How new CEFR mediation descriptors can help to assess the discussion skills of management students—Global and analytical scales

Irina Y. Pavlovskaya, St. Petersburg State University

Olga Y. Lankina, St. Petersburg State University

The article focuses on the assessment of mediation competence in the context of the Content and Language Integrated Learning (CLIL). We offer new assessment scales developed with the use of descriptors for mediation from the *CEFR Companion Volume* (2018). The approach to assessment of oral performance that we discuss combines global and analytical marks. For the majority of classroom teachers in Russia, this issue has become very important from two points of view: a) how to introduce new scales of mediation and connect them adequately with traditional speaking skills, described in the literature (Pavlovskaya 2017), and b) how to harmonize global assessment with analytical scales. The research is based on the experience of evaluating the mediation skills of students of the Graduate School of Management, St. Petersburg State University. The implications of the method for classroom teaching are discussed.

Keywords: mediation, oral performance, assessment, global and analytical marks, global achievement scale, analytical scale, CEFR descriptors, cognitive skills, relational skills, group discussion.

1 Introduction

CLIL teachers of management students always have to be on alert, looking for the 'soft' and 'hard' skills that students might need most. Mediation is partly a hard skill, because it is firmly based on proficiency in a foreign language as well as on the relevant professional knowledge, but it also covers the top 10 soft skills that are so attractive for employers (communication, flexibility, leadership, motivation, patience, persuasion, problem-solving abilities, teamwork, time management, work ethic) (*hard* skills *vs. soft* skills).

In our case, the aim of the classes is to develop language-related skills that managers may need at work. We think that facilitating and encouraging conceptual talks has become an important professional task of a manager. With this idea in mind, we focus on three task types: 1) how to facilitate discussions, 2) how to give persuasive talks, and 3) how to deliver business presentations. All of these tasks require mediation strategies.

Mediation, as it is defined in the CEFR Companion Volume (CEFR/CV), implies "passing on new information in an appropriate form; collaborating to construct new meaning; encouraging others to construct or understand new meaning, and creating the space and conditions for communicating and/or learning." (CEFR/CV 2018: 99). We also adopted the approach to learning as described by Brian North (North 2016: 9), who states that learners, and especially those who learn a foreign language, are usually confronted with the unknown, having to mediate new meanings to each other and thus find themselves challenged by situations that require reformulating a text or *mediating a text* (CEFR/CV 2018: 103-114). Alternatively, they have to *mediate concepts*, e.g. do problem solving, brainstorming and

concept development (CEFR/CV 2018: 114-119). The third type of mediation, *mediation of communication* (CEFR/CV 2018: 120-123), is less relevant to this particular environment, due to the fact that, linguistically and culturally, the students happen to be quite homogeneous.

It is crucially important to find an effective way of assessing the oral performance of students who are involved in group discussions on professional issues. To meet this challenge an empirical research setting has been employed using both global achievement and analytical scales.

2 Research setting

The research involves B.A. programme undergraduate students at the Graduate School of Management, St. Petersburg State University, Russia, and their teachers of English (See Table 1.)

Table 1. *Research participants*

1	Number of students	49
2	English language proficiency	B2+ / C1
3	Age	19 – 21
4	Department	Management
5	Teachers	3

The students speak advanced English and most of them have successfully passed IELTS, B2 First or C1 Advanced Cambridge exams. Within the university curriculum, they have two English classes a week, 90 minutes each. There are three teachers who have experience in rating speaking exams and who took part in a CEFR-linking project (familiarization, standardization training and cut-score setting). This background gives them a better understanding of new CEFR descriptors for mediation that are being used for assessment purposes within the research.

In the third and fourth semesters of their studies, students carried out a project on developing business plans for startups that they might launch in the future, for example, a family leisure club, online language courses, a waste collection company, a communal heating system or an urban park. Students worked in groups of three or four and presented their plans to the other groups. They facilitated discussions and gave persuasive talks. The most common classroom activity within this project was a discussion. During discussions, students informed their group members about the details of their business. For example, they explained how they created business budgets and estimated risks, or they asked for advice on how best to manage their startups.

