

Classroom-based assessment of group discussion: Challenges and opportunities

Olga Y. Lankina, St. Petersburg State University

Yulia V. Petc, St. Petersburg State University

<https://doi.org/10.37546/JALTSIG.CEFR3-8>

This article is open access and licensed under an Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

Collaborating in a group and leading group work are often used in the modern language classroom. However, it still seems relevant to find an effective way to assess group discussions and to identify problems that students experience when they have to mediate texts and concepts or collaborate on shared tasks. This paper analyzes the engagement of students with a task to mediate and discuss information in a group and how students approach the tasks at the B1 and B2 Council of Europe (CoE) or CEFR levels. The objective of this paper is to show that group discussions can be used for the formative and summative assessment of mediation skills as they are described in the CEFR Companion Volume (CEFR/CV; CoE 2018). We will describe the process of how learners can receive global achievement marks for mediation on par with analytical marks for interaction, discourse management, range, accuracy, and phonological control. We also provide an example of how assessment of a group discussion can be done by giving students a global achievement mark for mediation and analytical marks. The outcome of the research is a test that can be used in the classroom and to provide criteria for assessing mediation when it is part of a group discussion.

Keywords: mediation, classroom-based assessment, descriptors for mediation, global and analytical marks, CEFR/CV, group discussion

1 Introduction

Conveying information and discussing it is a prerequisite for effective cooperation. In order to understand the nature of communication in a group, we looked at the competencies that underlie group discussions. Interactional competence, which is a basis for group discussions, includes discourse competence, non-verbal communicative strategies, and strategic competence (Celce-Murcia et al. 1995). This is demonstrated when people “interact as speakers and listeners to construct meaning” (Fulcher and Davidson 2007: 49). These are the most social speech acts because they are co-constructed (McNamara 2006: 64). Researchers point out the complexity that this social element adds to co-constructed speech (Luoma 2004; Lazaraton 2014). Discussants construct the event together and share the right to influence the outcomes. Besides, in situations of information-related talk aimed at conveying information on a particular issue, speakers have to deploy several strategies to establish common ground, give information in bite-sized chunks, ensure logical progression, and maintain a comprehension check. (Luoma 2004: 20). These features of co-constructed speech pose certain challenges for students who learn how to mediate and discuss information in an academic environment.

CEFR frames mediation as a language activity in its written or oral form, which makes communication possible between people who are not able to communicate with each other directly (CoE 2001: 14). The co-construction of meaning while collaborating in a group is described in the (CoE 2018), which emphasizes the mediator as a social agent of communication. It is very important that when we use a language we combine several activities and switch between the modes of communication: mediation,

reception, production, and interaction (CoE 2018: 33).

Since 2001, linguistic, cultural, and social aspects of mediation have been thoroughly discussed and mediation has been linked to the context of school education (Zarate et al. 2004; Coste and Cavalli 2015). Mediating concepts in a collaborative work context implies, on the one hand, the learner's contribution to the effective work of the group towards achieving a shared objective, and on the other hand, his/her stimulating and developing ideas as a member of the group. The latter is characterized by the learner's ability to further develop other people's ideas and opinions, co-develop ideas, and evaluate problems and suggestions. The definition of mediation was developed by Coste and Cavalli and their proposition about the distinction between *cognitive mediation* (constructing or conveying meaning) and *relational mediation* (facilitating relationships) (Coste and Cavalli 2015: 28) became the key principle for designing the Global Achievement Scale used in this research. We also used mediation activities as listed by North and Piccardo (North and Piccardo 2016: 21) and described in the Illustrative Descriptor Scales (CoE 2018: 116) to write the contents of the Global Achievement Scale.

On the assumption that young adults should learn a foreign language and develop their social competencies together (Canale and Swain 1980; Long and Porter 1985; Sharan 1990; Slavin 1990; Pavlovskaya 2017), group work is a fundamental condition of language learning. It also plays an important role in the occupational and academic domains and therefore, has to be presented in an adequate format in language tests.

2 Research objective and participants

Having faced the task of teaching students to mediate texts and concepts, we noticed that students find it difficult to engage in mediation in group discussions and develop each other's ideas.

The objective of the research was to investigate the extent to which there is a deficiency of certain skills, i.e. developing other people's ideas among CEFR B1 and B2 level students, which does not allow students to hold discussions successfully.

