text likely to be comprehensible by learner at CEF level: B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit main idea/gist
- Item level estimated: B2
- Key: C
- Sample: 18-19-year old; senior secondary school; 10 years of English
- Per cent correct: 74,1%
- Discrimination (item/total correlation): .39

Task level estimated: B2

Comment: there was a good agreement about the item level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B1: Can identify the main conclusions in clearly signalled argumentative texts. Can recognize the line of argument in the treatment of the issue presented, though not necessarily in detail.

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.



Task 4 : (Items 1-4)

Political portals – just a new fad?

A new wave of for-profit “politics portals” -- Voter.com, Grassroots.com, SpeakOut.com, and others -- aim to do for civic life what Senator McCain’s ultimately unsuccessful primary campaign did for the recent American presidential race: shake things up a bit.

All of these sites are, for the most part, variations on a theme. Register and plug in information like your ZIP code or favorite issues, and they will funnel you user-specific news (mostly international news agency feeds and fluffy “exclusives”), candidate bios (= biographies) and legislative records, issue briefings from various interest and pressure groups, and even streaming video clips.

So far, though, Voter.com, Grassroots.com, and the rest, ultimately look like nothing so much as a pretty typical bunch of Internet start-ups. To them, the marketplace of ideas is just another marketplace, and political discourse is merely another form of commerce -- the buying and selling of candidates, ideas, and issues. And just as the Internet can make conventional commercial activities (for example, shopping for books) quicker and more convenient, so do sites like Grassroots.com and Voter.com seek to make politics quicker and more convenient. The new Internet companies argue that the decisions haven’t gotten any easier, participating in them has.

But is ease of participation the real problem? Arguably, Americans’ low voting activity is less a matter of convenience or inconvenience (how easy or difficult it is to vote) than of substance (who and what you get to vote for). That means the for-profit political sites have it backwards. In the end, they are not e-commerce businesses, but content businesses -- and some kinds of content just don’t sell. One-click shopping will never help Amazon.com sell more copies of ancient classics like Pliny the Elder’s *Natural History* (current sales rank: 66,467), no matter how convenient they make it.

In politics it is the content that matters in the end. Candidates are the contents. And, in spite of the novelty of a John McCain, if there’s one thing American voters have pretty consistently agreed upon during the past few decades, it’s that the content of American politics pretty much sucks.

Item 1

What do the web sites try to do?

- A Hook people by first offering cheap service
- B Attract people by catering to their interests
- C Appeal to people by free special offers

Content analysis

Text Characteristics:

- Text source: magazine
- Authenticity: authentic
- Discourse type: mainly argumentative
- Domain: public
- Topic: other/politics
- Nature of content: fairly extensive abstract content
- Text length 350 words
- Vocabulary: extended vocabulary



- Grammar: wide range of complex structures
- Text likely to be comprehensible by learner at CEF level: C1

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit main idea/gist
- Item level estimated: A2/B1
- Key: B
- Per cent correct: 90,5%
- Discrimination (item/total correlation): .31

Task level (including all the items) estimated: B2

Comment: Both the national and international experts overestimated the item level quite clearly.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

Item 2

Which of the following statements is true?

- A The sites speed up the turnover of political candidates
- B The present sites leave the author cold
- C The sites facilitate people's political decision-making

Content analysis

Item Characteristics:

- Operations: Make explicit inferences about the writer's mood
- Item level estimated: C2
- Key: B
- Per cent correct: 36,1 %
- Discrimination (item/total correlation): .45

Comment: Both the national and international experts underestimated slightly the item level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered



in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

C2: As in C1

Item 3

- What is the argument presented about political discussion?
- A It is pretty much the same as any other form of talk
 - B It makes people's views accessible to a broader audience
 - C It cannot be properly conducted via a computer

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit main idea/gist
- Item level estimated: C2
- Key: B
- Per cent correct: 43,3 %
- Discrimination (item/total correlation): .20

Comment: Both the national and international experts underestimated the item level by about one level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

C2: As in C2

Item 4

- Why is Pliny's Natural History referred to?
- A To show how easy it is to buy a classic on the Internet
 - B To applaud Amazon.com for its efforts to educate us
 - C To point out that it's the content that counts

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit detail
- Item level estimated: B1
- Key: C
- Per cent correct: 83,6 %
- Discrimination (item/total correlation): .43



Comment: Both the national and international experts overestimated the item level by about one level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B1: Can identify the main conclusions in clearly signalled argumentative texts. Can recognize the line of argument in the treatment of the issue presented, though not necessarily in detail.