The teachers tested the students at the beginning of the academic year to see how good the students were at holding group discussions. Then the students were divided into two cohorts, which we refer to as the 'Control Group' and the 'Experimental Group'. Both cohorts followed the standard program of English adopted by the University, but the Experimental Group did an additional component, which involved exercises in mediation and self-assessment with CEFR descriptors. Both cohorts had a similar time schedule of classes: four academic hours (45 minutes) per week, 15 weeks in a semester, which is a total of 120 hours per year. The discussions within the Experimental Group employed the techniques typical of mediating texts and mediating concepts, such as linking to previous knowledge, amplifying or streamlining the text, solving problems, inferring, etc. All of the students took an oral test at the end of the course.

3 Research question

The research question was as follows: How can we effectively integrate mediation into the set of criteria for oral assessment? We approached this question with the understanding that students complete a communication task successfully if they display good mediation skills. In addition, we expect them

to be intelligible, coherent and logical when presenting arguments, employ an appropriate range of grammatical patterns, have considerable lexical resources, and demonstrate sufficient accuracy of speech. Therefore, the analytical criteria should include (1) interaction, (2) discourse management, (3) range, (4) accuracy, and (5) phonological control. We also understand that the mediation, production and interaction skills are highly interdependent. Indeed, if students are not sufficiently intelligible or they have some problems with the accuracy or fluency of their speech, it would be highly unlikely that they could cope with a mediation task successfully.

The mediator reformulates, summarizes or streamlines information. At the same time he/she is trying to build rapport within the discussion group. That is why in order to assess mediation globally the assessor has to ask two questions: 1) has the student managed to convey information clearly, and 2) has the student facilitated the discussion and collaborated successfully to construct meaning? The answers to these questions help the assessor to decide on the global achievement mark for mediation. Consequently, the global achievement mark that evaluates the mediation skills describes (1) relaying information and (2) facilitating discussions and collaborating to construct meaning.

Keeping this in mind, we can suggest that the assessment of oral performance in a group discussion on professional issues would be effective if it includes awarding analytical and global marks, so that *five analytical marks* are given for 1) interaction, 2) discourse management, 3) range, 4) accuracy, and 5) phonological control, and the *global mark* is given for mediation.

4 Research methodology

The oral performance assessment scheme was developed for this purpose. Firstly, we outlined the skills of oral mediation that students need to acquire. In order to list the skills that we wanted to assess, we analyzed the needs of the students and mapped them onto the descriptors for mediation. We grouped *cognitive skills*, which cover relaying a text, shortening a text, and elaborating on the text (see Table 2), and *relational skills* (see Table 3), which refer to mediating concepts: facilitating collaborative interaction, collaborating to construct meaning, managing interaction, and encouraging conceptual talk (CEFR/CV 2018: 116-117; 119).

Table 2. *Cognitive skills*

1	*Relaying a text*
	• Can paraphrase and render its meaning.
	• Can adapt the style and change register to meet the needs of the recipient.
2	*Shortening a text*
	• Can highlight the key points.
	• Can choose the relevant information.
3	*Elaborating on the text*
	• Can link the issue to previous knowledge.
	• Can explain difficult notions.
	• Can explain relationships between ideas.
	• Can generalize to explain the meaning of examples.
	• Can provide examples to give meaning to abstract ideas.
	• Can use metaphors and idiomatic language to sum up.
	• Can transform complex notions used in the text into passages that are easy to understand.
	• Can speculate about the inferences used by the author.

Table 3. *Relational skills*

	Facilitating and managing collaborative interaction in groups
	Can define goals of the discussion.
	Can stimulate a discussion.
	Can steer a discussion towards a conclusion.
	Can conclude a discussion.
	Can show sensitivity to different perspectives in a group.
	Can organize a group discussion.
	Collaborating to construct meaning + encouraging conceptual talks
	Can present their ideas.
	Can invite reactions from other group members.
	Can further develop other people's ideas.
	Can participate in the group discussion accordingly, e.g. contributing to collaborative decision-making, highlighting issues, evaluating problems, elaborating points of view.
	Can encourage the other interlocutors to conduct conceptual talks.

Sets of tasks on professional topics were created for the training and final assessment. For the final assessment, students watched one of several videos on leadership; then they met in a group of five or six people who had watched different videos. They received a question for a discussion based on the problems raised in the video and the project that students were involved in. Students had to share their knowledge and experience about leadership styles, discuss a problem taking the role of a leader, and attempt to arrive at a conclusion. Those tasks were aligned to B2 CEFR level using the CEFR Grid for Speaking.