The research participants were undergraduate students in the Management Department of St. Petersburg State University (Russia; 91 people) and two trained examiners. The students' levels of English language proficiency varied between CEFR B1 and B2. In compliance with the standards for foreign language education applied at the university, students at the CEFR B1 and lower levels are taught General English. Students who have achieved the level of language proficiency equal to CEFR B2 or higher go on to study Academic English and English for Specific Purposes, particularly Business English in the Management department. Most of the students can be described as motivated and willing to learn the language for academic purposes (e.g., a semester abroad program, lectures in English delivered by non-native and native speakers), professional development (e.g., scientific research), and career opportunities.

Students were grouped into three cohorts according to the results of the placement test: B1 low, B1 average, and B2. The research tests were carried out in the middle of the academic year. By that time the students' proficiency had not changed considerably. The students worked in three-person groups. In cases, when it was not possible to organize a three-person group, groups of four were formed. Because the number of four-person groups was minimal, the impact of group size on the results was not taken into consideration. The details of the groups are presented in the table below.

3 Methodology – Test description and test marking

Three Speaking tasks were used for measuring oral performance in group discussions at the CEFR B1 and B2 levels. The assessment tasks were designed to conform to the foreign language curriculum content. Though the research participants were offered different tasks in terms of content, each task was tailored to comply with the students' language proficiency level and the course requirements. The tasks were designed to require the students to discuss a situation by mediating the background information, expressing opinions, and negotiating an agreement. For example, in General English lessons, students

worked in groups of three or four people. Each student had to present their charity event proposal and the group decided which event would be the most successful. More advanced students engaged in Business English study also worked in groups of three or four people, and had to discuss different leadership styles then choose those which they thought would best suit their educational institution. To familiarize the participants with this kind of task, they completed similar tasks with different course materials under the guidance of a teacher prior to the assessment stage.

Table 1. Student profile

Characteristics of groups	Cohorts		
	B1 low	B1 average	B2
3 students	5 groups	4 groups	16 groups
4 students	2 groups	2 groups	-
Total number of students	23	20	48
Type of English class	General English	General English	Academic Communication skills in English for Managers
Course details	Two semesters, 90-min face-to-face class twice a week	Two semesters, 90-min face-to-face class twice a week	Two semesters, 90-min face-to-face class twice a week.

Table 2. Test details

Test level	B1	B1 +	B2
Test topic	Charity (B1 test)	Education (B1 test)	Leadership (B2 test)
Input	Written (80-word role cards)	Written (150-word role cards)	Video (8-minute videos)
Preparation time	2 min	3 min	5 min
Output (speaking time)	10 min	12 min	15 min
Test length	12 min	15 min	28 min
Procedure	Each student is given a written stimulus, a unique option, and a task, then asked to present their options and discuss an issue.		The procedure is the same as the B1 level, except for input, which was a video.
Task focus	Mediating factual and general information, e.g., options for charity or education.		Mediating factual, general and abstract information, e.g., leadership styles.
	Using functional language to discuss options, invite to speak, ask straightforward questions, give reasons for views, repeat back to confirm mutual understanding, define the task, collaborate on a shared task, and maintain the focus of a discussion.		Using functional language to further develop other people’s ideas and opinions, present ideas and invite reactions, and consider different sides of an issue; organize a discussion, refocus it, highlight the main issue, and collaborate in decision-making.

Each student in a group was provided with a short input text on a common topic and was given 2-5 minutes to familiarize themselves with the content. After that, the texts were retrieved, and the students were required to convey the main idea of their input text ensuring all of the group members' comprehension. The task rubric was formulated in a manner that required students to facilitate discussion by inviting others to speak, express their agreement, and contribute towards fulfilling the goal. To create the conditions for uninterrupted communication, the teacher would refrain from guiding students through the task. The teacher evaluated the students, awarding Analytical marks and a Global Achievement mark. The tasks were recorded to provide the possibility for marking multiple times.

While discussing the topic, students had to relay information by processing, reformulating, summarizing, or streamlining it (cognitive mediation). At the same time, they aimed to build rapport with the other students in the discussion group, creating conditions that were instrumental for facilitating and moving the discussion towards accomplishing a shared communicative goal (relational mediation). Thus, mediation of a text, concepts, and communication became the global objective of the task and was reflected in the Global Achievement mark, which was awarded on the basis of such descriptors as relaying specific information in speech, facilitating collaborative interaction with peers, and collaborating to construct meaning.