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.



Task 5 : (Items 1-4)

Bobos

With *Bobos in Paradise*, David Brooks has introduced a new word “Bobo” -- into the lexicon, where it seems destined to take its place alongside such classic monikers as “yuppie,” “hippie,” and “WASP.” The book argues that the bohemian spirit of the sixties has merged with the acquisitive impulses of the eighties to yield, in the nineties, a hybrid bourgeois-bohemian (or “Bobo”) spirit of the age.

In the fifties and sixties, Brooks explains, the Protestant establishment fell victim to the new meritocratic beliefs which, thanks to the educational-testing movement, began conferring status according to educational achievement rather than inherited wealth and breeding. Although many of the bright Baby Boomers who gained access to top schools initially scorned wealth, these well-educated idealists often became rich in spite of themselves, as the information economy began lavishly rewarding their knowledge and education. In the nineties, these newly well-off antimaterialists found ways to reconcile their unexpected wealth with their high-minded ideals. The result was a new upper class whose “grand achievement” has been the creation of “a way of living that lets you be an affluent success and at the same time a free-spirit rebel.”

Bobos lavish their money not on luxuries, but on necessities like kitchens and bathrooms, splurging on marble shower stalls, and stainless steel refrigerators. The most popular Bobo leisure-time pursuits are strenuous or edifying (like hiking or ecotourism) rather than hedonistic (= keen on pleasure). And Bobo job offers claim to hold out potential for personal growth and self-discovery. Brooks dissects this new Bobo lifestyle with perceptive humor: “To calculate a person’s status, you take his net worth and multiply it by his antimaterialistic attitudes. A zero in either column means no prestige, but high numbers in both rocket you to the top of the heap. Thus, to be treated well in this world, not only do you have to show some income results; you have to perform a series of manoeuvres to show how little your worldly success means to you.... You will devote your conversational time to mocking your own success in a manner that simultaneously displays your accomplishments and your ironic distance from them. You will ceaselessly bash yuppies in order to show that you yourself have not become one. You will talk about your nanny as if she were your close personal friend, as if it were just a weird triviality that you happen to live in a \$900,000 Santa Monica house and she takes the bus two hours each day to the barrio.”

Item 1

How did Boboism emerge?

- A Testing became more reliable
- B Ability was rewarded
- C The Establishment found a new victim

Content analysis

Text Characteristics:

- Text source: magazine
- Authenticity: authentic
- Discourse type: mainly expository
- Domain: public/personal
- Topic: other/contemporary life



- Nature of content: mainly abstract content
- Text length 418 words
- Vocabulary: extended vocabulary
- Grammar: mainly complex structures
- Text likely to be comprehensible by learner at CEF level: C1

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit detail
- Item level estimated: B2
- Key: B
- Per cent correct: 65,9 %
- Discrimination (item/total correlation): .52

Task level estimated (all items) : C1

Comment: The national experts overestimated the item level by about one level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

Item 2

How can Bobos be characterized?

- A As antimaterialist losers
- B As rebellious and unwordly idealists
- C As reluctant winners

Content analysis

- Item type: Selected response/multiple choice
- Operations: Evaluate implicit opinion
- Item level estimated: C2
- Key: C
- Per cent correct: 49,2 %
- Discrimination (item/total correlation): .32

Comment: Both the national and international experts underestimated the item level by about one level.



Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.

Item 3

What does the quote from Brooks' book suggest?

- A Bobos are exceptionally skilled in handling math operations
- B Bobos hide their contempt of ordinary people in skilful irony
- C Bobos' lifestyles are elaborate social games

Content analysis

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit main idea/gist
- Item level estimated: C1
- Key: C
- Sample: 18-19-year old; senior secondary school; 10 years of English
- Per cent correct: 52,1 %
- Discrimination (item/total correlation): .36

Comment: Both the national and international agreed well with the empirical item difficulty level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.



Item 4

What do the word choices by Brooks and the author of the article imply about their attitude towards Boboism?

- A Irony mixed with some sympathetic humour
- B Sarcasm mixed with strong admiration
- C Disapproval mixed with badly concealed envy

Content analysis

- Item type: Selected response/multiple choice
- Operations: Evaluate implicit attitude/conclusion
- Item level estimated: B2
- Key: A
- Per cent correct: 74,4%
- Discrimination (item/total correlation): .24

Comment: Both the national and international experts overestimated the item level by about one level.

