Formative peer assessment to enhance primary school pupils' oral skills: Comparison of written feedback without discussion or oral feedback during a discussion

Abstract:

In this study, we seek to understand which form of peer assessment is most effective in helping pupils progress in their speaking skills: when the peers discuss the feedback they are giving in subgroups or when they provide it in writing and individually. We also want to analyze what the students say, for what purpose, and what elements of oral communication skills they talk about when they provide feedback. The sample consisted of 42 pupils aged 10-11 years in two conditions (same school): written feedback condition and discussed oral feedback condition. Results indicated that the pupils in the discussed oral feedback condition progressed more than their peers in the written feedback condition. In addition, in the written feedback condition the majority of comments were mainly descriptive observations, with a few prescriptive comments. On the other hand, in the discussed oral feedback condition the comments were mostly suggestions.

Keywords.

peer assessment – oral skills – primary school

Introduction

When children start primary school, although they mostly do not yet know how to read and write, they often do already know how to understand the oral messages addressed to them. They also already know how to express themselves in everyday conversations. As a result, priority is given to teaching reading and writing skills (Dupont & Grandaty, 2016; Simard, Dufays, Dolz, & Garcia-Debanc, 2014). Consequently, oral communication occupies a functional place in the primary classroom and is not usually the subject of explicit teaching/learning (Gagnon, De Pietro, & Fisher, 2017).

In addition, research has shown that oral communication is often not an instructional priority for teachers because they are unsure about how to teach and evaluate it (Authors, 2017; Dumais & Lafontaine, 2011; Sénéchal & Chartrand, 2011). Providing feedback to students on oral skills can also be tedious (Authors, 2019; Gagnon, de Pietro, & Fisher, 2017). To ease this burden on teachers, involving the pupil in the assessment process has been suggested. In this way, the pupil's learning objectives can be achieved with the aid of the teacher and their peers.

Although peer assessment has been identified as an important lever for learning (Gielen, Peeters, Dochy, Onghena, & Struyven, 2010), there is a lack of research available on peer assessment in primary education and certainly on peer assessment of oral skills. Hence, which peer assessment methods are most effective for the development of oral skills remains unclear. This study innovates by examining two peer assessment modalities for improving primary students' oral skills, specifically their speaking skills.

In this study, we seek to understand which form of peer assessment is most effective in helping pupils progress in their speaking skills: when they receive peer-negotiated oral feedback, or when they receive individual written feedback.

Teaching and assessing oral communication skills

Oral communication includes actions related to both speaking and understanding the messages heard. In a school situation, oral communication can take on three distinct aspects (Plane, 2015). First, it can have the status of a medium; that is, it is spontaneously used by people to verbalize and exchange information and ideas (Le Cunff & Jourdain, 2008), such as during the correction of class work. Second, it can have the status of reflexive oral communication, mobilized when metacognition is explicitly called for (Veenman, 2012), such as when students are asked to identify the steps they went through to complete a task or when they need to describe the strategies they have used. Third, oral communication can be a teaching/learning object. In this case, time is devoted to it in the class schedule, as with other school disciplines. This third aspect with its accompanying challenge of teaching oral skills to students is our focus. It is also the aspect that poses a problem for teachers (Simard et al., 2014). In fact, in schools, oral communication is frequently considered a complex discipline to teach (Simard et al., 2014). Studies that highlight oral communication teaching practices are few, especially for elementary school (Lafontaine & Messier, 2009; Sénéchal, 2017; Sénéchal, Dumais, & Bergeron, 2019). However, many researchers have emphasized the value of spending time on oral communication development in the classroom (Authors, 2019 Bergeron, Dumais, Harvey, & Nolin, 2015; Dumais & Lafontaine, 2011; Gagnon & Dolz, 2016; Kendeou, White, van den Broek, & Lynch, 2009; Storch & Whitehurst, 2002). Primarily, oral communication allows us to take part in our daily interactions with others, contributes to the development of thinking (Muijs, Kyriakides, van der Werf, Creemers, Timperley, & Earl, 2014), and predicts reading comprehension (Bianco, 2015). Moreover, not mastering oral language is an obstacle to learning, because it becomes a source of misunderstandings and/or cognitive overload (Sweller, Ayres, & Kalyuga, 2011). More broadly, instruction in oral communication can support literacy development (Allen, 2018; Lafontaine, 2013). Additionally, mastery of oral communication

helps to combat academic failure and social exclusion (Le Cunff & Jourdain, 2008; Nonnon, 2016).

Genre-based teaching of oral communication has been advocated (Dolz & Gagnon, 2008; Freedman & Medway, 1994; Horverak, 2016; Hyland, 2003, 2007), and mainly the learning of genres that students can be prepared for in advance (Jaubert, 2007), such as a presentation, a life story or a weather report. Other genres of oral communication, on the other hand, are spontaneous and cannot be prepared in advance.

Despite this recommendation, the few existing studies show that when oral skills are worked on in the classroom, a limited variety of genres are addressed. Oral presentations, recitation, large group discussions, reading aloud and questionnaires (to evaluate listening skills) top the list of practices (Authors, 2019; Nolin, 2015). This is due to several factors: the generality of the school curriculum, which does not specify genres or oral teaching objects in school curricula (Dumais, 2012; Lafontaine & Préfontaine, 2007); the scarcity of available instruction materials (Garcia-Debanc & Delcambre, 2001-2002; Sénéchal, 2017); and the status of oral skill as a general, transversal skill, mobilized in all disciplines and in class activities, and therefore not addressed instructionally for its own sake (Simard et al., 2014, p. 288). In addition, assessing oral skills is a challenge for teachers (Dumais, Lafontaine, & Pharand, 2017). They "have difficulty, resistance, in carrying out certified, and even formative, evaluations of pupils' oral productions" (de Pietro, Gagnon, & Fisher, 2017, p. 16).

In assessing oral communication skills, three types of capacities must be considered (Dolz, Pasquier, & Bronckart, 1993). First, there is the capacity for action. This refers to being able to adapt one’s speech according to the communication situation, which involves identifying the elements specific to different communication situations. Second, there is discourse capacity, which has two aspects. To communicate orally, one must elaborate upon ideas (Putman & Borko, 1997) and organize one's message correctly by ensuring its coherence (Dougan, 1996).

Third, there is linguistic-discursive capacity. This refers to the grammatical and lexical aspects that support text-generating processes (Bronckart & Dolz, 1999). It also includes verbal communication (volume, flow, articulation, and intonation of the voice), and non-verbal communication (eye contact, body posture, occupying space, use of a variety of supports to supplement one's speech) (Morreale et al., 1993). Studies have shown that rubrics are appropriate to assess oral skills (Dunbar, Brooks, & Kubicka-Miller, 2006). They allow evaluators to use an analytical evaluation system in which each component is rated individually and provide a valid judgment of the performance (Jonsson & Svingby, 2007; Lavoie & Bouchard, 2017; Pomplun, Capps, & Sundbye, 1998).

Formative assessment of oral communication: choosing peer assessment

Several forms of evaluation-related decisions have been identified by Allal (1991) and De Ketele (2010): "evaluate to guide new action to be taken", "evaluate to socially certify an action considered complete" and "evaluate to improve an ongoing action" (De Ketele, 2010, p. 26). The last of these refers to formative evaluation or assessment for learning (Earl, 2013). This type of evaluation intervenes during learning and regulates learning (De Ketele, 2010). Regulation through assessment is no longer considered as a particular event in learning, but as an integral part of it (Allal & Mottier Lopez, 2005). In this perspective, assessment and learning "are no longer separated, just as teachers and students become partners" (Younes & Gaime, 2012, p. 2).

This seems to us to be essential in the context of oral communication teaching/learning, because oral skills, much more than in any other discipline, are inseparable from the person: the body and voice are involved in all speaking (Garcia-Debanc, 1999). To evaluate oral communication, it is therefore necessary to use practices that are effective (Ko, Sammons, & Bakkum, 2014),

caring, and safe for pupils. Otherwise, there is a great risk that they will no longer dare to express themselves.

Peer assessment involves two or more students in the assessment of their respective acquisitions, approaches, progress, and/or difficulties (Gielen, Dochy, & Onghena, 2011; Wegmuller & Allal, 1997). Peer assessment can take three forms (Lafontaine, 2007): (1) normative, where peers assign a numerical value to the performance of the person being assessed; (2) qualitative, where students assess the work done through a commentary; and (3) descriptive, where some students observe others using specific criteria to identify their strengths and weaknesses. Descriptive comments are considered the most useful, because they are not judgmental and lead those receiving the feedback to improve (Brookhart, 2008). This form of peer assessment is the most interesting for us in the context of the formative assessment of oral communication, because the feedback provided by peers focuses on specific aspects. However, doing this type of evaluation is not self-evident and requires learning (Authors, 2017; Paulus, 1999). In their study, Leenknecht and Prins (2018) investigated whether the involvement of primary school students in defining assessment criteria and standards resulted in better peer assessment and a better style of feedback. Their results showed that it did. Notably, students who discussed the criteria and standards before giving feedback gave more positive and effective feedback than others. In addition, their study suggested that primary school students are able to develop explicit knowledge about standards and criteria for assessment.

Peer assessment brings several benefits. First, there are cognitive benefits. Indeed, peer assessment amplifies learning through two mechanisms: when one student evaluates another student, it allows him or her (the student doing the evaluation) to better understand the objectives to be achieved; feedback from a peer is given in his or her own words with suggestions and sharing of ideas, which creates a regulatory effect (Allal & Mottier Lopez,

2005; Hattie & Timperley, 2007). In this way, the person receiving the feedback will act upon it, and the person giving the feedback will also regulate his/her own performance based on the feedback that was given (Authors, 2018a). Second, there are metacognitive benefits. Indeed, students who evaluate must identify problems and make suggestions. This has been found to increase students' own skill level, because providing feedback increases "reflection and generalization to new situations, promoting self-assessment and greater metacognitive self-awareness" (Topping, 1998, p. 256). The ability to ask questions (both about the work as a whole, and about specific aspects) is being developed (Bostock, 2000; Lu & Law, 2012). Indeed, before being able to evaluate others effectively, an awareness of one's own mental processes is necessary. This leads to self-esteem and self-confidence. Third, there are social benefits. Students interact with each other and discuss their evaluation. This can develop teamwork skills. It can also develop verbal communication skills, such as sharing views, making or rejecting suggestions, etc. (Topping, 1998; van Gennip, Segers, & Tillema, 2010). Fourth, there are logistic benefits. Peer assessment saves teachers' time (Boud, 1989).