The Analytical criteria were interaction, discourse management, range, accuracy, and phonological control. The CEFR/CV descriptors for overall mediation, mediating concepts via collaborating in a group, leading group work, and mediating communication (CoE 2018: 101, 116-117, 119, 120-123) were adapted by the authors to write the Global Achievement Scale, while the Analytical Scale was drawn up without amending from Table 3: Common Reference Levels: qualitative aspects of spoken language use (CoE 2001: 28-29). The authors of the paper who were the students' teachers and examiners agreed on the 3 and 5-band performances which were used as standardized performances to refer to while assessing.

Both the Analytical and the Global Achievement scales had 5 points, where 1, 3, and 5 bands were described. Students' performance was reported in terms of 1-5 marks where '1' is the lowest and '5' is the highest mark.

4 Results and Analysis

The tests' results were analyzed using the Rasch Model (FACETS) and Classical Test Theory (ITEMAN).

Table 3. Summary statistics of ITEMAN and FACETS analysis

ITEMAN		FACETS											
Level	Topic	Number of students	SD (standard deviation)	Min Score	Max Score	Item mean (aver. of the criterion) marks	Alpha (KR 20)	SEM (standard error of measur.)	Mean Rater correlate. (PtMea)	Item mean	Mean Resd	Mean StRes	Mean S.D.
B1	Charity	23	4,734	13	30	3.485	0.951	1.044	0.81	3.48	0.00	0.02	1.02
B1+	Education	20	5,089	11	30	3.833	0.952	1.113	0.75	3.83	0.00	0.00	1.00
B2	Leadership	48	3,627	18	30	3.917	0.925	0.992	0.84	3.92	0.00	-0.01	0.99

As we can see in the table, the most difficult test, or the less able students, were the B1 cohort who took the Charity test (Item mean 3.48), while the least difficult test or the best-prepared students were those who took the B2 Leadership test. According to the FACETS User Manual, when the parameters of the test are successfully estimated, the mean Resd is 0.0. In our tests, it was 0.00. When the data fit the Rasch model, the mean of the Standardized Residuals (StRes) is expected to be near 0.0 (in our tests it varied from -0.01 to 0.02), and the Sample Standard Deviation (S.D.) is expected to be near 1.0 (1.02; 1.00; 0.99). The raters' correlation was quite strong and exceeded 0.75. If we look at the raters' agreement closely in Figure 1 below, we can see that mediation and interaction have a larger discrepancy than the other criteria, which can be explained by the fact that even though the raters had prepared for the assessment session, assessing mediation was still a new experience for them.