Finally, the criteria for the assessment scales were defined and their descriptors were adapted from those for B2 in the CEFR and CEFR Companion Volume. These descriptors were used in five-point analytical and global achievement scales for bands 1, 3 and 5.

Technically, the assessor listens to a group discussion (5-6 people), which continues for about 30 minutes and involves presenting the information that the students have researched or gained before. In addition, the students discuss conceptual issues. During the discussion, the assessor awards analytical marks to every student. After the discussion, the assessor gives students global achievement marks for mediation.

The discussions were recorded during the experiment. Subsequently, they were assessed by three raters. The first rater was the teacher, who conducted face-to-face assessments. The other two raters, also teachers, assessed the recorded performances. They used audio scripts to help identify students. These raters had undertaken tuning-in with standardized performances before assessing students' discussions. The aim of the tuning-in exercise is to remember what 'strong', 'average' and 'poor' performances are like and the raters did tuning-in exercises before each assessment session. The raters' correlation lay between 0.87 and 0.91 (Table 4).

Table 4. *Rater correlation*

RATERS	PEARSON
1 / 2	0.91
1 / 3	0.93
2 / 3	0.87

5 Results and discussion

The results of the experiment were statistically analyzed with the help of Excel and ITEMAN. The data provided by the three raters were collected and the average mark used further for calculations. The maximum score is 30.

Table 5. *Central trend measures and classical statistics for the two groups' scores*

| | Statistics | Method of Calculation | Diagnostic Test | | Final Test | |
			Experimental Group	Control Group	Experimental Group	Control Group
1	Number of Participants	No program	24	24	24	24
2	Mean	Excel	20.63	17.35	21.11	17.54
3	Mode (bimodal distribution)	Excel	18.33	Mode 1 15.00; Mode 2 18.00	18.67	20.00
4	Median	Excel	19.17	17.67	21.17	17.50
5	Standard Deviation	Excel	4.03	1.88	3.40	2.40
6	Skew	Excel	0.32	-0.02	0.13	-0.07
7	Kurtosis	Excel	-1.02	-1.22	-0.95	-1.41
8	Min Score/ Max Score	No program	13.00/27.00	13.00/ 22.00	14.00/25.00	14.00/21.00
9	Mean Item	ITEMAN	No.	2.89	3.69	2.92
10	Alpha	ITEMAN	0.95	0.90	0.94	0.83
11	SEM	ITEMAN	1.00	1.02	0.97	1.01

High values of Cronbach's alpha, showing internal consistency of characteristics (Table 5, no. 10) and the measurement error not exceeding 1.02 (Table 5, no. 11) indicate the reliability of the test.

The values of the minimum and maximum scores (Table 5, no. 8) as well as the standard deviation (Table 5, no. 8) indicate a greater homogeneity of the Control Group in comparison with the Experimental Group. It should be noted that at the final test both groups demonstrated a more uniform level of skills development, which is confirmed by a decrease in the standard deviation.

Some heterogeneity in the population of the groups is indicated by the flat-topped distribution, expressed by a small negative Kurtosis (Table 5, no. 7).

The absolute value of the Asymmetry in both groups is not significant and does not exceed 0,32 (Table 5, no. 6), while remaining positive in the Experimental Group and negative in the Control Group. This may indicate the presence of several students in the Experimental Group who are demonstrating higher level of skills development and some students in the Control Group with a lower level. Nevertheless, we can state that the difference between Experimental and Control groups did not exceed 0.54 points at the beginning of the experiment and 0.77 points at the end (Table 5, no. 9), and is not significant for the purposes of our experiment.

Further statistical characteristics of the holistic criterion 'mediation' and analytical criteria 'interaction', 'discourse management', 'variability', 'correctness' and 'phonological control' were calculated with the help of the program ITEMAN.

For the Diagnostic and Final tests, the correlation of scores by the six criteria with the Mean score (Table 5, no. 2) were calculated (see Table 6).

Table 6. *Criteria scores and mean score correlation*

Criteria	Experimental Group. Diagnostic Test	Experimental Group. Final Test	Control Group. Diagnostic Test	Control Group. Final Test
Mediation	0.83	0.81	0.86	0.80
Interaction	0.91	0.87	0.84	0.72
Discourse Management	0.93	0.87	0.70	0.58
Range	0.88	0.85	0.72	0.71
Accuracy	0.88	0.79	0.87	0.64
Phonological Control	0.76	0.86	0.49	0.38

As we can see in Table 6, almost all criteria scores strongly correlate with the Mean, except for the phonological control, which is not surprising, as pronunciation does not necessarily correlate with overall communicative proficiency. It is noteworthy that the Experimental and Control groups differ in the way that mediation, interaction, and discourse management statistics changed from diagnostic to final tests (see Figure 1).