According to Rodet (2000), feedback can have three different natures: informative, making suggestions or prescriptive. First of all, feedback can be information-giving. This is a matter of presenting findings, highlighting what is correct, and the like. Second, feedback may offer suggestions. This is a matter of formulating advice and proposals so that the person can improve. Third, prescriptive feedback gives actions to be taken or aspects to be addressed, often starting with "it is necessary" or "it should be". Rodet (2000) explained that "comments making suggestions seem to us to be preferred to informative comments, which themselves would be favored over prescriptive information" (p. 63, free translation).

In addition, four categories of content can be used in feedback (Hattie & Timperley, 2007; Rodet, 2000): (1) cognitive feedback about knowledge control and conceptual errors, which

relates to the task, and indicates to the student where he or she stands in the accomplishment of the task (task level); (2) methodological feedback, which refers to the use or not of effective strategies and is related to the processing of the task (process level); (3) metacognitive feedback, which places the learner's knowledge in a broader context (the course, the training program) and aims to develop the learner's control and autonomy (self-regulation level); and (4) emotional feedback, which refers to the person's state of mind (such as a feeling, emotion, feeling, mood) (self-level). Brookhart (2008) explained that feedback about the person is "generally not a good idea" (p. 21).

Feedback can be given either in writing or orally. Brookhart (2008) explained that written feedback is well suited for formal assignments, while oral feedback is suitable for formal as well as informal assignments. While some studies have compared the effectiveness of written and oral peer feedback, they have paid little attention to the feedback that can be given in the context of oral instruction. In addition, the participants in such studies have been students in higher education. Thus, in the context of writing activities, Suh (2005) compared the effects of oral and written feedback among L2 students at Korea University. For oral feedback, the teacher and peers were present and discussed in a talking circle. The written feedback was given by the peers. The results showed that for the 6 students whose data were analyzed, both types of feedback resulted in improved writing skills. Both also supported the promotion of positive attitudes among students and the development of social skills. Nevertheless, students improved more with oral feedback. The oral format allowed for more feedback than the written format, since several students gave their opinion at the same time, which was not the case with the written feedback. Moreover, the study showed differences between the subjects addressed in the two types of feedback. Thus, in their written feedback, the students focused on grammar-related aspects. Oral feedback focused on aspects related to the meaning, content and organization/coherence of the text. Sheen (2010) also drew attention to the fact that written

feedback, again in the context of a writing activity, is more complex for students to develop than oral feedback.

In their meta-analysis on the effectiveness of feedback, Biber, Nekrasova, and Horn (2011) collected a total of 306 articles. They considered the effectiveness of feedback for writing development; only 10 of those studies involved students aged 10-12. Their overall results indicated that oral feedback was more effective in these studies than written feedback. Students learning in their first language made significant gains in writing proficiency through oral feedback. This was not the case for written feedback.

In another context, van den Berg, Admiraal, and Pilot (2006) compared the written and oral feedback that students gave in seven different courses in the discipline of history. The study showed that students focused on different aspects in their oral as compared to their written feedback. Indeed, most of the students' written comments were product-oriented; students did not explain their evaluation comments, and did not offer many suggestions for improvement. Moreover, when the teacher provided an evaluation grid, students used it as a mandatory checklist to structure their written comments. In contrast, oral feedback was less product-oriented. Students explained their point of view and gave more advice. Sometimes, however, comments focused only on one specific aspect rather than on the different criteria. The authors concluded that "for PA [peer assessment] to yield the most complete feedback, a combination of written and oral feedback is essential" (p. 146).

This study

In this experimental study, we tested the implementation of two different peer assessment modalities during an instructional program in which students worked on their oral skills: a written feedback condition and a discussed oral feedback condition. Our ambition was to

determine whether one of these two modalities is more appropriate for the peer assessment of speaking skills. In the written feedback condition, feedback is given individually, in writing, and from one person to another, so that it remains intimate. The other way to do the evaluation is to let students discuss the oral performance in subgroups, with the speaker present. In this case, the discussion can take place directly, but the person being evaluated is in front of several others, who give their feedback.

We thus formulated two research questions: (1) how does the difference in peer assessment modalities affect the nature of the feedback produced? and (2) how does the peer assessment modality affect students' development of oral communication skills?

Method

Participants

The sample consisted of 42 pupils aged 10-11 years from the same school (a public school in the French-speaking community of Belgium), located in a well-off socio-cultural environment. The written feedback condition included 21 students (12 girls and 9 boys); and the discussed oral feedback group was composed of 21 students (12 girls and 9 boys). The experiment was conducted independently in the two groups. Consents were obtained through a specific request to the parents from the two classes in the sample.

In order to ensure that students were at the same level of oral language skills initially, we conducted a pre-test. Each student gave an oral performance alone, outside the classroom, for the assessor, who was the same researcher for all students. The instructions were: Speak for one minute about a familiar word drawn at random. Four elements of the oral performance were

assessed, the same as during the instructional program : communication intention, idea development, grammatical aspects, verbal communication[1]. The *t*-test revealed no significant differences between the two conditions (see Table 1). Effect sizes (Cohen's *d*) were calculated and interpreted following Cohen's (1988) cut-offs: .2, .5 and .8, which were considered small, medium and large effect sizes, respectively. The effect sizes in Table 1 are thus negligible (< 0.20) or small.

Table 1 – Descriptive statistics and comparison of initial pretest scores for elements of oral communication per condition (written feedback condition and discussed oral feedback condition)

	written feedback condition $n = 21$		discussed oral feedback condition $n = 21$		$t(40)$	p	Cohen's d
	M	*SD*	*M*	*SD*			
Communication intention (/3)	2.05	.23	2.05	.22	-.071	.944	.02
Idea development (/12)	5.21	2.20	4.57	1.57	-1.066	.293	.34
Grammatical aspects (/6)	1.68	1.57	1.43	.98	-.626	.535	.20
Verbal communication (/12)	9.63	1.54	9.38	1.43	-.534	.596	.17

Instructional program

To answer our research questions, we used the "itineraries" method of instruction for oral communications skills (Authors, 2015, 2018b, 2019). In this method, students give several oral versions of a speech, "reoralisation" (Authors, 2017, p. 188), to improve their mastery of a

[1] Two items were not assessed in the pre-test in connection with message organization and non-verbal communication. General organization was not assessed because the instruction was to speak for one minute on a familiar topic. There was no particular structure expected. Students could describe, narrate, make a riddle, use dialogue, etc. Non-verbal communication was not assessed in the pre-test to limit student stress. This was the first time that students had been asked to give a performance like this with an unfamiliar researcher. Adding a camera, to make sure that the criteria could be measured correctly, seemed possibly confusing for the students. We wanted to avoid the risk that they would also get confused during the rest of the instructional program.

chosen discursive genre (see Figure 1). They express themselves orally several times, and progress from one version to another. Interventions to support them and help them overcome obstacles are planned between these oral productions. The following occur in alternation: teacher scaffolding (focused on the knowledge necessary for the production of the genre in question), and feedback from peers. In addition, in this method, students are asked metacognitive questions (see Authors, 2016, 2019, submitted). They are posed by the teacher before, during and after the tasks. Examples of such questions are: What is your personal goal for this task?; When you encountered a difficulty, what did you do?; If you do the same activity again tomorrow, what strategies will you use to be more effective?; What are the reasons for this?

Students in both conditions therefore experienced the same oral teaching method, used by the same teacher (the children' regular instructor), with the only variable being the feedback modality. More concretely, the students experienced two consecutive tasks over a period of three months. The first task was completed before the Easter holidays. The second task followed the first, but was separated by a two-week vacation period. Each task took an average of 5 weeks, at a rate of 2 sessions of 55 minutes of work per week, which corresponded to 550 minutes of total time. The two tasks focused on the life story genre.

To launch the task, there was a project setup time. The teacher defined the task. Then, collectively, the parameters of the communication situation were clarified by examining a series of questions (for example, what is the speaking task, what is the purpose of the speech, who is the message intended for, when and under what conditions will the message be transmitted). A summary poster was produced. In the first task, a report of one’s day was worked on. The scaffolding provided by the teacher was related to the chronology of events (including adverbs),

and intonation. In the second task, the expression of personal wishes was developed. The scaffolding was devoted to the structure of this type of message and eye contact.

Pupils had two occasions per task to evaluate their peers' oral performances, for an equal amount of time (25 minutes) in each condition. Indeed, the pupils gave feedback to their peers four times in total, for a total time of 100 minutes. To help them do it, a teaching activity on feedback was included in the itineraries method. The teacher played a soundtrack for the students (speeches of the same genre as the one they were to produce). Together, the students had to find positive points and areas for improvement. This made it possible to develop a grid for criteria to talk or to write about when giving feedback to others. Students were given this grid to use to assess their peers (Appendix 1). The grid is composed of several aspects. First, students were asked to give a title to their comments (a kind of overall evaluation). Second, a series of elements were used to analyze the content of the message. Third, criteria for the organization and coherence of the discourse were present (introduction, development of the text with several personal events explained in chronological order, use of time connectors). Fourth, there was space for a few words about the overall quality of the message. Fifth, specific aspects of the oral message (intonation, eye contact) were considered. Finally, students could also make other comments, related to other aspects (directly related to tasks 1 or 2 and often related to the scaffoldings, such as the use of adverbs, for example).

This grid was provided prior to the first peer assessment moment. The students were able to ask any questions they had about the grid. They were also able to include other criteria that could be looked at, in addition to those in the grid provided by the teacher. This grid was then used during the 4 peer assessment moments.

In both conditions, each student received feedback from three or four peers. In the written feedback condition, students received feedback from every member of their subgroup. During the planned 25 minutes, each student, therefore, assessed two or three other students, depending on the number of students in their subgroup. The groups were initially chosen by the students and did not change throughout the work. Pupils wrote their comments individually, and then forwarded them to the evaluated student. This practice did not lead to discussion between the different members of the group, or between the pupils' assessors.

In the discussed oral feedback condition, students also worked in subgroups of 3-4 students. Each member received feedback from the other subgroup members on their oral performance. The members exchanged their views with the student being evaluated. The student was therefore present during the discussion, and could intervene if desired. The group spent about 6 minutes evaluating each member of the group. The students did not prepare their comments before the meeting. Comments were made only orally.

Instruments and procedure

Collecting students' feedback. To provide insight into whether the aspects on which students provided feedback to their peers differed according to condition (research question 1), for the discussed oral feedback condition, we recorded and transcribed the discussions during which the subgroups discussed what feedback to give to their peers. This corresponded to 415 minutes of recordings and 16,417 words. For the written feedback condition, we kept all of the students' written comments, also transcribe. The total comments amounted to 1,463 words.