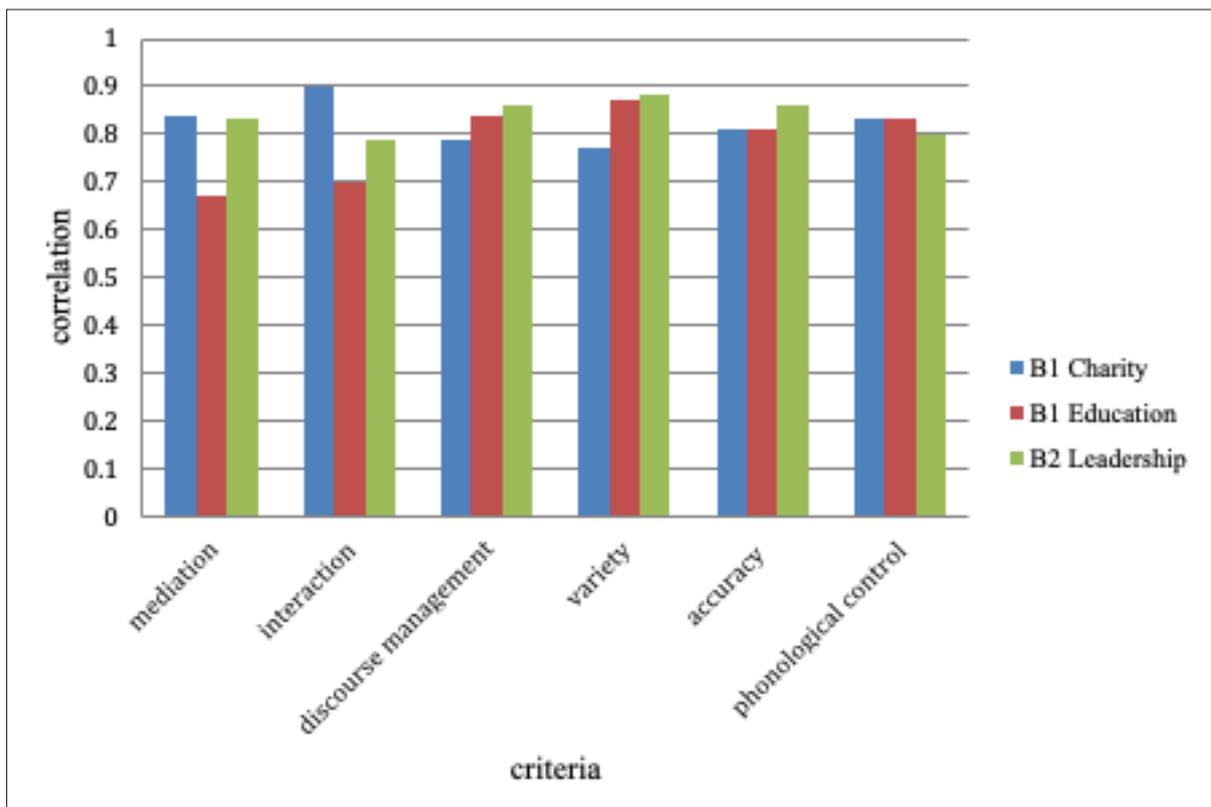


Figure 1. Raters' correlation.

Our aim was to analyze the factors that impede students' ability to co-construct the new meaning. In order to do that, we compared the amount of time students spent on presenting their own views (or producing their long turns) and collaborating on the task. The quantitative analysis of students' oral performances revealed that students at a low B1 level tended to retell the content of their input cards to other group participants almost without debating the issue. At the B1 level the following pattern of group performance prevailed: a series of long turns followed by a short and quite rudimentary collaboration phase (B1 low 0.8:0.2; B1 average 0.66:0.34; the numbers here represent the long turn and collaboration phases as they relate to the length of the test). This changes at the B2 level where mediation became naturally integrated into the discussion, and long turns made up only one-fifth of the total length of the task (0.2:0.8, the long turn and collaboration phases respectively), as shown in Figure 2. This integration of mediation manifests itself in the way discussants (1) took turns: they became more confident in balancing contributions from other group members with their own; (2) switched easily between modes of communication, so that mediation, interaction, and production in their speech became intertwined. For example, while discussing leadership styles, a student briefly outlined the contents of the video she

watched and then pointed out one feature she liked most about “silent leadership”, that is “connectedness with the community”. The other student built upon that by specifying the circumstances when “silent leadership” would suit the society best, i.e. “a period of prosperity”, according to how she understood it. The third student provided details from the video and his own experience and argued that “paternalistic leadership” seemed to be the most effective for a big company. As we can see, students blended their references to the input videos and their own opinions, thus mixing mediation with interaction. It was interesting to note this feature of a discussion at the B2 level, and to match it with how “Facilitating collaborative interaction with peers” is described in the CEFR/CV (CoE 2018: 115).