Rationale in terms of the Common European Framework

Scale: Reading for information and argument (CEFR, p. 70)

B2: Can understand articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints.

C1: Can understand in detail a wide range of lengthy complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied as well as stated opinions.



Listening Comprehension Items

The empirical item statistics are based on about 14,000 matriculation examinations takers in 2001. All the items were rated in terms of their CEF levels by 14 very experienced and CEF-familiarized exam markers. Some of the items were also rated by a team of five experienced experts representing five different European countries.

Listen to the text and answer the questions.

Task 1

Item 1.1

Why are Cathy and Richard discussing Keiko Kimura's book? Because of

- A reactions to previous programs.
- B requests from their audiences.
- C orders from their producer.

Content analysis

Text Characteristics:

- Text source: debates and discussions
- Authenticity: scripted
- Discourse type: mainly expository
- Domain: public/occupational
- Topic: other: world of work
- Nature of content: fairly extensive abstract content
- Text length: 1:56 minutes
- Vocabulary: rather extended
- Grammar: limited range of complex structures
- Text speed: fast
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: twice
- Text likely to be comprehensible by learner at CEF level: B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated (based partly on 13 raters's judgement): A2



- Key: B
- Per cent correct: 91,7%
- Discrimination (item/total correlation): .45

Task level estimated: B1/ B2

Comment: Both the national and international experts overestimated the item level by one level.

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

A2: Can understand and extract the essential information from short, recorded passages dealing with predictable everyday matters which are delivered slowly and clearly.

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 1.2

What is remarkable about the bi-annual gifts?

- A Lack of marketing
- B Parcel delivery
- C No cost-consciousness

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B1
- Key: B
- Per cent correct: 75,5 %
- Discrimination (item/total correlation): .49

Comment: Both the national and international experts overestimated the item level by about half a level.

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.



Item 1.3

- What is said about the buying of gifts?
- A Careful budgets are made
 - B A small percentage of salaries is spent
 - C Enormous amounts of money are used

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B1
- Key: C
- Per cent correct: 80,4 %
- Discrimination (item/total correlation): .35

Comment: Both the national and international experts agreed well with the item difficulty level.

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 1.4

- What does Richard suggest about the giving of gifts at work?
- A Giving gifts requires good taste
 - B One's readiness to sacrifice is valued
 - C Money is all-important in business

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit main idea
- Item level estimated: B2
- Key: B
- Per cent correct: 52,1 %
- Discrimination (item/total correlation): .21

Comment: Both the national and international experts agreed quite well with the item difficulty level.



Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 1.5

What does Cathy say about giving gifts?

- A It's close to bribery
- B It's regarded as natural
- C It's difficult to know who to give them to

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B1
- Key: B
- Per cent correct: 84,9 %
- Discrimination (item/total correlation): .31

Comment: The national experts overestimated the item level by about one level.

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.



Item 1.6

What possible problem do young American teachers face in Japan?
 A Their students expect gifts
 B Their seniors expect gifts
 C Their host families expect gifts

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B1
- Key: B
- Per cent correct: 76,9 %
- Discrimination (item/total correlation): .54

Comment: Both the national and international experts overestimated the item level by about a half level.

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.



Listen to the text and answer the questions.

Task 2

Item 2.1

What was said about Jasmine at the award ceremony?

- A She shows dedication and courage
- B She is hardworking and idealistic
- C She champions the cause of women lawyers

Content analysis

Text Characteristics:

- Text source: TV/radio documentaries
- Authenticity: authentic
- Discourse type: mainly expository
- Domain: personal/occupational
- Topic: other: world of work
- Nature of content: fairly extensive abstract content
- Text length: 2:17 minutes
- Vocabulary: rather extended
- Grammar: limited range of complex structures
- Text speed: normal
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: twice
- Text likely to be comprehensible by learner at CEF level: B2/C1

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit main idea
- Item level estimated: B2
- Key: A
- Per cent correct: 74,8 %
- Discrimination (item/total correlation): .53

Task level estimated: B2



Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

C1: Can understand a wide range of recorded and broadcast audio material, including some non-standard usage, and identify finer points of detail including attitudes and relationships between speakers.

Item 2.2

What does pro bono mean?