Measuring student progress. To measure their progress, the pupils took four tests (research question 2): before (test 1 and test 3) and after (test 2 and test 4) each of the two tasks (see Figure 1).

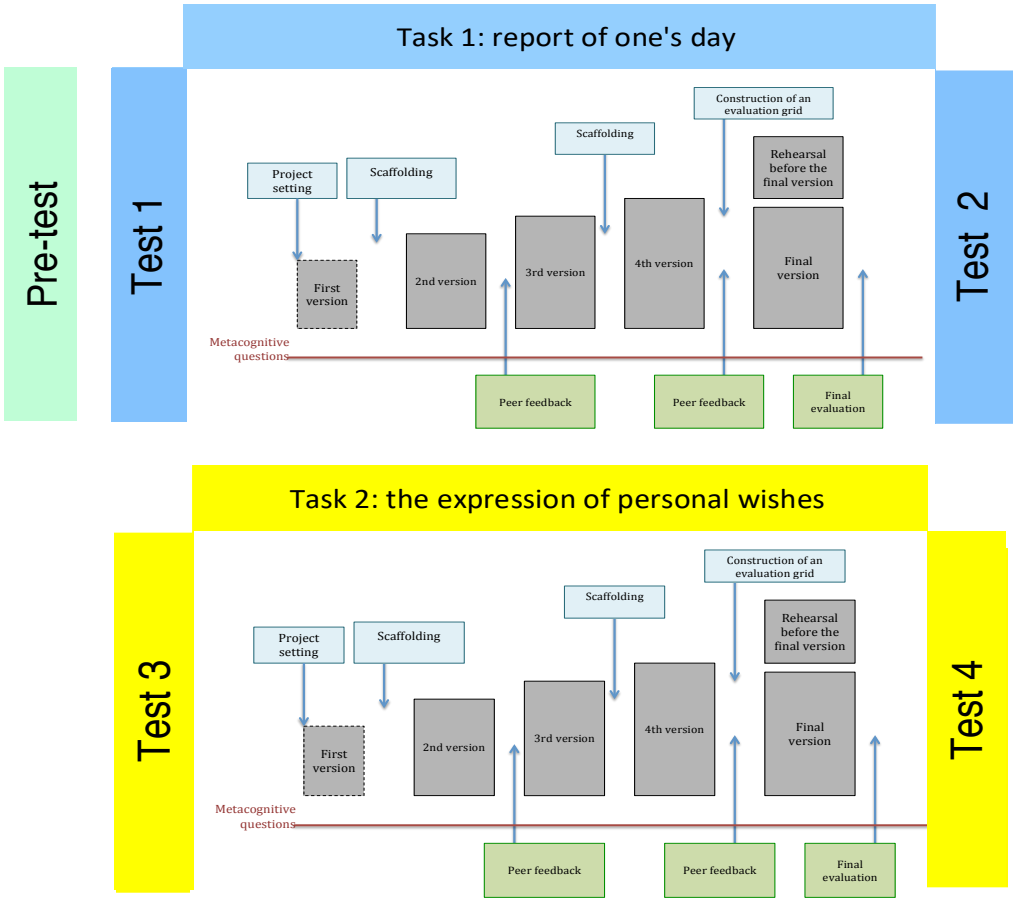


Figure 1 - Schematization of the instructional program

For each test, the pupils gave an individual oral performance based on instructions related to the life story genre. This was done outside the classroom with the researcher as the only audience. The preparation time was the same for each student (15 minutes for each preparation session). We checked that the students did not discuss this among themselves at the time, so that they would not influence each other. All of the pupils' oral performances were filmed and the content was transcribed. For test 1, students had to present on their last Monday, while for test 2, they presented on their last Thursday. For test 3, they were asked to express their personal wishes regarding objects to be acquired. For the final test, they had to express their wishes regarding the adoption of an animal (imaginary or real). Examples of student performances on test 2 and test 4 are shown in Appendix 2.

In total, 168 performances were analyzed using an evaluation matrix. This evaluation matrix included the six elements of oral communication related to language capacities (Authors, in press; Dolz et al., 1993): communication intention, idea development, message organization, grammatical aspects, verbal communication and non-verbal communication. Table 2 shows the details of the criteria used for each of the aspects and the weighting used.

Table 2 - Criteria used to evaluate students' oral communication

Aspects of oral communication	Criteria	How to evaluate
Adapt one's speech according to the communication situation (i.e., communication intention) /3	- compliance with the instructions - meeting the time limit - adapted language register	0 or 1 was assigned for each criterion.
Idea development /12	- number of sentences - number of words used - number of ideas - number of words per idea	Count. The results of the count were transformed into a Likert scale (see Appendix 2 for weightings).
Organization and coherence of the message /21	- introduction - development of actions - personal opinion - consistency in speech (chronology leading to a conclusion) - expressions using "I" all through the text	A rating from 0 to 3 was assigned to each criterion (0 = no, 1 = partial, 2 = yes, 3 = excellent[2]).
	- number of time connectors - number of different connectors[3]	Count. The results of the count were also transformed into a Likert scale (see Appendix 3).
Grammatical aspects /6	- number of syntactically correct sentences - number of adverbs used to qualify or reinforce statements[4]	Count. The results of the count were also transformed into a Likert scale (see Appendix 3).

[2] 2: yes, satisfactorily; 3: yes, remarkably (exceeding the expected requirements).

[3] Evaluated because it was taught in the scaffolding provided for in the itinerary method.

[4] Measured because they were supported in the first itinerary task.

Verbal communication /12	- volume - speech rate - voice articulation - intonation	A rating from 0 to 3 was assigned to each criterion (0 = no, 1 = partial, 2 = yes, 3 = excellent)
Non-verbal communication /6	- eye contact - body posture	A rating from 0 to 3 was assigned to each criterion (0 = no, 1 = partial, 2 = yes, 3 = excellent)

For the components where a count was performed, the score for each student was calculated using a weighting that took into account the minimum and maximum observed scores for each item (in the data set for each item; see Appendix 3). Based on the minimum and maximum marks obtained by all students in their oral performances, four levels of performance were established. This was done separately for each of the criteria. For example, regarding the number of sentences, the minimum number of sentences produced across all students was 3, while the maximum was 31. The 4 levels were based on these results. In order to avoid decimal values that did not allow us to easily classify a student's score in a level, we established ranges. Thus, for the number of sentences, level zero included students who produced between 0 and 8 sentences, level one included those who produced between 9 and 16 sentences, level two included those who produced between 17 and 24 sentences, and level four included those who produced between 25 and 32 sentences.

Data analysis

Feedback data. A content analysis was applied to the collected data. The steps identified by L'Ecuyer (1990) were used to analyze the qualitative data, using a mixed model (allowing categorization from theoretical frameworks, and emerging categories). The data were first divided into units of meaning by two researchers, who agreed on this division.

For the written feedback condition, a total of 1463 words were identified, which corresponded to 541 units of meaning. For the discussed oral feedback condition, 16417 words were counted, corresponding to 1368 units of meaning.

First, in both conditions, the units related to feedback about the person were identified and removed (200 units for the written condition and 212 for the oral condition), in alignment with the work of Brookhart (2008). In addition, for the discussed oral feedback condition, times when students were discussing other things were omitted from the analysis here (e.g., discussion about the meaning of a word or irrelevant aspects). This corresponded to 287 units of meaning. In the end, 311 units of meaning for the written feedback condition and 860 for the discussed oral feedback condition were kept. These remaining units underwent two types of analysis.

The first analysis was based on the nature of the feedback given (Rodet, 2000): informative feedback, making suggestions and prescriptive feedback. For this purpose, a chi-square test was used to determine whether the observed differences were significant.

The second analysis was based on what the students were discussing. The units listed for the different types of feedback were classified according to the six elements of oral communication related to language capacities (Authors, in press; Dolz et al., 1993), specified in the previous section and in Table 2: communication intention, idea development, organization and coherence of the message, grammatical aspects, verbal communication and non-verbal communication.

Numerical data. For the pre-test, both of the assessors evaluated all students using the observation grid. Following this, inter-rater reliability was calculated. Some criteria had moderate to low agreement. Landis and Koch (1977) established a satisfactory level of agreement as a Cohen's kappa above 0.61 and an ICC above 0.70. The disagreements were

related to the complexity of assessing oral language, as for some aspects, one of the assessors seemed to be more demanding than the other. The assessors therefore reviewed these items together for each student and negotiated to obtain an acceptable level of agreement. When the two assessors disagreed on students' performance, they listened to the recordings together again and after discussion, they agreed on the criteria. These criteria were then evaluated again. When inter-rater reliability was established (Kappa = 0.61 and ICC = 0.7), the lead rater assessed all tests for all students. As a result, inter-rater reliability was ensured for each of the criteria. The inter-rater reliability results obtained for each criterion are presented in Appendix 4. For the subsequent tests, although one researcher evaluated all recordings, a second researcher independently evaluated one third of these recordings (i.e., 8 students per class) to check for potential bias. The interrater reliability remained satisfactory for the subsequent tests.

The measurements thus obtained were processed using SPSS statistical software. A *t*-test was used to compare the averages for each task (test 1/test 2 and test 3/test 4), and to compare the scores on the test completed at the beginning of the instructional program (test 1) with the scores on the final test (test 4). In addition, following the recommendations of Green and Salkind (2017), an ANCOVA was carried out to examine the development of students' oral skills in relation to their condition. Partial η^2 effect sizes were interpreted according to the cut-offs presented by Green and Salkind (2017), with values larger than .01, .06 and .14 being interpreted as small, moderate and large, respectively. The results are presented in the following section.

Results

The results are presented according to the two research questions.

Question 1: How does the difference in feedback modality affect the kinds of feedback produced?

The ways of giving feedback differed from one condition to the other. The nature of the feedback given aligned well with the categories of Rodet (2000). The proportions of comments in each category (in counts and as percentages) are shown in Table 3. The χ^2 test performed shows that the difference is significant between the two conditions: $\chi^2(2) = 54.51$; $p < 0.001$. It can therefore be deduced that peers in each condition did not offer the same nature of feedback to the students.

Table 3 - Distribution of feedback categories (counts and percentages): observed numbers and expected theoretical numbers

		Informative feedback	Making suggestions	Prescriptive feedback	Total
Written feedback condition (n = 21)	Observed	287 **(92%)**	11 **(4%)**	13 **(4%)**	311 (100%)
	Expected	243 (78%)	48 (15.5%)	19 (6.5%)	
Discussed oral feedback condition (n = 21)	Observed	627 **(73%)**	173 **(20%)**	60 **(7%)**	860 (100%)
	Expected	671 (78%)	133 (15.5%)	54 (6.5%)	

First of all, in both classes, students formulated informative feedback (Rodet, 2000) with more or less the same frequency. In the written feedback condition (WFC), 92% of the comments were informative feedback, compared to 73% in the discussed oral feedback condition (DFC).