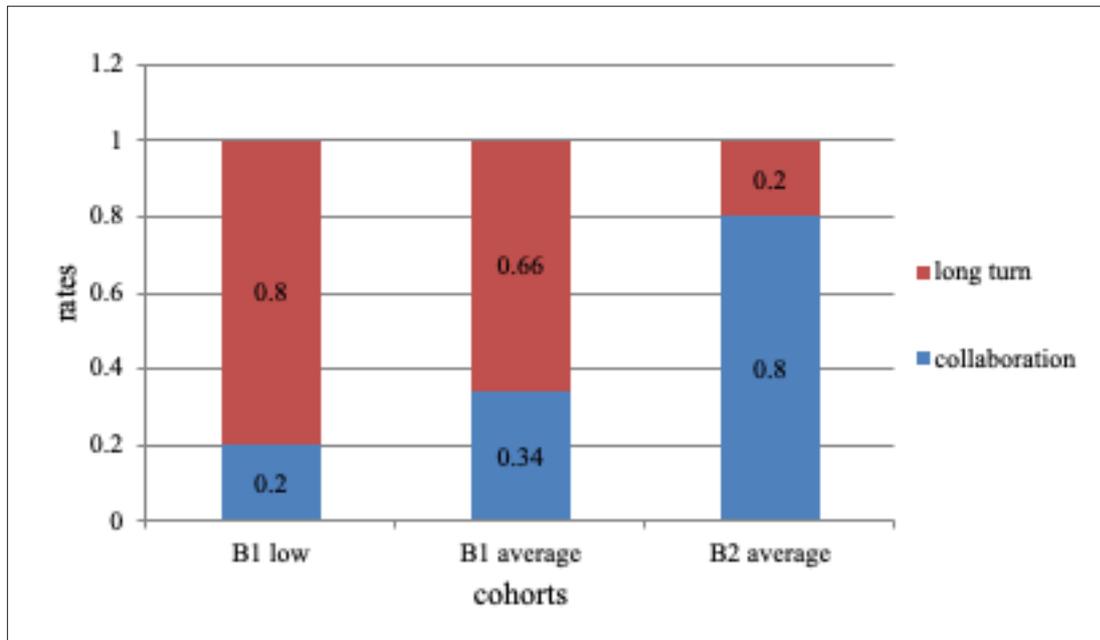


Figure 2. The rates of long turn and collaboration phases as they relate to the length of the whole test.

We also looked at how descriptors for mediation (collaborating in a group) work in discussions in the hope of better understanding the development of the students’ ability to build on what other people say. For this purpose, we made a list of 44 descriptors that were most suitable for our tasks and levels: descriptors for overall mediation, managing interaction, encouraging conceptual talk, and facilitating pluricultural space in CEFR/CV (CoE 2018: 101, 116, 119, 120-121). Then we shortlisted those descriptors to 17 that students were using in their speech and gave them ‘short names’, e.g., ‘collaborate on a shared task’ stands for “Can collaborate on a shared task, for example formulating and responding to suggestions, asking whether people agree, and proposing alternative approaches.” (B1. Facilitating Collaborative Interaction). ‘Consider different sides’ is used for the descriptor “Can consider two different sides of an issue, giving arguments for and against, and propose a solution or compromise.” (B2. Collaborating to Construct Meaning). (CoE 2018: 116). The occurrences for each descriptor were counted in each cohort and descriptors were sorted in order of frequency, as shown in Figures 2-4.

At the B1 low level, when faced with the problem of shared decision-making, students resorted to straightforward strategies such as stating the goal of the discussion (e.g., *We have to choose the best idea*), turn-taking, inviting partners to contribute to the discussion (e.g., *What do you think? Lena, what’s your idea?*), expressing agreement/disagreement (e.g., *Yes, I agree; No, I don’t like it.*) They also repeated back (A: *Do you mean that children from a hospital can cook and take part in the competition?* B: *They can, I think. But I say that all people who want can cook something and people choose a winner.*) and maintained the focus of the discussion (A: *Where will we get products for the competition?* B: *I’m sorry but we should talk not about where we take food or the place where we keep equipment ... we should think about what we can do to get some money for the children’s hospital.*) As we can see in Figure 3, the descriptors of B1

level dominate here; however, there are some instances of using descriptors from higher levels (B2; considering the pros and cons of an issue). The numbers in the figure show the occurrences for this particular cohort.

