Assessing Speaking at C1 Level Business English at MU

Jiřina Hrbáčková, Milan Boháček

Language Centre, Masaryk University Brno

Abstract

This paper deals with the standardisation process in specifying a new format of the Business English spoken test at the Faculty of Economics and Administration of Masaryk University, which was previously based on the assessment of integrated skills of reading and speaking. With a growing awareness of the CEFR levels, a need to enhance fairness, reliability and validity of testing arose and was embraced by the English language department team initiating a series of discussions resulting in an overhaul of the oral test format. The major change lies in a radical reduction of instruction input, a switch from a teacher-student interview to a peer-to-peer discussion pattern, thereby splitting the role of an interlocutor and rater and the use of analytical rating scales. A year from the implementation of this system, another round of discussions was held aiming to adjust the new format in order to enhance a more autonomous and higher-standard of language production. Recordings of real live tests have been collected for benchmarking purposes and to evaluate the result of the whole process. All signs indicate that the change was for the better with the new format being more capable of eliciting desired language and teachers' ratings more reliably and consistently.

Key words: oral test, assessment, reliability, validity, paired format, analytical rating scales

1. Introduction

The need for consistent, reliable and valid assessment of the language skills of business English students at the Faculty of Economics and Administration of Masaryk University resulted in a thorough revision of the previous practices in assessing oral performance. With the increasing awareness of the standard process of test development among teachers and test developers at the Faculty of Economics and Administration, the need for an in-depth revision of the long ago specified oral part of the examination arose. The immediate urgency to ensure compliance with the reliability and validity principles of measuring spoken language proficiency had become the topic of discussions over the future form of the tests.

With a series of theoretical seminars provided under the auspices of the Impact project, the staff came to understand that to rate fairly and consistently, a complete revamp of the oral test was essential. Up to then the consistency in testing speaking had solely lain in creating test assignments based on the curriculum of Business English courses which were used by all members of staff. The general consensus was in testing students on topics mirroring the curricular subjects.

The oral exam makes up one part of the overall final English examination, of which the then written part testing listening, grammar, vocabulary and translation was a major contributor to the final grade. Originally the standardisation efforts should have mainly concerned the written part, but as the whole process turned out to be rather radical, by the time the written part was newly specified, the then oral part would not have stood the challenge of being an equally integral part of the new overall test. The issue of fairness suddenly daunted on teachers and they wholly embraced the idea. We had what Douglas (2010) describes as a healthy scepticism about our measurement of language ability to be entirely trusted and we also needed confidence that our measurements were reasonably accurate.

2. Previous format of testing spoken performance

The oral performance test tasks were based on the comprehension of an extensive text that students were required to read and an ensuing discussion mainly drawing on students' understanding of that text. For many years we pursued an integrated format, and due to time constraints, the written part of the exam was void of reading comprehension and the oral exam compensated for that to a certain extent. This entailed the sensitive complication that students who failed to comprehend the text were also unable to answer questions related to it. This conditional correlation led to some students scoring low at the oral part, which, however, was not always due to their inability of spoken expression as much as to incomprehension of the text.

Furthermore, the format followed a teacher – student interview pattern, where insufficient space was allowed for students to demonstrate ability to produce their own language and react spontaneously. The relationship in the interview setting could not be one of parity and the production thus lacked a somewhat natural aspect of a real-life discussion. Another drawback lay in the simple fact that the teacher-interlocutor had to cope with the double role of an interviewer and a rater, which made it difficult to devote all one's attention to rating the performance.

Another major flaw the former version of the oral exam suffered was little focus on consistency of rating as the methodology of testing had been left at the discretion of individual teachers. It can be said that the rating was holistic but it is fair to add that the rating methodology had never been discussed or decided upon as being holistic. The status-quo originated in the generally accepted notion that being a good teacher also meant being a good rater and interlocutor at the same time. Rarely had the traditional attitude to testing been questioned by either teachers or students. However, with the increasing awareness of the CEFR levels among both teachers and students, the demand for consistent and reliable testing intensified and traditional testing methods no longer seemed sustainable.