Figure 1. Diagnostic test (blue) vs. Final test (red) results in a) Experimental and b) Control groups
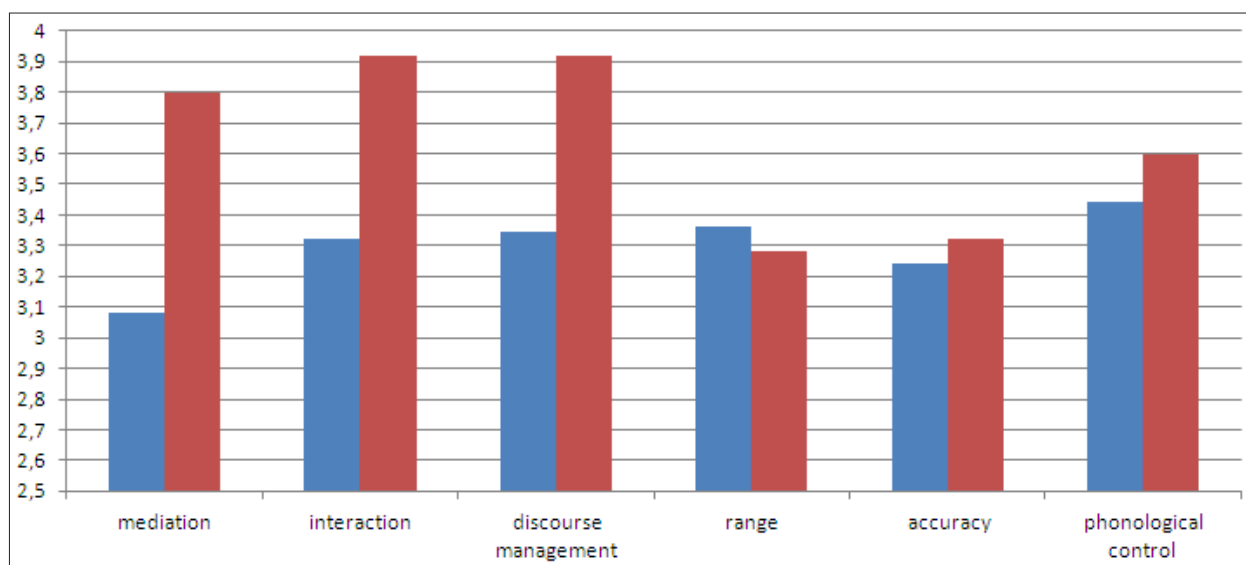
a) Experimental Group
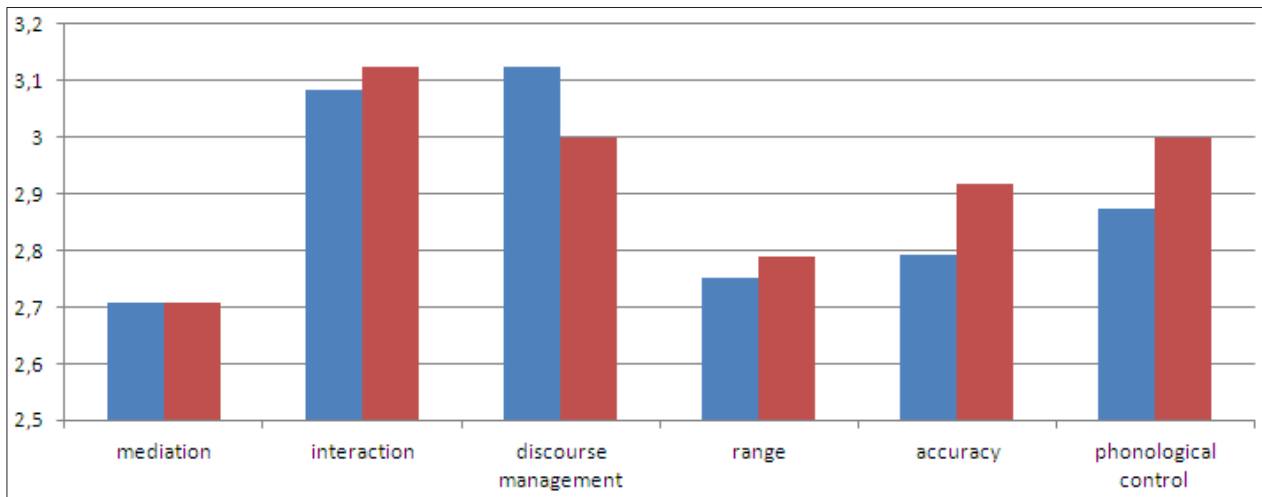
b) Control Group