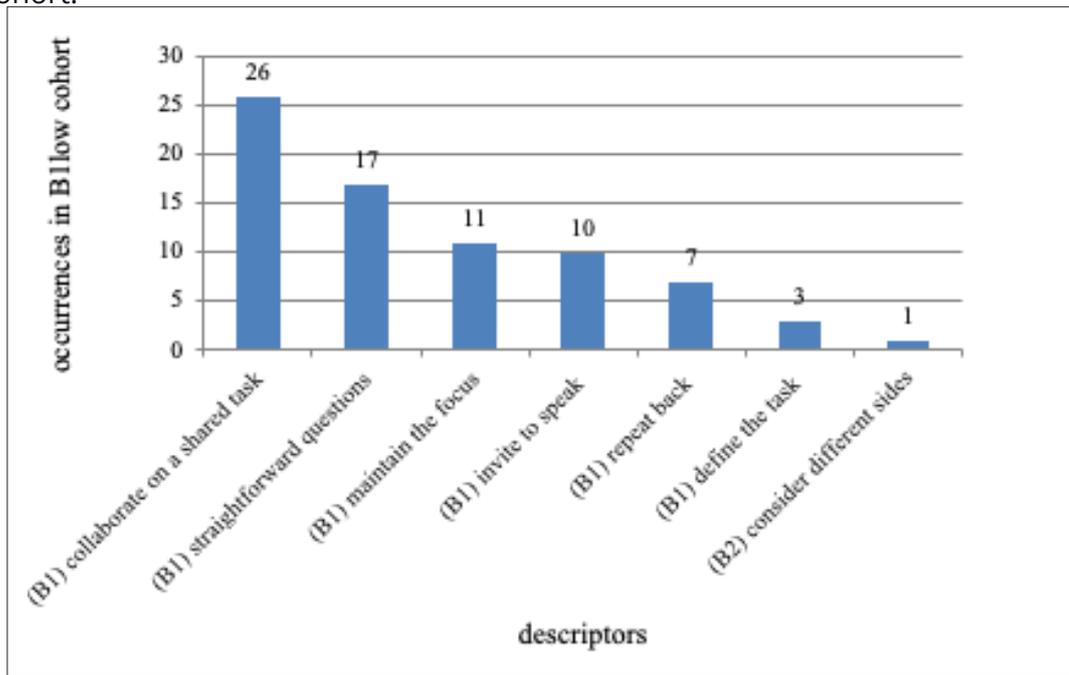


Figure 3. B1 Low cohort: descriptors distribution.

Figure 4 is related to the average B1 level, where one can find even more attempts to use sophisticated language functions, such as developing other people’s ideas and considering two different sides of an issue (e.g., A: *In my class, there were strong and weak students. B: I see what you are saying. It’s good to mix students, but it can be hard for those who are ambitious*); or invite other people’s reactions (e.g., A: *What do you think we can borrow from the educational systems of other countries?... B: I think it would be nice if school in Russia started at 9 o’clock or later like in the Netherlands. What do you think?*) In our tests, developing other people’s ideas as a language function was first used at this level (position number 8 in Figure 4).

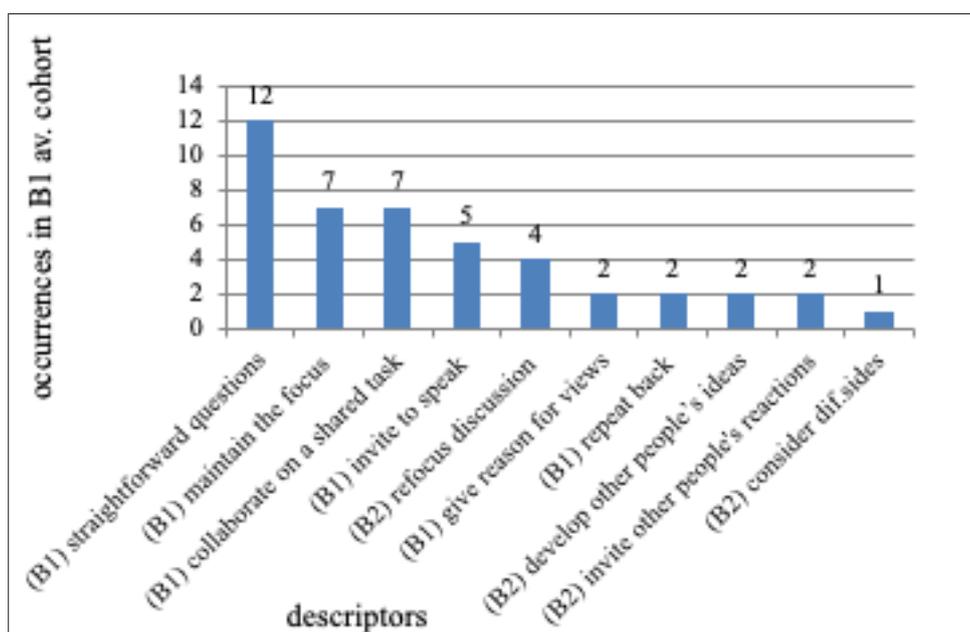


Figure 4. B1 Average cohort: descriptors distribution.