Most of the time, pupils made observations and they used "you" in most comments, as the examples below illustrate.

*Your structure is okay. (*Téo[5], WFC)

You put some connectors in your message. (Lise, WFC)

You're still looking too much at your paper. (Emile, WFC)

You often repeat the same connectors. (Amine, DFC)

In addition, in the discussed oral feedback condition, there were more comments that were making suggestions (Rodet, 2000), about 20% of the total, while only 4% of the comments in the written feedback condition provided this nature of feedback. That is an important difference. The following elements were present in these comments: action verbs, examples, tips, comparisons with previous products, highlighting of particular qualities, and so forth, as the examples below illustrate.

You can stop at periods, commas (Zoé, WFC)

Try to put an introduction. (Ben, DFC)

Good structure but try to put them in chronological order. (Phil, DFC)

Try to speak up. Imagine you're in front of fifty people. (Amy, DFC)

To improve your fluency you can repeat some words several times. The ones you find difficult to pronounce. (Tess, DFC)

Prescriptive feedback (Rodet, 2000) was present in both conditions, too: 7% of the comments in the written feedback condition and 4% in the discussed oral feedback condition. Examples of such comments are:

It is necessary to articulate! (Nicolas, WFC)

[5] The students' names have been changed to ensure anonymity.

You must look at the audience more often. (John, DFC)

Furthermore, it should be noted that in the discussed oral feedback condition, students did not mind having a face-to-face discussion about their oral performance with their peers. At the beginning of the experiment, their interjections were not very frequent. The assessed students rather took a listening posture. Thus, over a 6-minute discussion, the evaluated student spoke on average once during the first exchange to show his or her agreement. After they did so, they interjected a little more often (three times on average during the second and third discussions, four times on average during the last exchange). They mentioned their agreement or asked for clarification. They did not question the feedback given.

In short, in the written feedback condition, students gave essentially informative feedback. They also gave more positive feedback to encourage their peer. In the discussed oral feedback condition, students also make suggestions to help peers to improve.

The second content analysis was carried out with the aspects of oral communication as a filter. Table 4 shows the number of sense units (and percentages) listed for these six elements, relative to the three nature of feedback used.

Table 4 - Number of sense units (percentages) for the six aspects of oral communication by the nature of the feedback, per condition.

		Informative feedback	Making suggestions	Prescriptive feedback	TOTAL
Written feedback condition	Communication intention	0	0	0	0
	Idea development	46 (15%)	2	2	50 (16%)
	Organization and coherence of the message	153 (49%)	1	1	155 (50%)
	Grammatical aspects	14 (5%)	1	3 (1%)	18 (6%)

	Verbal communication	44 (14%)	7 (2%)	6 (2%)	57 (18%)
	Non-verbal communication	30 (10%)	1	0	31 (10%)
	TOTAL	287 **(92 %)**	12 **(4%)**	12 **(4%)**	311 (100%)
Discussed oral feedback condition	Communication intention	0	0	0	0
	Idea development	103 (12%)	17 (2%)	8 (1%)	128 (15%)
	Organization and coherence of the message	327 (38%)	95 (11%)	8 (1%)	430 (50%)
	Grammatical aspects	34 (4%)	27 (3%)	26 (3%)	87 (10%)
	Verbal communication	94 (11%)	17 (2%)	9 (1%)	120 (14%)
	Non-verbal communication	69 (8%)	17 (2%)	9 (1%)	95 (11%)
	TOTAL	627 **(73%)**	173 **(20%)**	60 **(7%)**	860 (100%)

The analysis showed that the elements of oral communication addressed were more or less the same in both conditions. Students tended to focus their feedback on what was learned from the scaffolding. Thus, the majority of students in both conditions dealt with the general organization of the text and the use of temporal connectors to structure the discourse (50% in both conditions). Idea development was also approached about as often by students in both conditions (16% and 15%), as was non-verbal communication (10% and 11%). Verbal communication was addressed more in the written feedback condition (18%) than in the discussed oral feedback condition (11%). The opposite is observed for grammatical aspects, which were addressed more in the discussed oral feedback condition (10%) than in the written feedback condition (6%). Students did not address the intention to communicate. There were no comments that consider the communication instructions, nor did any comments provide feedback on the adaptation of the message to the interlocutor. This limitation may be due to the fact that the grid used by the students did not have all the parameters present in the teacher/researcher grid.

Question 2: How does feedback modality affect students' development of oral communication skills?

The progress of students in the written feedback condition

Table 5 presents the oral skills pretest-posttest results for the written feedback condition.

Table 5 – Repeated measures *t*-test results, descriptive statistics and effect sizes for observed variables in the written feedback condition for tests before and after each task ($n = 21$)

	Task 1				Task 2				Between tests 1 and 4	
	Test 1	Test 2	*t*(20)	Cohen's *d*	Test 3	Test 4	*t*(20)	Cohen's *d*	*t*(20)	Cohen's *d*
	M (SD)	*M (SD)*			*M (SD)*	*M (SD)*				
Communication intention (/3)	2.33 (.48)	2.43 (.51)	-.81	.25	2.33 (.57)	2.52 (.51)	-1.16	.36	-1.71	.53
Idea development (/12)	3.95 (1.80)	5.19 (1.94)	-3.04**	.94	4.05 (1.56)	3.81 (2.04)	.44	.14	.25	.08
Organization and coherence of the message (/21)	8.33 (1.50)	12.29 (3.07)	-5.50***	1.70	10.05 (1.75)	11.62 (2.54)	-2.66*	.82	-5.29***	1.63
Grammatical aspects (/6)	1.71 (1.35)	2.57 (1.08)	-2.69**	.83	2.38 (1.28)	1.86 (1.39)	1.56	.48	-0.41	.13
Verbal communication (/12)	9.76 (1.18)	11.71 (.56)	-6.77***	2.09	11.43 (.68)	11.00 (1.00)	2.63*	.81	-4.81***	1.49
Non-verbal communication (/6)	3.33 (.66)	3.86 (.96)	-2.06	.63	3.90 (1.00)	4.48 (1.03)	-2.25*	.69	-4.73***	1.46

* $p < .05$
** $p < .01$
*** $p < .001$

In terms of progress between the beginning of the instructional program (test 1) and its end (test 4), three elements evidently improved and show large effect sizes. The organization and coherence of the message improved from an average of 8.33 out of 21 ($SD = 1.50$) at test 1 to 11.62 ($SD = 2.54$) at test 4 ($t(20) = -5.29$, $p < .001$, $d = 1.63$). The mean for verbal communication improved as well, from 9.76 out of 12 ($SD = 1.18$) at test 1 to 11.00 ($SD = 1.00$) at test 4 ($t(20) = -4.81$, $p < .001$, $d = 1.49$). The mean for non-verbal communication also

improved, from 3.33 out of 6 (*SD* = .66) at test 1 to 4.48 (*SD* = 1.03) at test 4 ($t(20) = -4.73$, $p < .001$, $d = 1.46$).

For the first task (test 1/test 2), four elements out of the six measured showed significant progress and a large effect size. Idea development went from an average score of 3.95 out of 12 (*SD* = 1.80) at test 1 to 5.19 (*SD* = 1.94) at test 2 ($t(20) = -3.04$, $p < .01$, $d = .94$). The mean score for message organization and coherence showed a large improvement as well, from 8.33 out of 21 (*SD* = 1.50) at test 1 to 12.29 (*SD* = 3.07) at test 2 ($t(20) = -5.50$, $p < .001$, $d = 1.70$). Grammatical aspects went from an average score of 1.71 out of 6 (*SD* = 1.35) at test 1 to 2.57 (*SD* = 1.08) at test 2 ($t(20) = -2.69$, $p < .01$, $d = .83$). The mean for verbal communication improved from 9.76 out of 12 (*SD* = 1.18) at test 1 to 11.71 (*SD* = .56) at test 2 ($t(20) = -6.77$, $p < .001$, $d = 2.09$). For idea development, data analysis showed that mean scores improved for three out of the four criteria in this category. Indeed, the average number of sentences used in the speech was 13.29 (*SD* = 4.04) at test 1, while at test 2, the average was 18.05 (*SD* = 5.55), ($t(20) = 4.22$, $p < .001$). The average number of words used also increased from 110.81 (*SD* = 52.43) at test 1 to 137.57 at test 2 (*SD* = 41.93), $t(20) = 2.88$, $p = .009$. The average number of ideas increased as well from 11.67 (*SD* = 2.76) at test 1 to 13.86 (*SD* = 3.12) at test 2 ($t(20) = 2.83$, $p = .01$). For message organization, progress was due to an increase in the mean number of connectors of different tests from 5.52 (*SD* = 1.91) to 8.10 (*SD* = 3.15), ($t(20) = 4.75$, $p < .001$). Grammatical aspects also showed improvement, due to an increase in the mean number of correct sentences from 11.19 (*SD* = 3.93) to 17.33 (*SD* = 5.09), ($t(20) = 5.30$, $p < .001$). Verbal communication also improved significantly on all four competencies in this category, although the one concerning intonation improved significantly more than the others: only five students scored 3 points (the maximum) at test 1, compared to 19 students at test 2.

For the second task (test 3/test 4), two elements out of the six measured showed significant progress. For message organization, the average score increased from 10.05 out of 21 (*SD* = 1.75) to 11.62 (*SD* = 2.54), ($t(20) = -2.66$, $p < .05$, $d = .82$), which is considered a large effect. And for non-verbal communication, there was a medium-sized increase, from 3.90 out of 6 (*SD* = 1.00) to 4.48 (*SD* = 1.03), ($t(20) = -2.25$, $p < .001$, $d = .69$).

Progress in the organization and coherence of the message can be explained by improvement on the criterion evaluating the number of time connectors, with an increase from an average of 5.52 (*SD* = 1.91) to 8.10 (*SD* = 3.15), ($t(20) = 4.75$, $p < .001$). For non-verbal communication, the criterion that improved considerably was eye contact. There was a significant and large decrease for the verbal communication aspect, from 11.43 out of 12 (*SD* =.68) to 11.00 (*SD* = 1.00), ($t(20) = 2.63$, $p < .05$, $d = .81$).