- A The best possible legal help is given
- B Foreigners receive special attention
- C Clients pay nothing

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B1
- Key: C
- Per cent correct: 88,5 %
- Discrimination (item/total correlation): .56

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

C1: Can understand a wide range of recorded and broadcast audio material, including some non-standard usage, and identify finer points of detail including attitudes and relationships between speakers.



Item 2.3

What is this discussion?
 A A job interview
 B An interview for the court
 C An interview on the radio

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit main idea/gist
- Item level estimated: B1
- Key: C
- Per cent correct: 85,9%
- Discrimination (item/total correlation): .53

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 2.4

What is Jasmine's reaction to the public acknowledgement?
 A She felt ashamed
 B She felt indifferent
 C She felt honoured

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit main idea/gist
- Item level estimated: B2
- Key: C
- Per cent correct: 71,1%
- Discrimination (item/total correlation): .46



Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 2.5

What does Jasmine say about legal advice?

- A It should be available to all
- B It should be paid by the state
- C It should stress human rights above all

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Recognise and retrieve explicit detail
- Item level estimated: A2/B1
- Key: A
- Per cent correct: 90,2%
- Discrimination (item/total correlation): .39

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

Item 2.6

Did Jasmine always wish to specialise in human rights?

- A Not really, it just more or less happened like that
- B Yes, but she realised that there are other important considerations
- C Yes, however, she understood that the economy still deserved to be the first priority

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about explicit detail
- Item level estimated: B2
- Key: B
- Per cent correct: 51,9%
- Discrimination (item/total correlation): .19

Task level estimated: B2



Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Item 2.7

What does the interviewer suggest about the lawyers?

- A Lawyers are usually seen to be keen on making money
- B Lawyers base their thinking on concrete anecdotal cases
- C Lawyers struggle to strike a balance between conflicting values

Content analysis

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Infer implicit detail
- Item level estimated: B1
- Key: A
- Per cent correct: 77,7%
- Discrimination (item/total correlation): .55

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.



Task 3

Item

What does Joan do?
 A She forecasts the weather
 B She repeats what she heard
 C She argues with Phil

Content analysis

Text Characteristics:

- Text source: interpersonal dialogues and conversations
- Authenticity: scripted
- Discourse type: mainly descriptive
- Domain: personal
- Topic: relations with other people
- Nature of content: mostly concrete content
- Text length: 19 seconds
- Vocabulary: mostly frequent words
- Grammar: mainly simple structures
- Text speed: normal
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: once
- Text likely to be comprehensible by learner at CEF level: B1/B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit opinion
- Item level estimated: B1
- Key: B
- Per cent correct: 64,5%
- Discrimination (item/total correlation): .31

Task level estimated: B1/ B2

Comment: Both the national and international experts overestimated the item level by about half a level.

Rationale in terms of the Common European Framework

Scale: Understanding conversation between native speakers (CEFR, p. 66)

B1: Can generally follow the main points of extended discussion around him/her provided speech is clearly articulated in standard dialect.

B2: Can keep up with an animated discussion between native speakers.



Task 4

Item

What does the woman talk about?
 A Mental illness
 B Fraud
 C Priceless art

Content analysis

Text Characteristics:

- Text source: Interpersonal dialogues and conversations
- Authenticity: pedagogic
- Discourse type: mainly narrative
- Domain: personal/public
- Topic: relations with other people
- Nature of content: mainly abstract content
- Text length: 20 seconds
- Vocabulary: mostly frequent words
- Grammar: mainly simple structures
- Text speed: normal
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: once
- Text likely to be comprehensible by learner at CEF level: B1/B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit opinion
- Item level estimated: C1
- Key: B
- Per cent correct: 30,3%
- Discrimination (item/total correlation): .37

Task level estimated: B2

Comment: The national experts underestimated the item level by about one level (the international experts...) half a level.

Rationale in terms of the Common European Framework

Scale: Understanding conversation between native speakers (CEFR, p. 66)

B2: Can keep up with an animated discussion between native speakers.

C1: Can easily follow complex interactions between third parties in group discussion and debate, even on abstract, complex unfamiliar topics.