The last point to complement the overall picture is the C1 CEFR level of the final English language examination that students at the Faculty of Economics and Administration take after completion of a four-semester Business English course. The level itself is an issue that has been generating a lot of debate regarding the scope of testing at the faculty, where a set of at least 8 extensive tests is necessary to administer for each term to the number of students finishing the compulsory language tuition. It is not financially nor humanly feasible to conduct a standard setting process to verify the level. Therefore, despite the success rate of the test being relatively high (90% in 2013, 86% in 2012 and 75% in 2011), it would be hypocritical to claim that all students passing the examination actually comply with the requirements of the C1 level.

3. Need for change

As the staff grew aware of the can-do statements of the CEFR levels and simultaneously learnt more about the standard test development procedure, they came to understand that to ensure consistent and reliable testing of its students, a complete revamp of the whole test was inevitable. There was little inter-rater, let alone intra-rater reliability. Students were rated and interlocuted by the teacher whose lessons they attended during the two years of study and therefore there was little doubt that the performance had invariably been assessed impartially. The situation was favourable for students and placed relatively low demands on teachers.

The staff of the language department gradually initiated a series of discussions and negotiations about the future form of the oral examination. By 2012 there was a unanimous consensus about the necessity of a major overhaul, the principal requirement being the separation of reading comprehension from testing oral competence. A survey into available formats of international examinations was made. Following suit of the good practice of BEC higher and other Cambridge English exams, the staff agreed that the oral examination should consist of two parts, a monologue and a discussion between two students.

The idea of testing the two aforementioned formats required that students be tested in pairs to eliminate the influence of a teacher as a discussion partner and to allow students to manifest their grasp of the language. In testing, according to Brown and Ducasse (2009), peer or group tasks involving peer-to-peer interactions have increasingly been used since the late 1980s. This justifies our observation about the role of a teacher as a partner in discussion and a rater at the same time affects students' performance in comparison with peer-to-peer interaction. The expectations were also for eliminating the influence of the teachers' double role, where in the teacher-student interview format the role of a rater is suppressed in order to engage in a discussion rather than focus on assessing the performance, which is not the case in peer-to-peer interaction. To pre-test the newly defined format of the oral exam, the team decided to run a period of trial examinations where not only students but also teachers would be paired to split the roles of a rater and interlocutor and allow for more consistent and reliable testing and eliminate the initial concern about one teacher not being able to rate two students at the same time.

Peer-to-peer interaction seems to better reflect the interactive curricular activities that all staff consider to be an integral part of tuition. The decision to assess spoken competence as described above was more than well-received and it raised anticipations about the outcome, i.e. about the way students would tackle the tasks and about results they would be achieving. According to Brooks (2009), in the paired format, test-takers perform better and demonstrate their facility in negotiating meaning and communicating with another language learner, co-constructing better, richer performances through their interaction. This is an aspect which would be a desired outcome of our efforts and which will have to be explored further after finalising the format. It should be mentioned that teachers were curious and eager to explore the new unfamiliar method of assessment and eager to see the outcome.

4. Specification of the new format

The next step in the process consisted in specifying the test tasks. This was done by first defining the construct and selecting relevant topics covered in the seminars over the four semesters. The overall examination is a pro-achievement type of test and so is the oral part, testing both the ability of a student to use the language naturally while engaging in a discussion and to use specific vocabulary pertinent to the 17 topics selected. Task instructions were written, discussed in a team, reviewed, rechecked and drawn to their final form for the 2013 spring examination period. Further steps involved designing the two test tasks for each of the 17 topics, an example of which is shown below.

Table 1 shows testing assignment for unit 12, Securities, consisting of two optional questions in the first, monologue part of the exam, and a role-play specifically designed for each of the two participants in the discussion – student A and student B. As obvious from the table, students were provided with background information to the situation and several points they were expected to present and discuss. Roles often consisted of persuading each other or one of the partners in the discussion, which was soon found to be limiting the choice of communicative language functions that students could use.

Table 1: The new format assignment of the oral part of the examination for question 12, Securities

12. Securities

Monologue

1. When you have money to invest, what kind of securities will you invest it in? Why?
2. What is equity financing and debt financing and why would a company prefer one over the other?

Role-play

Student A

You are an Investment portfolio manager and you manage investment portfolios for rich clients. You meet with one of them to present a new investment opportunity: you want your client to invest in Berganian government bonds. Bergania is a country which has problems with rebels, violence and civil rights. The bond has a 7-year maturity and offers a 12% coupon. The country has huge mineral reserves and has potential for economic growth if stability is achieved. Persuade your client to invest, with regard to the following facts:

- You know that the government wants to use the capital raised to buy weapons to fight against the rebels.
- You secretly made a deal with the government which ensures you get a share of the capital raised.
- Your client considers you a friend and trusts you.

