Language Proficiency Evaluation in the Age of AI: The Resurgence of Oral Production Skill

Maud Pillet, Snejina Sonina

Department of Language Studies, University of Toronto Scarborough, Canada.

Abstract

This article addresses a recent shift in the evaluation of language proficiency in our university courses towards the greater significance of oral production skill. Having compared language-practice and content-based French courses, we concluded that this shift represents a positive change in our teaching practice. We thus share our findings and thoughts on oral production and offer practical advice about the type of assessments and evaluation schemes that allow for the accurate and fair measurement of this skill.

Keywords: oral production; language proficiency; evaluation.

1. Introduction

About a century ago, Ferdinand de Saussure warned linguists about the unrightful place that written language tended to occupy in common opinion. "But the written word becomes so closely interwoven with the spoken word, of which it is the image," he taught in his *Cours de linguistique générale* in 1906-1911, "that it ends up usurping the leading role…" ("Mais le mot écrit se mêle si intimement au mot parlé dont il est l'image, qu'il finit par usurper le rôle principal …" ; de Saussure 1968, 45). In the same passage, he compared the study of a written image of the spoken word—instead of the direct study of the latter—to trying to know someone by looking at their photograph, rather than their actual face. A hundred years later, it turns out, our university language courses were still mostly doing exactly that: trying to judge our students' language proficiency based on their written exams (which equated 35-50 percent of the final mark in 2006-2011 at out department). Yet, slowly but surely, the situation has been changing.

Several factors contributed to the elimination of final written exams in the language practice courses at our department: First, the action-based approach (ABA) made the

evaluation of grammar skills quasi impossible because of its focus on practice instead of theory. Second, the COVID-19 years forced us to find more diverse methods of evaluation because take-home written exams did not yield reliable results. Third, the arrival of machine-translation and AI available on minuscule electronic devices like the Apple watch, hidden cameras, and ear-fitting listening devices made even in-class written exams less reliable than 10 years ago.

In this article, we will argue that the events that contributed to this change have only precipitated the inevitable and improved our evaluation schemes, whose re-evaluation was long overdue. We will also give examples of the most successful formative and summative assessments that contribute to the development and a better evaluation of the oral production skill, as well as the evaluation rubrics best suitable for each assignment.

2. New Trends in Language Teaching and Proficiency Evaluation

In theory, all four language skills need to be evaluated to assess language proficiency. However, in practice, passive language skills – reading and listening – are evaluated, while active language skills – writing and speaking, and especially speaking – are often left aside by university placement tests because they are harder to evaluate and usually require a live examiner and, thus, additional resources. While TOEFL has experimented with involving AI and automated systems in the evaluation of recorded answers, the test still involves human raters, just like most English and French proficiency tests. Thus, these official tests prove costly and not suitable for placing students in university courses or evaluating their proficiency at the end of a course.

Yet several new approaches that gained popularity in this century favour oral production and its evaluation as part of language proficiency. As we mentioned in the introduction, with the adoption of ABA, our language courses shifted from written final exams to more graduated formative assessments, which now include oral examinations. This seems to be a general tendency for courses using action-based and task-based language teaching (TBLT), while the latest development is the use of technological resources to improve oral production through self-study in preparation of oral activities (Serna Silva 2022). The appearance of Multimedia TBLT (Bava Harji & Gheitanchian 2017) and production-oriented approaches (POA) thus placed even greater emphasis on oral production (Wen 2024) and its integration in the blended teaching of languages (Zhang 2023).

In addition to automated translation tools, the 20s of this century brought a new challenge to university language courses – GenAI. In an article on the impact of GenAI on higher education,

A. Duane has summarized the provisional outcomes of research on the challenges of AI's availability for university courses: "Methods of teaching, learning and assessment will have to radically change" (Duane, 2024, 15). This conclusion coincides with our experience in teaching university language courses: we have already changed our methods of teaching both in French- language courses and content-based courses taught in French to more emphasize oral skills; our students have learned to use multimedia resources to improve their oral production; and our methods of assessment have been heavily tilted towards oral assignments and oral exams. We see these changes as positive, and we believe they have improved our teaching and assessment methods. We will share our most successful new practices below.