At higher levels, discussions became lengthy and detailed, and students demonstrated a wider repertoire of exploited strategies. They showed their ability to organize and manage collaborative work (e.g., *Today we're going to talk about different types of leaders.... And first of all, we should understand what type of leadership everyone watched a video about. Let's talk about different types of leaders and then we'll discuss what leadership style can be used at our school*). Students co-developed ideas, offered suggestions, compared different points of view, summarized, etc. (A: *The type of leadership depends on a person's characteristics. B: Apart from a leader's personality, social milieu plays an important role in making this or that type of leadership popular. C: Right, so we have personal and social factors here. Which would be more important?*) Even though developing other people's ideas as a language function was used more often by the B2 students, it still found itself at the tail of the descriptors distribution (position number 8 in Figure 5).

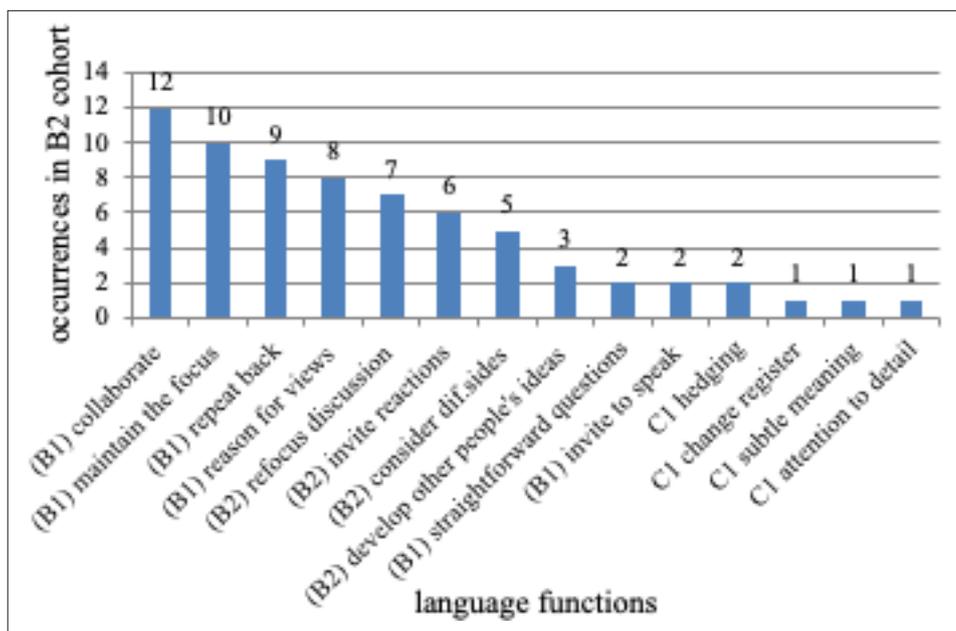


Figure 5. B2 cohort: descriptors distribution.

5 Conclusions

The tests that have been described in the paper were designed with the aim of assessing students and identifying gaps in their language skills including mediation. This type of test task can be used for the formative and summative assessment of mediation in group discussions, and language proficiency in general.

The descriptor analysis showed which descriptors were used by students often and successfully, and which descriptors students did not pick. Judging by the marks that students received, we can see which language skills had been developed and to which extent as well as which skills were lacking. In this way, the use of CEFR/CV descriptors can help identify problem spots in students' skill development, and backward design can be used to cover the lacunae.

We also observed changes in language behavior with the progression of the language level: (1) if students are given a task to mediate and discuss some information, at lower levels they tended to complete the task in two distinct phases, i.e. relaying information first and then interacting to solve the task; while at higher levels relaying information and interacting were blended and continued throughout the task; (2) the variety of descriptors for mediation which can be attributed to a speech act increases from lower to higher levels. These features of students' language behavior can give an examiner or a teacher a good sign when determining the level of a test-taker.

6 Limitations and areas of further research

The descriptors that were not observed in the test were either not required for the task, or students lacked the skill to use them. Backward design in planning the curriculum can be used to cover the outstanding descriptors. In backward design, educators first identify learning outcomes, then write tasks for assessment, and finally create learning activities to achieve the desired results (Wiggins and McTighe 2008: 17-18).