You start the conversation.

Student B

You are a wealthy investor. Your investment portfolio manager has arranged a meeting with you because she/he wants to inform you about a new investment opportunity. You welcome a new investment opportunity because you decided to create a special fund for future education of your children. Listen to your portfolio manager's proposal and decide whether to invest, with regard to the following facts:

- You are always interested in where your investment is going.
- You are a pacifist.
- Your investment portfolio manager is your friend.

Your partner will start the conversation.

In the first part, students choose one of the questions and have one minute to prepare. This is followed by a two-minute autonomous language performance. The teachers do not interrupt or ask additional questions so that they can also rate the aspect of independence and fluency of a student's speech. On completing the first part of the oral test, one student in the pair is invited to choose the topic for the dialogue. After a two-minute preparation, a discussion covering the roles follows. Students are then asked to wait until the rater discusses the performance with the colleague, the interlocutor, and presents the mark to the students justifying the decision if necessary.

Originally the team intended to pre-test the newly specified test on a sample group of target student population before using it in real testing situations, but the discontent and reluctance to use the old versions resulted in a unanimous consensus to introduce the new format straight away. Partly it was due to the examination period being by all means a hectic part of the academic year and the old version oral tests requiring that teachers repeatedly went through the texts underlying the oral test tasks, which was a time-consuming and unnerving obligation, and partly it was because the new format seemed promising in making students work harder and depriving teachers of the double role of interviewer and rater.

5. Implementation process

The very first rounds of oral examinations confirmed that the format suited our purposes much better than the previous one. Little vocabulary input was provided to students that would enable them to either misunderstand or recycle the language and so their performance was indeed a much better indicator of their communicative ability. The split role of teachers as interlocutors and raters proved to be another substantial asset and although initially planned only to set off the radically new format, the faculty acknowledged the benefit of increased reliability and mutual support of colleagues in assessing the performance and anchored the presence of two testers at the exam in the specifications. The predicted time economy, in which two students complete their tasks in approximately 20 minutes, is counterbalanced by the double-teacher format.

6. Revision of the new format

By the end of the spring 2013 exam period, the teachers had already acknowledged that the new format, however enthusiastic they were about it, needed revision to satisfy the requirements for eliciting a truly independent oral performance. Two major objections were raised as to the validity of the test tasks with the former being a limited range of language that some of the tasks forced students to produce and the latter calling for an even more independent production. It was pointed out that mainly weaker students used the input instructions thoroughly to their advantage repeating the language of the assignment. This was by no means something to punish students for as they were expected to discuss the topic based on the roles designed. For the topic of Securities (table 1), several key expressions of the assignment could be used and so rating range and accuracy of vocabulary (by the newly developed rating scales, as will be discussed further) did not lead to the expected extent of autonomous production.

Another point in discussion was the imbalance between the two roles. In some cases the role of one student was superior to the other. This, as we learnt by experience, might have disadvantaged those students who drew the role of a subordinate in the corporate hierarchy. Similarly, if a role entailed the need to persuade the other student about a certain point or a solution, we agreed that the role disadvantaged the second student who had to focus powers on defending their point of view. Therefore the disproportion in the roles arose as another aspect that needed reviewing.

In terms of language production time, the overall performance of a single student accounted for about 5 to 6 minutes, which the team evaluated as insufficient to allow for objective and reliable assessment. Consequently, an additional test task capable of generating more language in novel ways was searched.

As an invaluable contribution facilitating the revision process turned out to be the recordings of the oral part recorded for the purposes of benchmarking and project archiving. When discussing the performances during the benchmarking sessions, some of the issues described above surfaced quite naturally calling for remedial action. Furthermore, as we presented some of the recordings at a seminar on the role of interlocutors for external opinion, these became the subject to scrutiny of external language teachers who confirmed our suspicion of the format not eliciting representative chunks of language and the required CEFR level.