In the end, significant progress was identified for five out of the six elements on at least one of the two tasks. The aspect of adaptation to the communication situation did not improve, as it was already good from the beginning. This can be related to the instructional scheme itself, in which the teacher and the students together began by specifying the parameters of the communication situation, before the students developed their first version.

The progress of the pupils in the discussed oral feedback condition

Table 6 presents the oral skills pretest-posttest results for the discussed oral feedback condition.

Table 6 - Repeated measures *t*-test results, descriptive statistics and effect size for observed variables in the discussed oral feedback condition for tests before and after each task ($n = 21$)

	Task 1				Task 2				Between tests 1 and 4	
	Test 1 *M (SD)*	Test 2 *M (SD)*	*t*(20)	Cohen's *d*	Test 3 *M (SD)*	Test 4 *M (SD)*	*t*(20)	Cohen's *d*	*t*(20)	Cohen's *d*

Communication intention (/3)	2.43 (.51)	2.67 (.48)	-1.56	.48	2.62 (.50)	2.48 (.51)	1.14	.35	-.37	.11
Idea development (/12)	4.29 (2.55)	5.62 (2.84)	-1.86	.57	4.24 (2.34)	4.48 (1.72)	-.45	.14	-.35	.11
Organization and coherence of the message (/21)	8.19 (2.21)	12.29 (2.57)	-6.37***	1.96	9.10 (2.26)	13.38 (2.44)	-6.69***	2.06	-8.08***	2.49
Grammatical aspects (/6)	1.43 (1.29)	3.19 (1.72)	-5.02***	1.55	2.52 (1.75)	2.29 (1.52)	.65	.20	-2.69*	.83
Verbal communication (/12)	9.90 (1.76)	11.05 (.97)	-3.17**	.98	11.43 (.68)	11.05 (1.32)	1.19	.37	-3.06**	.94
Non-verbal communication (/6)	3.81 (1.21)	4.05 (1.24)	-0.93	.29	4.57 (.98)	5.19 (1.08)	-2.77*	.86	-4.32***	1.33

* $p < .05$
** $p < .01$
*** $p < .001$

In terms of progress between the beginning (test 1) and the end (test 4) of the instructional program, significant improvement was noted in four elements, with a large effect size for each. This is quite similar to what was seen for the written feedback condition, in which three elements improved significantly and showed a large effect size. The same three elements improved in the discussed oral feedback condition, plus improvement in the grammatical aspects as well. The organization and coherence of the message went from an average score of 8.16 out of 21 (SD = 2.21) at test 1 to 13.38 (SD = 2.44) at test 4 ($t(20) = -8.08$, $p < .001$, d = 2.49). The average score for grammatical aspects showed a large increase as well, from 1.43 out of 6 (SD = 1.29) at test 1 to 2.29 (SD = 1.52) at test 4 ($t(20) = -2.69$, $p < .05$, d = .83). The average score for verbal communication grew from 9.90 out of 12 (SD = 1.76) at test 1 to 11.05 (SD = 1.32) at test 4 ($t(20) = -3.06$, $p < .01$, d = .94). And non-verbal communication showed an increase from 3.81 (SD = 1.21) at test 1 to 5.19 (SD = 1.08) at test 4 ($t(20) = -4.32$, $p < .001$, d = 1.33).

For the first task (test 1/test 2), three elements out of the six measured showed significant progress and a large effect size. The organization and coherence of the message showed progress from 8.19 out of 21 (SD = 2.21) at test 1 to 12.29 at test 2 (SD = 2.57), ($t(20) = - 6.37$,

$p < .001$). Grammatical aspects improved from 1.43 out of 6 (SD =1.29) at test 1 to 3.19 (SD = 3.19) at test 2 ($t(20) = -5.02$, $p < .001$, $d = 1.55$). The average score for verbal communication improved from 9.90 out of 12 (SD = 1.76) at test 1 to 11.05 (SD = .97) to test 2 ($t(20) = -3.17$, $p < .01$, $d = .98$). The improvement in the organization and coherence of the message ($t(20) = -6.37$, $p < .001$, $d = 1.96$) can be explained by an increase in the mean number of time connectors from 5.81 (SD = 2.68) to 8.24 (SD = 2.30), ($t(20) = 3.14$, $p = .005$). Grammatical aspects development can be explained by the improvement in two items: the mean number of correct sentences increased from 10.19 (SD = 5.34) to 17.29 (SD = 6.62), ($t(20) = 4.87$, $p < .001$), and the mean number of adverbs increased from 11.76 (SD = 6.68) to 17.90 (SD = 7.65), ($t(20) = 3.71$, $p = .001$). Improvements in verbal communication can be explained by an increased number of students who scored well for their speech rate (at test 1, eight students expressed themselves using a varied flow rate and appropriate pauses, compared to 18 at test 2), their intonation (at test 1, 13 students expressed themselves in an expressive and varied manner, compared to 16 at test 2), and their volume (at test 1, 16 pupils spoke audibly, compared to 20 at test 2).

For the second task (test 3/test 4), two elements showed large improvement: the organization and coherence of the message and non-verbal communication. For the organization and coherence of the message, the mean score at the beginning of test 3 was 9.10 out of 21 (SD = 2.26) and rose to 13.38 (SD = 2.44) at test 4 ($t(20) = 6.69$, $p < .001$, $d = 2.06$). This improvement came from progress in certain areas. More students added an introduction to their text (0 students with an introduction at test 3, nine at test 4). Students did better at development of their ideas (13 students developed their ideas correctly at test 3, compared to 21 at test 4). They also invested themselves in the message by using “I” (seven pupils at test 3, 16 at test 4) and by giving their personal opinions (0 students at test 3, nine at test 4).

For non-verbal communication, at test 3, students attained an average score of 4.57 out of 6 (SD = .98), which improved at test 4 to 5.19 (SD = 1.08), $t(20) = 2.77$, $p = .01$, $d = .86$. For non-verbal communication, scores for the relevant criterion greatly improved, from five students showing adequate eye contact at test 3 to 18 at test 4.

Two elements did not show any improvement during the instructional program: the intention to communicate and the development of ideas. Although we can make the same assumption as before concerning the lack of improvement in communication intention, we have no concrete facts to help us understand why there was no significant improvement in the development of ideas.

Comparison of the progress in the two conditions

This comparison is structured in two parts. First, based on the results presented above, we investigate whether students in each condition were able to use the elements they had learned in another task. We also compare the students' results across the curriculum. Second, we compare the group scores in both conditions on the different tests and for each task.

To see if the students were able to use the elements they had learned in a different task, test 1 and test 3 were compared. As shown in Tables 5 and 6, students in both conditions improved. This means that students were able to use the skills they learned, regardless of the feedback modality. Nevertheless, students in the written feedback condition showed significant improvements in four elements (message organization, grammatical aspects, verbal and non-verbal communication), while in the discussed oral feedback condition, only three elements showed significant improvement (grammatical aspects, verbal and non-verbal communication).

However, as mentioned above, when students' results are compared over the entire instructional program, from test 1 to test 4, it is clear that students in the discussed oral feedback condition performed better. Indeed, while students in both classes showed large improvements in three elements (message organization, verbal communication and non-verbal communication), students in the discussed oral feedback condition made greater progress in two out of these three elements. For message organization, in the written feedback condition, the mean went from 8.33 out of 21 (SD = 1.50) to 11.62 (SD = 2.54), ($t(20)$ =-5.29, $p < .001$, d = 1.63), an increase of 3.29; in the discussed oral feedback condition, the mean went from 8.19 (SD = 2.21) to 13.38 (SD = 2.44), ($t(20)$ = -8.08, $p < .001$, d = 2.49), an increase of 5.19. For non-verbal communication, in the written feedback condition, the mean went from 3.33 out of 6 (SD = .66) to 4.48 (SD = 1.03), ($t(20)$ = -4.73, $p < .001$, d = 1.43-, an increase of 1.15; in the discussed oral feedback condition, the mean went from 3.81 (SD = 1.21) to 5.19 (SD = 1.08), ($t(20)$ = -4.32, $p < .001$, d = 1.33). Therefore, as the discussed oral feedback condition also showed significant progress in grammatical aspects, those in this condition made greater progress overall as compared to those in the written feedback condition.

In short, the instructional program based on the itineraries method improved some of the oral communication skills of the pupils in our sample. In addition, when feedback was discussed between students during peer assessment times, the results were better. To verify this, a covariance analysis was performed to compare the scores at the different tests for each task: test 1 versus test 2, and test 3 versus test 4. Change over the entire instructional program was also observed: test 1 versus test 4. The results are presented in Table 7. The covariate was the overall score in the initial test and the dependent variable was the overall score in the later test.

The independent variable was the condition (discussed oral feedback condition versus written feedback condition).

Table 7 - Covariance analysis

Covariance analysis at test 2 ($N = 42$)

Source	Sum of square	*df*	Mean square	$F(1, 40)$	*Partial Eta Squared*
Test 1	120.91	1	120.91	4.59 *	.105
Condition	3.80	1	3.80	.14	.004
Error	1026.61	39	26.32		

Covariance analysis at test 4 ($N = 42$)

Source	Sum of square	*df*	Mean square	$F(1, 40)$	*Partial Eta Squared*
Test 3	236.10	1	236.10	9.83**	.201
Condition	122.31	1	122.31	5.09*	.115
Error	936.76	39	24.20		

Covariance analysis at test 4 ($N = 42$)

Source	Sum of square	*df*	Mean square	$F(1, 40)$	*Partial Eta Squared*
Test 1	197.74	1	197.74	7.91**	.169
Condition	114.39	1	114.39	4.58*	.105
Error	975.12	39	25.00		

* $p < .05$

** $p < .01$

The results indicate that condition mattered, and that discussed oral feedback led to better results. For the first task, the differences were not significant. However, they were for the second task: $F(1,40) = 5.09$, $MSE = 24.20$, $p < .05$, partial $\eta^2 = .115$. They were also significant over the whole instructional program: $F(1,40) = 4.58$, $MSE = 25.00$, $p < .05$, partial $\eta^2 = .105$,

which is a moderate effect. On test 4, 10.5% of the variance in scores was due to the experimental condition that students belonged to, after controlling for the scores at the pretest. Figure 2 shows an overview of the change in student results for the two conditions over the four tests. It can be seen that the students in discussed oral feedback condition indeed have a better growth curve.