2.1. New Approach and Evaluation Methods in Beginner French Courses

Our approach to beginner courses and requirements for pronunciation differ from the general CEFR (the Common European Framework of Reference) approach to oral production. While CEFR presupposes the gradual improvement of pronunciation – from inaccurate but understandable at the A1 level, still imperfect at A2 and clear at B2, to near-native at the C1 and flawless at the C2 levels – we observed that pronunciation does not improve by itself when students progress through our university courses without taking a specialized phonetics course. In fact, since the ABA does not include any conscious work on phonetics, students reinforce their erroneous articulation through their years of study, and their pronunciation skills do not evolve as described by CEFR. We therefore approach pronunciation scientifically from the start – while working on their A1 level, our beginners master the articulation of all French phonemes, as well as all reading and prosodic rules of the French language. All these criteria are included in the final evaluation of the reading test and a video project.

We call our new approach to teaching beginners Theory-and-Practice, or TaP (Sonina 2023). In teaching oral production, it differs from other approaches through the inclusion of linguistic theory – especially articulatory phonetics and the International Phonetic Alphabet (IPA). Naturally, for these classes, we simplify the theory and tailor it to the classroom's practical needs. However, the structured scientific approach to pronunciation ensures a quick mastery of all the particularities of French phonetics in

just one semester (12 weeks, or 6 weeks in intensive summer courses). These results are consistent with the experience of using IPA to teach English oral production described by our Algerian colleague (Beghoul 2017).

2.1.1. Evaluation of Oral Production in Introductory French I

Throughout the semester, students work on their pronunciation through e-exercises and graded recordings. For practice, they are encouraged to use all means available, including soundtracks of dictionaries and of Google Translate, as well as the dictate function available in Word. They also have the opportunity to sign up for individual consultations with TAs to practice their graded assignments.

At the end of the course, they have two oral assignments to demonstrate their ability to speak French without reading (final video project) and their mastery of French phonetics (reading test). The video is evaluated based on the following criteria:

- Creativity – 2% (Say something original, important, or make it interesting to watch)
- Grammar – 2% (Double-check your grammar during individual consultations)
- Vocabulary – 2% (Favour vocabulary that you have learned in the course)
- Pronunciation – 2% (Respect rhythmic groups and the articulation of important phonemes)
- Fluency – 2% (Memorize your lines and speak fluently. Do not read!)

To evaluate the pronunciation more accurately, we decided to introduce a reading test, which allows TAs to check the most important French phonemes and prosodic rules in the same context. Students have 30 minutes to practice and record a text that they have never seen before.

They have to mark the rhythmic groups, shade the silent letters, and indicate the consonants participating in a liaison with red colour, as well as highlight any reading difficulties by bolding, as shown below on a reading-test practice template.

The oral production evaluated in the video is not spontaneous and only 4 percent out of 10 percent of the mark represent the evaluation of speaking per se – fluency and pronunciation. However, together with the evaluation of the pronunciation on the reading test (5%), short oral presentations during tutorials (10%), and recordings produced during the semester (5%), the total evaluation of oral production skills adds up to almost 25 percent of the final mark.
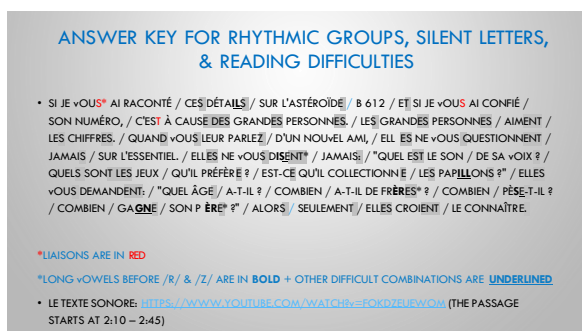
Figure 1. Reading-test template with examples of helpful highlights.

2.1.2. The Evaluation of Oral Production in Introductory French II

On the next level, the final video project is longer and requires cooperation with classmates, though the evaluation criteria remain the same. As for the reading test, it is conducted by TAs, so students prepare their reading in class and do not have an opportunity to listen to the text read by a computer program. Moreover, at the end of the reading, they have to answer two comprehension questions. The evaluation in class is possible on this higher level because students make fewer mistakes due to articulation: they review all the articulatory and prosodic difficulties learned in the Introductory French I, this time phonologically structured, and, as a result, most students master French pronunciation close to perfection. However, we still find that recording each student's performance is helpful, because the possibility of playing the recording many times over allows all mistakes to be marked accurately and to grade the test fairly after comparing students' success on the test.

2.2. New Evaluation Methods in Intermediate and Advanced French Courses

Following the evaluation methods applied in beginner-level courses, our university's intermediate and advanced French courses place an increasing emphasis on oral production. This shift reflects a progression in pedagogical goals, aiming to equip students with the skills necessary for effective communication in authentic contexts. The final course in this sequence, fourth-year Oral French, entirely dedicated to oral communication, represents a culmination of this focus. As students advance through the levels, evaluation methods evolve to prioritize interaction and improvisation, with evaluation criteria adapted to the specific course level. By encouraging spontaneity and adaptability, these methods not only strengthen language proficiency but also foster a

deeper engagement with the learning process, motivating students to transfer their linguistic knowledge to diverse real-life scenarios.