Task 5

Item

The woman says that teachers are like
 A policemen
 B prisoners
 C jail keepers

Content analysis

Text Characteristics:

- Text source: Interpersonal dialogues and conversations
- Authenticity: pedagogic
- Discourse type: mainly argumentative
- Domain: personal/public
- Topic: other: contemporary life
- Nature of content: mainly abstract content
- Text length: 32 seconds
- Vocabulary: mostly frequent words
- Grammar: mainly simple structures
- Text speed: normal
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: once
- Text likely to be comprehensible by learner at CEF level: B1/B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit main idea/gist
- Item level estimated: B1
- Key: B
- Per cent correct: 78,9%
- Discrimination (item/total correlation): .50

Task level estimated: B1/B2

Comment: The national experts agreed quite well with the empirical item difficulty level.

Rationale in terms of the Common European Framework

Scale: Understanding conversation between native speakers (CEFR, p. 66)

B1: Can generally follow the main points of extended discussion around him/her provided speech is clearly articulated in standard dialect.

B2: Can keep up with an animated discussion between native speakers.



Task 6

Item

What would the man say next?

A Me too! Let's do something about it.

B You don't say! Perhaps we should have a photograph made.

C That's interesting! Has this happened to you before?

Content analysis

Text Characteristics:

- Text source: Interpersonal dialogues and conversations
- Authenticity: pedagogic
- Discourse type: mainly narrative
- Domain: personal
- Topic: relations with other people
- Nature of content: mostly concrete content
- Text length: 15 seconds
- Vocabulary: mostly frequent words
- Grammar: mainly simple structures
- Text speed: normal
- Number of participants: two
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: once
- Text likely to be comprehensible by learner at CEF level: B1/B2

Item Characteristics:

- Item type: Selected response/multiple choice
- Operations: Make inferences about implicit detail
- Item level estimated: B2
- Key: A
- Per cent correct: 71,7%
- Discrimination (item/total correlation): .44

Task level estimated: B2

Comment: The national experts agreed well with the empirical item difficulty level.

Rationale in terms of the Common European Framework

Scale: Understanding conversation between native speakers (CEFR, p. 66)

B1: Can generally follow the main points of extended discussion around him/her provided speech is clearly articulated in standard dialect.

B2: Can keep up with an animated discussion between native speakers.



Task 7

Question A

Listen and answer the questions in English:

Why does the speaker predict Mr Strew will win?

Content analysis

Text Characteristics:

- Text source: news broadcasts
- Authenticity: genuine
- Discourse type: mainly narrative
- Domain: public
- Topic: other: politics
- Nature of content: fairly abstract
- Text length: 1:05 minutes
- Vocabulary: rather extended vocabulary
- Grammar: limited range of complex structures
- Text speed: normal
- Number of participants: one
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: twice
- Text likely to be comprehensible by learner at CEF level: B1/B2

Item Characteristics:

- Item type: Short answer question (sentence)
- Operations: recognize explicit detail
- Item level estimated: B1

Task level estimated: B2

Rationale in terms of the Common European Framework

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Scoring key 0-2. To obtain 2 points the answer has to mention either that (a) the opponent has performed badly, or that (b) Strew's campaign has been better.



Task 8

Question B

What did the boy do when he was caught? Mention two things.

Content analysis

Text Characteristics:

- Text source: news broadcasts
- Authenticity: genuine
- Discourse type: mainly narrative
- Domain: public
- Topic: other: crime
- Nature of content: mostly concrete content
- Text length: 38 seconds
- Vocabulary: rather extended vocabulary
- Grammar: limited range of complex structures
- Text speed: normal
- Number of participants: one
- Accent/standard: standard pronunciation
- Clarity of articulation: normally articulated
- How often played: twice
- Text likely to be comprehensible by learner at CEF level: B2

Item Characteristics:

- Item type: Report in own words
- Operations: Recognise and retrieve explicit detail
- Item level estimated: B2

Task level estimated: B2

Rationale in terms of the Common European Framework

Rationale in terms of the Common European Frame of Reference:

Scale: Listening to audio media and recordings (CEFR, p. 68)

B1: Can understand the information content of the majority of recorded or broadcast audio material on topics of personal interest delivered in clear standard speech.

B2: Can understand most radio documentaries and most other recorded or broadcast audio material delivered in standard dialect and can identify the speaker's mood, tone, etc.

Scoring 0-2. For full marks, the answer has to mention two things from the following three points: (a) the boy tried to pull a gun/shoot, (b) the boy threw the drug packets under a parked lorry, (c) the boy explained that the money was a birthday present. For one point, one of these points has to be mentioned, or a general mention about the boy resisted the police. The answer has to refer to concrete action, not to a plan or intention. The tense has to be past tense.