After the termination of the exam period, the specifications of the new format were adjusted to comply with the decisions made about them in the course of the exam period in order to prompt students to more extensive and superior language production. The major change lied in shortening the prompts the students received on drawing a topic. Table 2 shows the result of the revision process for question 12, Securities, to compare with the first version. The monologue was transformed into a simple statement requiring students to reflect and expand on the importance/significance/drawback/role etc. of a certain aspect of a specific business topic. This was expected to allow the test takers to broaden the range of functional language production as is expected of C1 level students, i.e. to prompt them into “speculating about the causes and consequences, weighting advantages and disadvantages of different approaches”(CEFR).

In order to obtain a more rateable sample we also added, what is now called part 2, a list of 10 questions specific for each topic, of which the interlocutor would choose two or three appropriate ones for a particular testing situation. This is to represent the achievement part verifying students’ knowledge of the subjects covered in the course of the four semesters. The third part remained a discussion between two students, as seen in table 2, with the input now reduced to the absolute minimum, just enough to set the scene and assign the task. The previously two different roles for each student were eliminated and combined into one providing the paired students with equal starting point in the discussion while not disadvantaging either of them. Students are presented with a situation and asked to discuss it with the partner while covering all three points. The time limit to do so is 3 to 4 minutes with two minutes for preparation. The bareness of instruction, the teachers believe, will coerce students to rely on their acquisition of language more than in the case of the previously designed format. Teachers expect that this will better serve the purpose of verifying whether the test taker reached the C1 level, i.e. whether they "have a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions, with little obvious searching for expressions or avoidance strategies and where only a conceptually difficult subject can hinder a natural, smooth flow of language" (CEFR).

Table 2: The assignment for the oral part of the examination, question 12, Securities.

12. SECURITIES

Monologue

*Choose **ONE** of the following questions and answer it. Your preparation time is **2 minutes** before you start. To fulfill the task, you are supposed to talk for **2 minutes**:*

1.	The importance of diversifying investment
2.	Ways of raising capital for further growth

Masarykova univerzita
Centrum jazykového vzdělávání na Ekonomicko-správní fakultě
BPJ_JI4A Jazyk I/4 - Angličtina

Teacher questions:

1. What is the difference between equity financing and debt financing?
2. What types of securities are the most common ones?
3. What happens on the stock exchange?
4. What is the role of a stockbroker?
5. How does short selling work?
6. How does insider-trade dealing work?
7. What affects stock value in free market economies?
8. What kinds of investment do you consider safe?
9. What are the advantages of investing in bonds over shares?
10. How do venture capitalists get profit from their investments?

12. INVESTMENT AND SECURITIES

Dialogue

***Role-play** the following situation using the guidelines below. The length of your role-play* *should be **3 - 4 minutes**:*

A company is considering investing a percentage of last year's net profit of your company in order to generate returns for future expansion. Discuss the situation with your partner and decide • **where the company should invest the money** • **what returns might be expected** • **the risk involved in the selected options**

Since the reviewed format has not been tested yet, the presumptions will not be confirmed or disclaimed until the end of the 2014 examination period. The team have developed, discussed and incorporated propositions of other colleagues to the revisited test tasks with a new set of students' test tasks on output. As the Impact project obligates us to another round of benchmarking sessions, new sample recordings will be made and discussed following the examination period. Teachers do believe that the outcome will bring the desired result and students' production will outperform that elicited by the previous tasks in terms of scope and quality.

7. Rating scales

All the efforts to enhance quality testing and standardise the outcome would be to little avail had we not had a set of rating scales for B1-C1 CEFR levels developed by a small team made up of the language centre teachers. Spoken performance analytical rating scales development had been enshrined in the project documentation at the onset of the project as an indispensable part to anchor the standardised format of new tests.

As already mentioned, the rating of spoken performance had been rather arbitrary with little respect to inter-rater or intra-rater reliability, i.e. little consistency of rating students' performance by different teachers or even the same teacher on different occasions. No language department of any of the 9 faculties of the university followed a thoroughly specified set of rating criteria to assess students' performance as that had not been considered essential. Teachers examined students they taught and it was difficult to detach the "emotional bond" that quite naturally originated, especially in the classroom environment where communicative methods with many interactive activities make the core of tuition. Therefore, the reliability aspect was side-lined and rarely questioned. Unsurprisingly, the imposition of the analytical rating scales met with a wave of protests from some of the teachers/testers who argued that there was no need for such scales as the existing procedure served the purpose just as well. As for the complexity of the issue, rooted in a variety of motivations of individual teachers, and lengthiness of the debate over the need of the imposition of the scales' criteria on the assessment process over the entire language centre, this aspect will not be discussed further here.