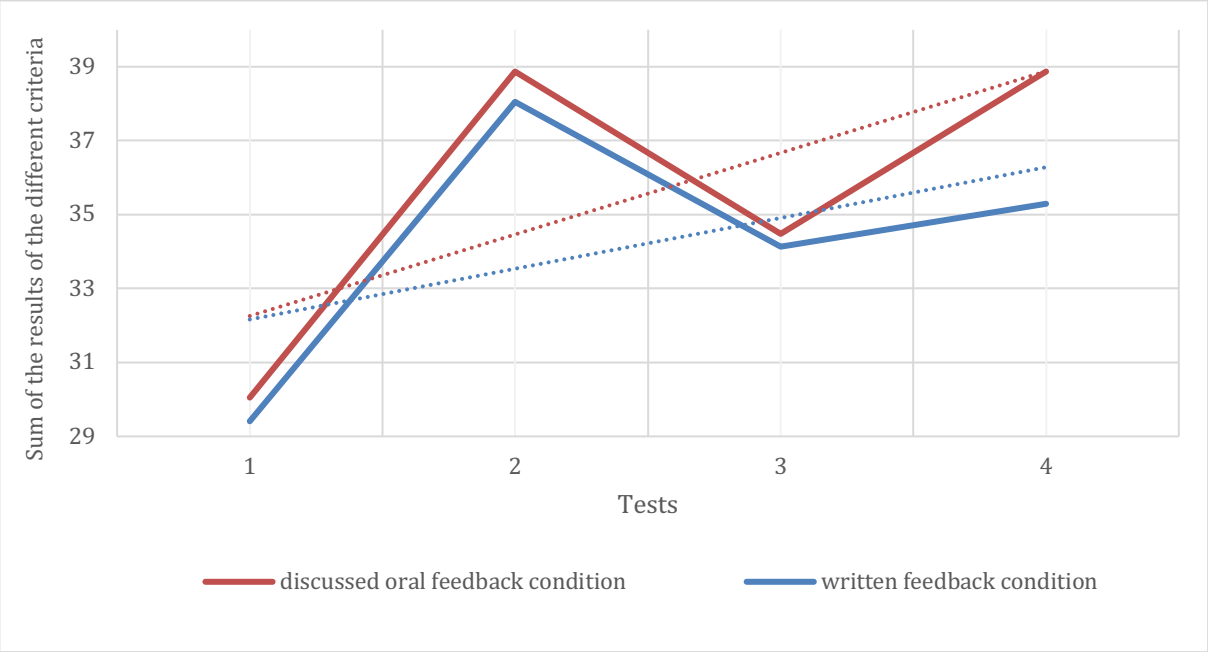


Figure 2 – Change in students' scores over the four tests, by condition

Discussion and conclusion

In this contribution, we used a method for teaching oral communication skills based on a pedagogical approach addressing oral genres (Dolz & Gagnon, 2008; Freedman & Medway, 1994; Horverak, 2016; Hyland, 2003, 2007). This method encourages repeated oral performances, teacher support, peer assessment, and metacognition. In this method, the pupils are involved in giving feedback to their peers several times. We tested two modalities: feedback given individually in writing, or given orally by a small group of students having a discussion.

It seemed relevant to us to test these two modalities in the specific case of oral communication, and more specifically, the ability to speak where the whole person is "exposed".

In both cases, in this context of formative evaluation (Earl, 2003), the majority of comments had an informative nature (Rodet, 2000). In addition, in both conditions, peer assessment was mostly focused on certain aspects of the teacher's scaffolding. Indeed, students commented most on the organization and coherence of the text. They also addressed verbal and non-verbal communication and, least often, grammar. It therefore seems that peer assessment is a way of directly applying what has been seen during instruction. Students also address, but little, the development of ideas, while the communication intention is not addressed by the students at all.

Moreover, the results show that in both conditions, the oral skills of the pupils improved. Indeed, our results show improvement in four elements in both conditions: message organization, grammar, verbal communication and non-verbal communication. On this point, in conjunction with our concern to be able to offer teachers teaching methods for oral communications skills, and the lack of methods identified in the literature (Sénéchal, 2017), we can say that the itineraries method used enabled the students in the two conditions we worked with to progress in these skills, as we have also shown elsewhere (Authors, in press; Authors, submitted). Nevertheless, the aspect relating to the intention to communicate did not improve. Perhaps this is due to the fact that the results were already good at the beginning in both groups. Perhaps this is why it was also not addressed in the peer feedback. We attributed this initial good result to the instructional program and the students' project activity. Its purpose was to highlight the parameters of the communication situation. In addition, the development of ideas improved, but this improvement was only significant in the written feedback condition during

the first task. Students spent no time in class working on this aspect with the teacher, unlike the other elements of oral communication. Thus, students did not have any support on how to develop ideas. This may suggest a way to improve this instructional program.

Besides, there are differences between the two conditions. When measuring progress for each task and for the instructional program as a whole, the results showed that the discussed oral feedback condition made the most progress overall. These differences in the two conditions can be related not only to the discussion between the group members, but also to the difference between having oral or written feedback.

Indeed, in the discussed oral feedback condition, the results were therefore better overall. This is consistent with the conclusions of Suh (2005) and Biber, Nekrasova, and Horn (2011). There are several reasons for this. First, when students discussed the feedback, they provided more comments than in the written feedback condition. This may be due to the fact that writing feedback, as Sheen (2010) reminded us, is more complex than giving it orally. Second, it also seems, contrary to what one might think, that participating in the discussion led students to become more aware of the aspects to be modified than when they read the written feedback received alone. The oral discussion could have the strength to be more convincing than the written feedback. Suh (2005) explained that facing people has a greater impact than reading about their point of view. Third, in accordance with van den Berg, Admiraal and Pilot (2006), this modality led students to adopt a supportive posture, suggesting improvements (Rodet, 2000). This way of doing things is present in 20% of the meaning units in the discussed oral feedback condition, compared to 4% in the written feedback condition. Nevertheless, the frequency of feedback of this nature remained low. This suggests that it would be interesting,

if not necessary, to teach students to give feedback to others and to work with them on the different ways to do so. In particular, this has already been suggested by Leenknecht and Prins (2017) and Paulus (1999).

In conclusion, the results indicate several findings: 1) oral skills can be developed through dedicated teaching; 2) when students give feedback to their peers, it is essentially informative, focused on what is right or wrong; 3) when students discuss an oral performance with the speaker in small groups, the feedback includes more suggestions than when it is written; 4) when students discuss an oral performance with the speaker in small groups, possible improvements are more important. Previous studies that have looked at oral versus written feedback (e.g., Biber, Nekrasova, & Horn, 2011; Sheen, 2010; Suh, 2005) have often focused on improving writing. Our study is innovative in that it focuses on the development of oral skills. This area has been little investigated to date, both in terms of teaching and evaluation (Sénéchal et al., 2019).

Let us point out some limitations and perspectives regarding this study. First, we cannot attribute the effects noted in the discussed oral feedback condition to a particular feature of the situation. Indeed, we cannot at this stage say whether it was the fact of having received the comments orally, or whether it was the peer discussion itself that led to the improvements. We plan to replicate this study with an additional condition. Thus, in a new condition, students will receive recorded oral feedback from each evaluator. This will neutralize interactions between evaluators. We did not implement this modality here, because we were looking to compare modalities that are easy for teachers to implement. The literature shows that they are reluctant to teach oral communication and need modalities that are straightforward. This third modality seems more complex to implement, particular because it requires equipment (such as tablets or

computers), and teachers will have to use these technologies with their students. However, it nevertheless deserves to be investigated.

Second, our sample is an important limitation, since its size (42 participants) does not allow us to generalize the results. We intend to reproduce this experiment at a larger scale to confirm our findings, particularly by exploring other school environments.

Third, another limitation is that we measured student progress at four times, but the final measurement (test 4) was very close to the last learning session. It would be interesting to know what students actually retain from this instructional program, by reassessing them later in their school curriculum. In addition, it could be verified whether the skills addressed in this method are put into practice in real classrooms when students are asked to tell a life story. This would make it possible to further verify the transfer of this learning (Presseau & Frenay, 2004). Fourth, we do not know at this point where the progress in the discussed oral feedback condition comes from. Did students improve their oral performance because the feedback received was more precise and formulated in terms of advice, or did they improve because collaboratively discussing the performance of others prompted self-evaluation of their own performance? Or did students improve because they used both the elements discussed in the group and the feedback received, as we have shown for rewriting activities (Authors, 2019)? Therefore, this can be a focus of future research. Fifth, the two tasks concerned oral performance related to a life story. Other genres of discourse could be explored, making it possible to diversify approaches and, in the long term, to make it possible to develop a rigorous and precise curriculum to guide teachers in oral communications skills instruction and assessment.

Notwithstanding these limitations, this study is innovative in that, on the one hand, it places oral language at the center of learning as a real object of teaching and evaluation and, on the other hand, because it investigates the ways that students can provide feedback in this context.

Moreover, allowing pupils to repeat and revise their oral performance several times, as the itineraries method does, means that progress is possible and measurable. Finally, in addition to the suggestions for future research mentioned above, it is worthwhile to further examine the discussion between peers when they provide feedback: how do students react, contradict each other and adjust, and what discussion processes are at stake? This may also require some attention and may be related to the extent to which students are able to give effective feedback (on what they have listened to) and to benefit from the feedback (what they are listening to).

References

Allal, L. (1991). *Vers une pratique de l'évaluation formative* [Towards a practice of formative evaluation]. Bruxelles: De Boeck.

Allal, L., & Mottier Lopez, L. (2005). Formative assessment of learning: A review of publications in French. In OECF (Ed.), *Formative assessment – Improving learning in secondary classrooms* (pp. 241-264). Paris: OECD-CERI Publications.

Allen, N. (2018). Développer la littératie volet oral pour favoriser la transition interordre entre le primaire et le secondaire: quelques considérations théoriques [Developing oral literacy to support the interorder transition from elementary to secondary school: some theoretical considerations]. *Canadian Journal of Education*, *40*(4), 638-664.

Authors. (2015).

Authors. (2016).

Authors. (2017)

Authors. (2018a).

Authors. (2018b).

Authors. (2019).

Author. (in press).

Author. (submitted).

Bergeron, R., Dumais, C., Harvey, B., & Nolin, R. (2015). *La didactique du français oral du primaire à l'université* [The didactics of oral French from primary school to university]. Côte Saint-Luc, Québec: Editions Peisaj.

Bianco, M. (2015). *Du langage oral à la compréhension de l'écrit* [From oral language to reading comprehension]. Grenoble: Presses Universitaires de Grenoble.

Biber, D., Nekrasova, T., & Horn, B. (2011). The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. *ETS Research Report Series*, *2011*(1), i-99

Bostock, S. (2000). Student peer assessment. Keele University, *Learning Technology*. Accessed at: https://s3.amazonaws.com/academia.edu.documents/32570036/Student_peer_assessment_-

Boud, D. (1989). The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education, 14,* 20-30.