2.2.1. Evaluation of Oral Production in Language Practice at the Intermediate Level

At this level, oral production is assessed through group presentations (20%) and individual interviews (25%). Group presentations play a crucial role in assessing students' ability to work collaboratively and apply their linguistic knowledge in a structured format. The evaluation criteria for these presentations include:

- Narration Skills – 5% (Recounting and describing events or situations effectively)
- Vocabulary Use – 5% (Demonstrating a contextually appropriate lexicon)
- Grammar Use – 5% (Employing simple but appropriate grammatical structures)
- Pronunciation – 5% (Articulating clearly with correct rhythm)

In addition to group presentations, students undergo an individual interview, which evolves in complexity between first- and second-year courses. In the first course of the series, students prepare a short presentation on a course-related theme, followed by a guided discussion. In the next, an additional role-play component evaluates their ability to interact in more dynamic scenarios. In addition to the evaluation criteria applied to presentations, the interviews include points for social interaction, information exchange, and spontaneity.

2.2.2. Evaluation of Oral Production in Language Practice at the Advanced Level

In the higher-level courses, both group presentations and individual interviews are evaluated more rigorously. In presentations, in addition to the criteria listed above, fluency and presentation skills are evaluated individually, while the group work, representing 20 percent of the mark, is evaluated based on timing, organisation, originality, and the quality of visual aids. This dual evaluation ensures a balanced focus on both teamwork and individual performance. Additionally, students must submit a short video commentary on their peers' presentations, further reinforcing their oral skills.

The final course of the sequence entirely centred on oral communication incorporates a variety of innovative tasks designed to challenge students and help refine their skills. A task called Eloquence Capsule requires students to deliver a five-minute oral presentation on an idiomatic expression from the textbook, linking it to a social issue. This task evaluates their ability to research, synthesize, and present information creatively. The next task, Radio Show Simulation, allows students to collaborate,

exchange ideas, and demonstrate their ability to sustain a dynamic conversation during 15 minutes of "show time." Finally, the typical final interview consists of a 30-minute preparation period followed by two components: an argumentative presentation and a debate. These tasks assess students' ability to articulate opinions, defend ideas, and interact dynamically. On this level, all previous evaluation criteria remain in play, but more emphasis is given to coherence, linguistic precision, spontaneity, and task adherence.

2.3. New Evaluation Methods in Content-Based French Courses

Most of our content-based courses also include oral presentations and voice recordings as part of the final mark. For example, in Business French, weekly presentations account for 10 percent of the final mark, and the final oral presentation – a simulated job interview, which includes a self-introduction and a few "spontaneous" responses – accounts for another 10 percent of the final mark. At the beginning of the semester, the ungraded first attempt at self-introduction is recorded and submitted to the instructor who points out the pronunciation mistakes and assigns each student exercises useful to correct them. The "spontaneous" responses represent answers to typical job-interview questions. Students receive a list of questions in advance, but they do not know which exact questions will be asked during the simulation of the interview. Those unsure about their pronunciation can practice their answers using computer programs offering voiced versions of French sentences. The responses are evaluated based on purely oral criteria, 1 percent of each is given by the instructor, while the other represents the average of the evaluations provided by other students present in class:

- Intelligibility – 2% (Is the answer easy or difficult to understand?)
- Style – 2% (Is it simple and direct or complicated and tangled?)
- Tempo – 2% (Is it too fast, too slow, or just right?)
- Articulation – 2% (Is it clear and neat or unclear and inaccurate?)
- Intonation – 2% (Is it too monotonous, too agitated, or helpful for emphasis?)

Following the successful practice of an ungraded first-try recording in Business French, and due to the declining reliability of written exams, we also implemented entrance and exit recordings in our French Phonetics and Phonology course, which used to be purely theoretical. These recordings are evaluated based on the mastery of all phonological distinctions and prosodic particularities of French and now represent an integral part of the final mark.

3. Conclusion

When we compared our assessment methods, both between different courses and between the evaluation in each course now and 10 years ago, we found that within a decade, our assessment schemes had changed significantly in all language courses. The main tendency is apparent: all our language courses are now evaluated without final written exams, and all evaluation schemes include an oral examination or a voice recording. This alone proves the tendency of oral production regaining its rightful place in language proficiency. Two recent French monographs on the subject also confirm this tendency through the study of new didactic developments (De Pietro 2017) and of language facts pertinent to oral production (Cappeau 2021).