The group format (3 or 4 people) and the lack of interference of the Interlocutor into group work creates an authentic context, which can be viewed as a positive feature of this type of exam. At the same time, one can argue that the absence of the Interlocutor can disadvantage shy or lower level candidates; however, we believe that the situation when candidates have an opportunity to communicate freely provides plenty of room for mediating communication and its assessment. It may also have a positive effect on teaching and sharpening those skills in the classroom setting.

A larger sample of oral performances and the involvement of more assessors would increase the reliability of the measurements. Also, a more refined focus on the nature of collaboration in a group can give more information on how to teach and assess group discussions. For example, it might be worthwhile researching whether the cooperative and competitive modes of communication affect successful task completion.

We have noticed that the skills for mediating concepts are quite often underdeveloped as a social skill among the target population of students who lack the ability to articulate their stance clearly. This can be viewed as a problem area which requires the attention of educators and test developers. These tests, as well as other similar tests, can be offered as a tool of measurement of progress on the way to working out these problems.

7 References

- Canale, Michael & Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1). 1–47.
- Celce-Murcia, Marianne, Zoltan Dörnyei & Sarah Thurrell. 1995. Communicative competence: a pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6(2). 5–35.
- Coste, Daniel & Marisa Cavalli. 2015. *Education, mobility, otherness. The mediation functions of schools*. Strasbourg: Council of Europe, Language Policy Unit. DGII.
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. Strasbourg: Council of Europe. <https://rm.coe.int/16802fc1bf> (accessed 20 August 2020).
- Council of Europe. 2018. *The CEFR Companion Volume with new descriptors*. Strasbourg: Council of Europe. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989> (accessed 03 August 2020).
- Fulcher, Glenn & Fred Davidson. 2007. *Language Testing and Assessment. An advanced resource book*. London and New York: Routledge.
- Lazaraton, Anne. 2014. Second language speaking. In Celce-Murcia M., D. M. Brinton & M. A. Snow (eds.), *Teaching English as a second or foreign language*, 4th ed., 106–120. Boston: Cengage Learning.
- Long, Michael H. & Patricia A. Porter. 1985. Group work, interlanguage talk, and second language acquisition. *TESOL Quarterly* 19(2). 207–228.
- Luoma, Sari. 2004. *Assessing speaking*. Cambridge: Cambridge University press.
- McNamara, Tim & Carsten Roever. 2006. *Language Testing: The social dimension*. Malden, MA: Blackwell.
- North, Brian & Enrica Piccardo. 2016. *Developing illustrative descriptors of aspects of mediation for the Common European Framework of Reference (CEFR)* <https://www.rm.coe.int/developing-illustrative-descriptors-of-aspects-of-mediation-for-the-co/1680713e2c> (accessed 03 August 2020).

- Pavlovskaya, Irina (ed.) 2017. *Актуальные вопросы языкового тестирования. [Modern issues of language testing]*. 2. St. Petersburg, 684 p. URL <http://hdl.handle.net/11701/7188> (accessed 03 August 2020).
- Sharan, Shlomo. 1990. *Cooperative learning: theory and research*. New York: Praeger Publishers.
- Slavin, Robert E. 1990. *Cooperative learning: Theory, research and practice*. Englewood Cliffs, NJ: Prentice-Hall.
- Wiggins, Grant & Jay McTighe. 2008. *Understanding by Design*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Zarate, Genevieve, Aline Gohard-Radenkovic, Denise Lussier & Hermine Penz. 2004. *Cultural mediation in language learning and teaching*. Strasbourg: Council of Europe Publishing.

Software

- ITEMAN <http://www.assess.com/iteman/>
- FACETS <https://www.winsteps.com/minifac.htm>
- FACETS User Manual, <https://www.winsteps.com/facetman/index.htm>

8 Biography

Olga Y. Lankina is a researcher. Her field of research includes language testing and mediation. She holds a specialist degree in applied linguistics and works as an ESP teacher and teacher of English for academic purposes at St. Petersburg State University (Russia). Olga is a Speaking Examiner and examiner trainer. In 2013-2014 she took part in linking the University test to the CoE. She is a member of EALTA and UKALTA.

Yulia V. Petc holds a Master's degree in Pedagogy and Linguistics. She works as a language test developer at the Language testing department of St. Petersburg State University (Russia) and an ESP teacher and teacher of English for academic purposes at St. Petersburg Mining University (Russia). Yulia is a Speaking Examiner. In 2013-2014 she took part in linking the University test to the CoE. She is a member of EALTA.