However, as it is the task of teachers to provide quality teaching, it is also their responsibility to certify that students are assessed reliably and fairly. Luckily, it should be stressed that the debate at the faculty of Economics and Administration was a constructive one and that the whole staff soon understood that any effort in standardising the format would only stand as a cosmetic remake scratching the surface, not enhancing the validity and reliability aspects, should there be no firm grounds for assessing reliably. The staff also understood that using the rating scales for assessment and learning to apply them to the individual performances uniformly is a time and energy consuming process, which they have all agreed to commit themselves to.

The rating scales are divided into four criteria, each consisting of several components. Each criterion ranges from 0 to 3 points, i.e. the maximum number of points a student can reach is 12. Since students can get a maximum of 20 points at the oral part of the exam at the Faculty of Economics, a conversion table has been developed to calculate a corresponding number of points from the 12-point scale. Decimal numbers are always rounded up to the higher number to advantage rather than disadvantage students. The rating scales were originally intended as a single assessment tool for all faculties, but recently it has turned out that the needs of individual faculties might be specific and the original scales might need to be modified to reflect specific requirements of test formats of these faculties.

The basic assessment criteria include task completion, lexical and grammatical means and pronunciation for B1 – C1 CEFR levels. For each of the level, the components differ slightly to reflect its specifics. For task completion at the C1 level there are four components – explanation of the task, engagement in discussions, argumentation and autonomy. For lexical means these are range and accuracy of general/specific/academic vocabulary; and complex grammatical structures of specific and academic language and a wide range of cohesive devices for grammatical means. The last criterion, pronunciation, is composed of pronunciation and intonation. Each component is further defined on the scale from 0 to 3 points, being graded progressively to differentiate test-takers' performance.

During the 2013 spring exam period teachers were making acquaintance with the rating scales, setting off the analytical approach to assessment with a benchmarking session. With the adoption of the new format of the oral test, on their own accord teachers started using the scales to check and practise their applicability. The split role of raters and interlocutors enabled them to focus on analytical assessment and soon we acknowledged the benefit of having a firm basis for reliable assessment. The 2014 spring exam period is the first exam period for which the use of the scales will become compulsory.

The whole standardisation process has raised a range of questions, with many being solved in the course of the project and others waiting to be answered. Setting the score for failed/passed decision about a performance is one of those that cannot be answered in the current environment with the cut-score setting process by far exceeding time and financial limits. The English language teachers at the Economics Faculty have decided to cut the score at 12 points (out of the maximum 20) a level where, they are convinced, the performance would comply with the C1 CEFR level.

8. Conclusion

The standardisation of the assessment process that has been a part of the Impact project has resulted in a complete revamp of the oral part of the final examination at C1 CEFR level at the Faculty of Economics and Administration of Masaryk University. A major shift towards assessment of spoken performance, independent of other language skills, proposed in gradual steps received impatient reception by the English language team who seized the opportunity to implement the valuable knowledge of basic principles of valid and reliable testing into practice. The theoretical knowledge acquired in a series of seminars provided under the auspices of the Impact project helped to build a solid theoretical base at all faculties of the University. Teachers at the Economics Faculty felt it pertinent to question the old practice and devoted a lot of time and energy to change it for the better. The new oral test format being free of other skills, teachers having the assessment tool in the form of analytical rating scales and a newly established peer-to-peer format are all a huge development. The fact that teachers opted for splitting the role of interlocutor and rater, though at the expense of time sacrifice, is also an important aspect in the perception of fair and reliable assessment as a priority in higher education.

References

[1] DOUGLAS, Dan. *Understanding language testing*. London: Hodder Education, 2010, ix, 156 s. ISBN 9780340983430.

[2] Brooks, L. Interacting in pairs ina test of oral proficiency: Co-constructing a better performance. Language Testing, 2009 26 (3), 341-366

[3] Brown, A. & Ducasse, A. M. *Assessing paired orals: Raters' orientation to interaction*. Language Testing, 2009 26 (3), 423-443

[4] http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf Cambridge University Press. ISBN : HB 0521803136 - PB 0521005310 www.uk.cambridge.org/elt