Figure 1 presents the data of the diagnostic test (blue) versus the final test (red) in the two groups. The upper bar chart shows the results obtained from the group that had some additional practice with descriptors for mediation and the lower bar chart gives information about the group, which did not have this additional practice. From left to right we have twin bars of mediation, interaction, discourse management, range, accuracy, and phonological control. We can see that the performance of the group who worked with CEFR descriptors is slightly better than in the Control Group. This difference is quite small, but consistent. The difference is also stronger in relation to the three communicative criteria as opposed to the three linguistic criteria. These data may indicate the effectiveness of a set of exercises for the development of oral mediation skills in group discussion that was used in the Experimental Group.

6 Conclusion

The main conclusion is that the global achievement mark for mediation and the analytical marks are interrelated and we can support our analytical marks with the global mark for mediation and vice versa. To some extent, this approach can be regarded as *efficient* because it helps the assessor to self-check. The main implication of shifting from teaching communication to teaching mediation is the *increased focus on the collaborative development of new ideas*. By elaborating the concept of mediation and introducing mediation activities into the classroom, we facilitate passing on and receiving knowledge, and, most importantly, increase the autonomy of learners.

We realized that 'leading group work: encouraging conceptual talk' is a kind of activity that is often thought to be the responsibility of teachers, whereas the CEFR urges us to include it in students' repertoires, thus making them more independent. Working with CEFR descriptors can improve their social and collaborative skills.

Apart from these conclusions, some other interesting observations were made. For example, we noticed how mediation abilities develop with the progression of CEFR levels. At B2 level, students normally cannot grasp the totality of a complex abstract idea. Rather, they isolate two or three notions and explain them. At higher levels, students can mediate a concept in all its complexity as a whole. This could be a good indicator of students' level of language proficiency.

7 References

Association of Language Testers in Europe (ALTE). 2014. *The CEFR Grid for Speaking.* https://www.alte.org/resources/Documents/ALTE_CEFR_Speaking_grid_tests2014.pdf (accessed 6.11.2018).

Council of Europe. 2018. *The CEFR Companion Volume with new descriptors.* https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989 (accessed 6.11.2018).

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*, 2001. Strasbourg: Council of Europe. https://rm.coe.int/16802fc1bf (accessed 6.11.2018).

Hard Skills vs. Soft Skills: What's the Difference? https://www.thebalancecareers.com/hard-skills-vs-soft-skills-2063780 (accessed 6.11.2018).

North, Brian and Enrica Piccardo, 2016. *Developing illustrative descriptors of aspects of mediation for the Common European Framework of Reference (CEFR).* https://rm.coe.int/developing-illustrative-descriptors-of-aspects-of-mediation-for-the-co/1680713e2c (accessed 6.11.2018).

Pavlovskaya, Irina (ed.) 2017. *Актуальные вопросы языкового тестирования. [Modern issues of language testing].* Vol.2. St. Petersburg. http://hdl.handle.net/11701/7188 (accessed 08.04.2019).

8 Biographies

Prof. Irina Y. Pavlovskaya, Doctor of Philology works as a professor at St. Petersburg State University, Department of Foreign Languages and Linguodidactics. Her fields of research include language testing, language teaching methodology, phonetics and phonology, and phonosemantics. She has published more than 100 research and pedagogical articles. She supervised 13 successfully defended PhD papers (candidates' dissertations). She is a member of ILTA, EALTA, IATEFL (International Language Teaching Association), and the founding president of SPELTA (St. Petersburg English Language Teachers' Association).

Olga Y. Lankina of St. Petersburg State University holds a specialist degree in applied linguistics. She is a researcher, CLIL teacher and teacher of English for academic purposes at St. Petersburg State University, Russia. Olga is a Speaking examiner and examiner trainer. She is a member of EALTA. In 2013-2015 she was involved in the CEFR-linking project for the final test in English at St. Petersburg State University.