Brookhart, S. (2008). *How to give effective feedback to your students*. ASCD: Alexandria, VA

Bronckart, J. P., & Dolz, J. (1999). La notion de compétence : quelle pertinence pour l'étude de l'apprentissage des actions langagières ? [The notion of competence: what relevance for the study of the learning of language actions?]. *Raisons éducatives*, *2*, 27-44.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Communauté Française. (2013). *Socles de compétences* [Skill bases]. Bruxelles: Administration Générale de l'Enseignement et de la Recherche Scientifique, Service Général du Pilotage du Système Educatif.

De Ketele, J.M. (2010). Ne pas se tromper d'évaluation [Do not make the wrong assessment]. *Revue française de linguistique appliquée*, *15*(1), 25-37.

Dougan, A.M. (1996). Student assessment by portfolio: One institution's journey. *The History Teacher, 29*, 171-178.

Dolz, J., & Gagnon, R. (2008). Le genre du texte, un outil didactique pour développer le langage oral et écrit [The gender of the text, a didactic tool for developing oral and written language]. *Pratiques. Linguistique, littérature, didactique*, (137-138), 179-198. doi 10.4000/pratiques.1159

Dolz, J., Pasquier, A., & Bronckart, J. P. (1993). L'acquisition des discours: Émergence d'une compétence ou apprentissage de capacités langagières diverses? [Speech acquisition: The emergence of a skill or learning of various language skills?]. *Études de linguistique appliquée, 92*, 23-37.

Dumais, C. (2012). Vers une progression des objets d'enseignement/apprentissage de l'oral qui s'appuie sur le développement des élèves [Towards a progression of oral teaching/learning objects based on student development]. In R. Bergeron & G. Plessis-Bélair (Eds.). *Représentations, analyses et descriptions du français oral, de son utilisation et de son enseignement au primaire, au secondaire et à l'université* [Representations, analyses and descriptions of oral French, its use and teaching at the primary, secondary and university levels] (pp. 19-34). Côte Saint-Luc: Editions Peijac.

Dumais, C., & Lafontaine, L. (2011). L'oral à l'école québécoise : portrait des recherches actuelles [Oral communication in Quebec schools: a portrait of current research]. *Revue suisse des sciences de l'éducation*, *33*(2), 285-302.

Dumais, C., Lafontaine, L., & Pharand, J. (2017). " J'enseigne et j'évalue l'oral": pratiques effectives au 3e cycle du primaire ["I teach and evaluate oral communication": actual practices at the 3rd cycle of primary school]. In J.-F. de Pietro, C. Fisher, & R. Gagnon

(Eds.), *L'oral aujourd'hui: perspectives didactiques* [Oral communication today: didactic perspectives] (pp. 151-174). Namur: Presses Universitaires de Namur.

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, *31*(2), 115-128.

Dupont, P., & Grandaty, M. (2016). De la dichotomie oral enseigné pour apprendre à la dialectique oral travaillé-oral enseigné [From the oral dichotomy taught to learn to the oral dialectic worked-oral taught]. *Repères, 54*, 7-16.

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.

Freedman, A., & Medway, P. (1994). *Learning and teaching genre*. Portsmouth: Boynton/Cook.

Gagnon, R., De Pietro, F., & Fisher, C. (2017). Introduction. In J.-F. de Pietro, C. Fisher, & R. Gagnon (Eds.), *L'oral aujourd'hui: perspectives didactiques* [Oral communication today: didactic perspectives] (pp. 11-40). Namur: Presses Universitaires de Namur.

Gagnon, R., & Dolz, J. (2016). Enseigner l'oral en classes hétérogènes : quelle ingénierie didactique [Teaching oral communication in heterogeneous classes: which didactic engineering]? *Les Dossiers des Sciences de l'Education, 136,* 109-130.

Garcia-Debanc, C. (1999). Evaluer l'oral [Evaluate oral communication]. *Pratiques, 103/104*, 193-212.

Garcia-Debanc, C., & Delcambre, I. (2001-2002). Enseigner l'oral [Teaching oral communication]. *Repères, 24/25*, 3-21.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304-315.

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137-155.

Green, S., & Salkind, N. (2017). *Using SPSS for Windows and Macintosh* (8th ed.). Pearson.

Halté, J.-F., & Rispail, M. (2005). *L'oral dans la classe : compétences, enseignement, activités* [Oral communication in the classroom: skills, teaching, activities]. Paris: L'Harmattan.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Horverak, M. O. (2016.) Students' and teachers' perceptions on writing instruction inspired by genre-pedagogy and Systemic Functional Linguistics. *Proceedings of CECIL's 5, Olomouc, Czech Republic,* 58-73

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing, 12*(1). 17-29.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*(3). 148-164.

Jaubert, M. (2007). *Langage et construction de savoirs à l'école, un exemple en sciences* [Language and knowledge building in schools, an example in science]. Bordeaux: Presses Universitaires de Bordeaux.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*, 130-144.

Kendeou, P., White, M.-J., van den Broek, P., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*(4), 765-778.

Ko, J., Sammons, P., & Bakkum, L. (2014). *Effective teaching: A review of research and evidence*. England: CfBT Education Trust.

Lafontaine, L. (2007). *Enseigner l'oral au secondaire. Séquences didactiques intégrées et outils d'évaluation* [Teach oral communication in high school. Integrated didactic sequences and evaluation tools]. Montréal (Québec): Chenelière Education.

Lafontaine, L. (2013). Reading and oral literacy in disadvantaged communities: Quebec kindergarten, first and second grade teachers' perceptions and practices, *International Journal of Humanities*, *10*, 69-80.

Lafontaine, L., & Messier, G. (2009). Les représentations de l'enseignement et de l'évaluation de l'oral chez des enseignants et des élèves du secondaire en français langue d'enseignement [Representations of oral instruction and assessment in teachers and secondary school students in French as the language of instruction]. *Revue du Nouvel-Ontario, 34,* 119-144.

Lafontaine, L., & Préfontaine, C. (2007). Modèle didactique descriptif de la production orale en classe de français langue première au secondaire [Descriptive didactic model of oral production in French as a first language in high school]. *Revue des Sciences de l'Education, 33*(1), 47-66. doi: 10.7202/016188ar

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. Retrieved from http://www.jstor.org/stable/2529310

Lavoie, C., & Bouchard, É. (2017). In De Pietro, J-F., R. Gagnon et C. Fisher (Eds.). *Formation universitaire à l'évaluation de l'oral : regard sur la capacité d'autoévaluation de futurs enseignants. L'oral aujourd'hui : perspectives didactiques* (pp. 259-274) [University training in oral evaluation: a look at the self-evaluation capacity of future teachers. Oral communication today: didactic perspectives]. Namur, Belgique: Presses universitaires de Namur.

L'Écuyer, R. (1990). Méthodologie de l'analyse développementale des contenus. Méthode GPS et concept de soi [Methodology of developmental content analysis. GPS method and self-concept]. Québec, Canada : Presses de l'Université du Québec.

Leenknecht, M. J. M., & Prins, F. J. (2018). Formative peer assessment in primary school: the effects of involving pupils in setting assessment criteria on their appraisal and feedback style. *European Journal of Psychology of Education,* 33, 101-116.

Lu, J., & Law, N. (2011). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science* 40(2), 257–275

Le Cunff, C., & Jourdain, P. (2008). *Enseigner l'oral à l'école primaire* [Teaching oral communication in primary school]. Paris: Hachette Education.

Morreale, S. P., Moore, M. R., Taylor, K. P., Surges- Tatum, D., & Hulbert-Johnson, R. (1993). *The Competent Speaker Speech Evaluation Form.* SCA: Annan- dale, VA.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning, *School Effectiveness and School Improvement*, *25*(2), 231-256

Nolin, R. (2015). Portrait des pratiques déclarées d'enseignement et d'évaluation de l'oral en classe du primaire au Québec [Portrait of the declared practices of oral instruction and assessment in primary school classes in Quebec]. In R. Bergeron, C. Dumais, B. Harvey, & R. Nolin, *La didactique du français oral du primaire à l'université* [The didactics of oral French from primary school to university]. Canada: Peisaj.

Nonnon, E. (2016). 40 ans de discours sur l'enseignement de l'oral. Enseignement/apprentissage de la langue, des textes et des discours [40 years of speeches on the teaching of oral communication. Teaching/learning of language, texts and speeches]. *Pratiques* [online], 169-170. Accessed [le 10 novembre 2016] http://pratiques.revues.org/3115; DOI: 10.4000/pratiques.3115

Paulus, M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265-289.

Plane, S. (2015). *Pourquoi l'oral doit-il être enseigné* [translation]*?* Retrieved from Cahiers pédagogiques of http://www.cahiers-pedagogiques.com/Pourquoi-l-oral-doit-il-etre-enseigne

Pomplun, M., Capps, L., & Sundbye, N. (1998). Criteria teachers use to score performance items. *Educational Assessment*, *5*(2), 95-110

Presseau, A., & Frenay, M. (2004). *Le transfert des apprentissages : comprendre pour mieux intervenir* [The transfer of learning: understanding to better intervene]. Laval: Les Presses de l'Université.

Putnam, R., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*(1), 4-15.

Rodet, J. (2000). La rétroaction, support d'apprentissage? *DistanceS* [Feedback as a learning tool?]. Montréal: TÉLUQ/UQAM. Accessed https://edutice.archives-ouvertes.fr/edutice-00000482/document

Sénéchal, K. (2017). Quels obstacles à la mise en oeuvre de séquences didactiques destinées à l'enseignement de l'oral [What are the obstacles to the implementation of didactic sequences for oral instruction?]? In J.-F. De Pietro, C. Fisher, & R. Gagnon (Eds.), *L'oral aujourd'hui: perspectives didactiques.* [Oral communication today: didactic perspectives] (pp. 177-199). Namur: Presses Universitaires de Namur.

Sénéchal, K., & Chartrand, S.-G. (2011). Etats des lieux de l'enseignement du français (ELEF) [State of play of French language teaching]. *Les cahiers de l'AQPF, 1*(1). 4-5.

Sénéchal, K., & Chartrand, S.-G. (2012). Représentations et pratiques de l'enseignement de l'oral en classe de français : changement et constantes depuis 25 ans [Representations and practices of oral teaching in French classes: changes and constants over the past 25

years]. In R. Bergeron & G. Plessis-Bélair (Eds.). *Représentation, analyse et description du français oral, de son utilisation et de son enseignement au primaire, au secondaire et à l'université* [Representation, analysis and description of oral French, its use and teaching at the primary, secondary and university levels] (pp. 185-199). Côte Saint-Luc: Editions Peisaj.