Certainly, there are still problems concerning official proficiency tests, which still either leave oral production aside or involve an interaction with an examiner (e.g., DELF/DALF tests) to evaluate speaking. For example, Evalang, used in our department for placement, does not evaluate any active language skills, but is based only on passive skill evaluation – reading and listening. However, it is possible to treat these tests as approximate and preliminary and then, in cases where it proves important, add to them an oral or written examination of active skills. We found that oral examinations do not always need to be thorough or lengthy – they just need to include all the components relevant to the given level, and above all, they need to be recorded, so it is possible to replay student's answers to ensure the accuracy of evaluation.

The innovative evaluation methods based on oral production implemented in our university's intermediate and advanced French courses are designed to foster authentic linguistic skills and to prepare students for real-world communication. By emphasizing interactive and improvised tasks, these approaches ensure that students develop both confidence and competence, while maintaining strong motivation throughout their learning journey. It is worth mentioning that students appreciate an opportunity to work with oral production. According to the end-of-term survey in 2024, 90 percent of students in Business French answered "YES" to the question "Do you think that correct pronunciation is important for your career?" while only 10 percent answered "more or less" and 0 percent chose "NO." Moreover, 71 percent of students found the initial evaluation of their pronunciation in the recorded first-try self-presentation "very helpful," 14 percent found it "quite helpful," 10 percent "more or less helpful," and only 5 percent did not find it helpful or have not done the first-try recording.

Overall, oral production proved to be the best way to evaluate student language proficiency, not only because it is AI-proof but also because it can be AI-enhanced in preparation for oral tasks and because it allows testing various language skills in one assessment. However, while working on this article, we concluded that we needed to do additional work on inclusiveness and on offsetting the anxiety experienced by some students while taking oral exams, especially when they are recorded and when the marks matter for their success. For now, we place our hopes on the preliminary, ungraded recording of the oral presentation, since this practice has proved successful in our content-based courses. We have also considered setting up a pronunciation clinic at our department, with students coming from diverse backgrounds and levels of language training. We will share our new experiments with the organization of oral exams in the next article. For now, we hope that the examples of formative assessments and marking schemes that we provided in this article will help language teachers to work more broadly with oral production and to administer oral examinations more often and more effectively.

References

Bava Harji, M., & Gheitanchian, M. (2017). Effects of Multimedia Task-Based Teaching and learning approach on EFL learners' accuracy, fluency and complexity of oral production. *TOJET The Turkish Online Journal Of Educational Technology*, 16(2), 25–34.

Beghoul, Y. (2017). Using the International Phonetic Alphabet in Teaching English Oral Production and Reception to Algerian Learners as a Potentially Efficient Solution to the Sound Spelling Incoherence. *Majallat al-ʿulūm al-insānīyah* (Constantine, Algeria), ﺏ(48), 63–78.

Cappeau, P. (2021). *Une grammaire à l'aune de l'oral ?* Presses universitaires de Rennes.

De Pietro, J.-F. et al. (2017). *L'oral aujourd'hui : Perspectives didactiques*. Presses universitaires de Namur.

Duane, A. (2024). Can Artificial Intelligence Complete My Assessment? A Student Led Initiative to Stress Test the Academic Integrity of University Course Work Using Generative AI. In: *10th International Conference on Higher Education Advances (HEAd'24)*. Valencia, 18-21 June 2024. https://doi.org/10.4995/HEAd24.2024.17143

Saussure, F. de. (1968). *Cours de linguistique générale*. Paris : Payot.

Serna Silva, I. (2022). From Language Teaching to Language Assessment with the Help of Technological Resources: Higher Education Students and Oral Production. *ELIA*, 22, 93–123. https://doi.org/10.12795/elia.2022.i22.04

Sonina S. (2023) Computerising and Intellectualising an Action-Based Approach for University Language Courses. In: *The Future of Education*, *13th Edition*. https://conference.pixel-online.net/library_scheda.php?id_abs=5986

Wen, Q. (2024). *A Production-Oriented Approach to Teaching Foreign Languages: Does a Post-Method Era Need a New Approach?* (1st ed.). Palgrave Macmillan.

Zhang, Y. (2023). Enhancing Oral Production in Integrated English Blended Teaching through a Production-Oriented Approach: An Action Research Study. *International Journal of Emerging Technologies in Learning*, 18(19), 61–71. https://doi.org/10.3991/ijet.v18i19.42477