Sénéchal, K., Dumais, C., & Bergeron, R. (2019). *Didactique de l'oral : de la recherche à la classe, de la classe à la recherche* [Oral didactics: from research to class, from class to research]. Québec: Editions Peisaj.

Simard, C., Dufays, J.L., Dolz, J., & Garcia-Debanc, C. (2014). *Didactique du français langue première* [Didactics of French as a first language]. Bruxelles: De Boeck Supérieur.

Sheen, Y. (2010). Introduction: The role of oral and written corrective feedback in SLA. *Studies in second language acquisition*, *32*(2), 169-179

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934-947.

Suh, J. (2005). Peer feedback interactions in EFL compositions: Written feedback versus oral feedback. *ENGLISH TEACHING (영어교육)*, *60*(3), 91-116

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. doi:10.3102/00346543068003249

Van den Berg, I., Admiraal, W., & Pilot, A. (2006) Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education*, *31*, 341-356. http://dx.doi.org/10.1080/03075070600680836

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, 20(4), 280–290.

Veenman, M. (2012). Metacognition in science education: Definitions, constituents, and their intricate relation with cognition. In A. Zohar & Y.-J. Dori (Eds.), *Contemporary trends and issues in science education metacognition and learning* (pp. 21-36). London: Springer.

Wegmuller, E., & Allal, L. (1997). LEX-EVAL: lexique d'évaluation dans le cadre de la scolarité obligatoire [lexicon of evaluation in the context of compulsory education]. *Résonances, 3*, 14-15.

Younes, N., & Gaime, E. (2012). L'évaluation formative en contexte : points de vue d'enseignants, points de vue d'élèves [Formative evaluation in context: teacher's points of view, student's points of view]. *Diversité, VEI*, *169*, 161-166.

Appendix 1 – Grid used by students to evaluate their peers (task 1)

	My comments
Overall commentary	....
Content of the message	
Do we know who it is?	....
Do we know what we're talking about?	....
Do we know where the events are happening?	....
Do we know when it's happening?	....
Organization and coherence of the message	
Introduction	
Several personnal events are presented?	....
Are the events presented in chronological order?	....
Are time connectors used?	....
Message comprehension	
Quality of the message	....
Présentation	
Is the speaker putting intonation into his speech?	....
Any other comments?	

Appendix 2 : test example 2 and test example 4

Example of test 2 performance

"I'm going to present my day to you on Thursday, March 29, 2018.
Well, at the beginning of the day, I got ready for school. Like everyone normally does on a Thursday. Then, well, well, I went to school. We were supposed to go to class but Mrs. Stephanie was not there. So he replaced Mrs. Jessica for the first two hours. During those two hours, students from third B or third A came to our class because their teacher wasn't there. We worked a lot. After the two hours, it was recess. After recess, instead of going back to the classrooms where we were supposed to be, Mrs. Caroline kept us in class. And some of us got a little crazy. Then we had gym to, uh, rehearse the fancy fair show. It was really great! Then it was recess. We had recess right after we ate our bowl of rice. Because it was rice bowl day. We had three refills. But I don't see the point of having a rice bowl day if we could get a refill. After number uh we, after recess we did Dutch all afternoon. It was nice because we talked about Easter. And we didn't do the same thing all the time. After like after that big day, I was very tired and I went to sleep. Voilà!" (Alice)

Example of test 4 performance

"If I had a unicorn, I could go to school with her and my friends would all be jealous. My favorite animal would be the unicorn and I would do my report on it. She would go get the stars and bring them back to my room. And the next morning, when she wakes me up, I would be in a good mood and I would go to school in a good mood with nothing on my mind.
But I don't have a unicorn. I only have a brother. He's wonderful. "(Achilles)

Appendix 3 - Weightings used for the assessment of oral skills: complement to Table 2

Idea development

	Weightings	0	1	2	3
Number of sentences	Min = 3 Max = 31	0 → 8	9 → 16	17 → 24	25 → 32
Number of words used	Min = 35 Max = 300	0 → 75	76 → 150	151 → 225	226 → 300
Number of ideas	Min = 2 Max = 17	0 → 5	6 → 10	11 → 15	16 → 20
Number of words per idea	Min = 6.67 Max = 44.67	0 → 12.9	13 → 24.9	25 → 36.9	37 → 48

Organization and coherence

	Weightings	0	1	2	3
Number of time connectors	Min = 1 Max = 24	0 → 6	7 → 12	13 → 18	19 → 24
Number of different connectors	Min = 1 Max = 12	0 → 3	4 → 6	7 → 9	10 → 12

Grammatical aspects

	Weightings	0	1	2	3
Number of syntactically correct sentences	Min = 3 Max = 27	0 → 7	8 → 14	15 → 21	22 → 28
Number of adverbs	Min = 1 Max = 39	0 → 10	11 → 20	21 → 30	31 → 40

Appendix 4 - Inter-rater reliability results obtained for each criterion

Variables		Task 1 – pre-test		Task 1 – post-test		Task 2 – pre-test		Task 2 – post-test	
		Results SPSS	**Interpretation**	**Results SPSS**	**Interpretation**	**Results SPSS**	**Interpretation**	**Results SPSS**	**Interpretation**
1.	**Communication intention → ICC**								
1.1	Compliance with the instructions	constant		constant		constant		constant	
1.2	Respect for the time limit	k=0.949 (p=0.000)	excellent agreement	k=1.000 (p=0.000)	excellent agreement	k=0.875 (p=0.000)	excellent agreement	k=0.871 (p=0.000)	excellent agreement
1.3	Adapted language register	constant		constant		constant		constant	
2.	**Idea development → ICC**								
2.1	Number of sentences	ICC = 1 (p=0.000)	guaranteed reliability	ICC = 1 (p=0.000)	guaranteed reliability	ICC = 0,990 (p=0.000)	guaranteed reliability	ICC = 0,996 (p=0.000)	guaranteed reliability
2.2	Number of words used	ICC = 1	guaranteed reliability	ICC = 1.000	guaranteed reliability	ICC = 1	guaranteed reliability	ICC = 1	guaranteed reliability
2.3	Number of ideas	ICC = 0.950 (p=0.000)	guaranteed reliability	ICC = 0.853 (p=0.000)	guaranteed reliability	ICC = 0.970 (p=0.000)	guaranteed reliability	ICC = 0.987 (p=0.000)	guaranteed reliability
2.4	Number of words per idea	ICC = 0.969 (p=0.000)	guaranteed reliability	ICC = 0.789 (p=0.002)	guaranteed reliability	ICC = 0.911 (p=0.000)	guaranteed reliability	ICC = 0.991 (p=0.000)	guaranteed reliability
3.	**Message organization → ICC**								
3.1	Introduction/Start situation	k= 0.693 (p=0.000)	acceptable agreement	k= 0.642 (p=0.000)	acceptable agreement	k= 0.714 (p=0.000)	acceptable agreement	k= 0.719 (p=0.000)	acceptable agreement
3.2	Development of actions	k=0.337 (p=0.003)	poor agreement	k=0.754 (p=0.002)	acceptable agreement	k=0.652 (p=0.000)	acceptable agreement	k=1.000 (p=0.000)	excellent agreement
3.3	Personal opinion	k=0.656 (p=0.000)	acceptable agreement	k=0.639(p=0.001)	acceptable agreement	k= constant (p=0.000)	excellent agreement	k= 0.804 (p=0.000)	acceptable agreement
3.4	Consistency in speech	k=0.605 (p=0.000)	intermediate agreement	k=1.000 (p=0.000)	excellent agreement	k=1.0 (p=0.000)	excellent agreement	k=1.0 (p=0.000)	excellent agreement
3.5	Number of time connectors	ICC=0.977 (p=0.000)	guaranteed reliability	ICC=0.989 (p=0.000)	guaranteed reliability	ICC=0.991 (p=0.000)	guaranteed reliability	ICC=1.000 (p=0.000)	guaranteed reliability
3.6	Number of different connectors	ICC=0.956 (p=0.000)	guaranteed reliability	ICC=0.945 (p=0.000)	guaranteed reliability	ICC=0.989 (p=0.000)	guaranteed reliability	ICC=0.995 (p=0.000)	guaranteed reliability

3.7	Expression using "I", speaks of oneself	k=1 (p=0.000)	excellent agreement	k=1.000 (p=0.000)	excellent agreement	k=0.897 (p=0.000)	excellent agreement	k=0.742 (p=0.000)	acceptable agreement
4. Grammatical aspects → ICC									
4.1	Number of syntactically correct sentences	ICC = 0.925 (p=0.000)	guaranteed reliability	ICC = 0.979 (p=0.000)	guaranteed reliability	ICC = 0.989 (p=0.000)	guaranteed reliability	ICC = 0.990 (p=0.000)	guaranteed reliability
4.2	Number of adverbs	ICC = 0.989 (p=0.000)	guaranteed reliability	ICC = 0.995 (p=0.000)	guaranteed reliability	ICC = 0.997 (p=0.000)	guaranteed reliability	ICC = 1.000 (p=0.000)	guaranteed reliability
5. Verbal and non-verbal communication → ICC									
5.1	Volume	k=0.863 (p=0.000)	excellent agreement	k=constant (p=0.000)	excellent agreement	k=constant (p=0.000)	excellent agreement	k=constant (p=0.000)	excellent agreement
5.2	Speech rate	k = 0.671 (p=0.000)	acceptable agreement	k = 0.763 (p=0.000)	acceptable agreement	k = 0.901 (p=0.000)	excellent agreement	k = 1.000 (p=0.000)	excellent agreement
5.3	Voice articulation	k = 1 (p=0.000)	accord excellent	k = 1 (p=0.000)	accord excellent	k = constant (p=0.000)	excellent agreement	k = constant (p=0.000)	excellent agreement
5.4	Intonation	k = 0.750 (p=0.000)	acceptable agreement	k = 0.667 (p=0.008)	acceptable agreement	k = 1.0 (p=0.000)	excellent agreement	k = 0.765 (p=0.002)	acceptable agreement
5.5	Eye contact	k= 0.744 (p=0.000)	acceptable agreement	k= 1.000 (p=0.000)	accord excellent	k= 0.812 (p=0.000)	excellent agreement	k= 0.916 (p=0.000)	excellent agreement
5.6	Body posture	k=0.592 (p=0.000)	intermediate agreement	k=constant (p=0.000)	accord excellent	k=constant (p=0.000)	excellent agreement	k=constant (p=0.000)	excellent agreement