Regular Paper

The JALT CALL Journal ISSN 1832-4215
Vol. 18, No.1 Pages 137–161
https://doi.org/10.29140/jaltcall.v18n1.601

Castledown

Statistical analysis of the impact of the e-learning platform Furago on French learners' listening skills

Albéric Deriblé
aderible@gmail.com
University of Tokyo, JAPAN

Éric Wiel
eric.wiel@gmail.com
Tokyo University of the Arts, JAPAN

The current study focuses on an e-learning platform called Furago designed to provide language instructors with exercises complementing activities already implemented in their classes. Its objective was to assess the impact the use of Furago could have on the listening skills of first-year Japanese university students. Statistical analyses that were used to outline the correlations between the main test variables (i.e., the grades achieved by students on two listening comprehension tests) and several predictor variables pertaining to the platform, such as the score the participants had accumulated or the consistency in which they use it, were all found to be statistically significant, albeit with correlation coefficients that were rather weaker than expected. Despite its users' slow margin of progression, as evidenced by the linear model conducted during this study, such results confirmed the value of Furago as a means to improve students' listening skills, especially, it would seem, at the very beginning of foreign language acquisition. Unexpectedly, these results also suggested that the users' improvement might have more to do with the overall amount of time they put in their study (reflected here by their scores on Furago) than their consistency.

Keywords: French, e-learning, Furago, listening skills, statistics

Introduction

The recent Coronavirus pandemic has made foreign language instructors all over the world rethink their teaching practices and reconsider the value of e-learning tools. The current study focuses on an Internet platform called

Furago (from the Japanese colloquial term referring to the French language) that generates a wide variety of exercises and drills according to the level of its users. Primarily designed to be used in parallel with traditional classroom activities, it operates autonomously and frees the teacher from the burden of correcting and/or grading assignments. The platform moreover keeps track of its users' performance and indeed rewards them by adding 40 points or so to their total score for each series of exercises they complete.

The objective of the current study was to assess the impact the consistent use of Furago could have on the listening skills of first-year Japanese university students. To that end, statistical analyses were used to outline the correlations, if any, between the main test variables, the grades achieved by students on two listening comprehension tests, and two sets of predictor variables. The research team considered firstly whether or not the subjects had used Furago prior to the listening comprehension tests, and secondly, for those who had: the pricing model they had registered for, the consistency in which they use the platform Furago, the type of exercises they had done and the overall score they had accumulated on the platform up to the day of each test.

In the article, we give a brief overview of the platform's contents and operating system and present the students' profiles, the type of listening tests they were subjected to and the manner in which these tests were taken. The grades received by the participants of this study and the overall data collected from the platform are then shown in greater detail, after which we present the results brought forth by the statistical analyses made possible from the available data. Finally, the results of the current study are discussed, where possible reasons for some of its more surprising findings will be discussed.

Description of the platform Furago

Online format and contents

The e-learning platform on which the current study was conducted is designed to provide language instructors with exercises complementing activities already implemented in their classes. It was originally created in 2012 by one of the authors of this study (second author). It then underwent a series of transitions until 2016. At each stage, changes were made to modernize and improve the existing interface to foster motivation in its users and, ultimately, to facilitate its use alongside traditional in-class educational settings. It has been hosted since 2016 on a portal named www.furago.education. It is however scheduled to migrate to www.furago.io in 2023.

It contains about 2000 exercises that can be mapped out according to the lexical or grammatical progression of the particular textbook used in class. The platform was designed to meet several requirements. First, it was meant to provide a clear vision of the tasks to be carried out. Second, it had to allow for autonomous practice of the target language (French). Third, it was designed to boost regular reviewing of its contents. Furago users can indeed be asked to redo exercises: once new exercises (discovery phrases) have been discovered

and done for the first time, the reviewing process imbedded in the platform system might include them in the user's personal revision list (revision phrases), and bring them forth again at repeated intervals to ensure mastery of a particular set of lexical or grammatical items.

Intended use of the platform

Furago is designed to assist instructors of French as a second language by providing students with learning materials that can be used autonomously outside the classroom. To the most eager students, it indeed offers contents that can be practiced on freely for dozens of hours. The teachers can still however set, for those of their students who might not be as well disposed, particular objectives that are to be met along the course of the university semester. Such goals can consist in a minimum amount of points to be accumulated on the platform, or a given list of assignments to be completed before the end of the term. It should be noted that the platform is not meant to provide the teachers with qualitative feedback on their students' performance. The platform was designed under the assumption that, by simply interacting and engaging with them in class, each language instructor knows best his or her students' strengths and weaknesses. It therefore does not automatically generate specific remediation tools to address individual students' needs. Instructors can nevertheless, when assigning series of exercises to their students, choose specific aspects of the target language to be worked on, such as oral or written comprehension, vocabulary or grammar.

Three conditions for an effective and autonomous study

Learner autonomy being the ultimate goal of the platform, its creator identified three types of need for its prospective users. First, users need to be able to clearly identify the object of their study. Second, users need to be given regular feedback on how well or poorly they are doing. Third, users need to be presented with a revision plan that takes into account their particular shortcomings.

Exercises adapted to the contents studied in class. The platform covers the first of these needs by permitting language instructors to map the existing exercises into various lists. These lists closely follow the grammatical and/or lexical contents of the textbook used in class. Each series of exercises has a "tag" attribute which stands for a content or a skill label such as: "grammar," "phonetics," or "listening comprehension." These tags enable language instructors to create lists specifically tailored to their teaching objectives. Moreover, they can choose from more than 2000 exercises in the database. Screenshot 1 in Figure 1 shows a list of exercises created to match the contents of a French as a second language textbook named *Interactions*. In this example, we can see that the user has completed the three sets of exercises that were assigned to

the first lesson in the textbook. We can also see that the exercises of the other lessons remain to be done.

Precise and detailed feedback. The feedback requirement is easily met by presenting the correct answers as well as the transcripts (in the case of listening comprehension drills) whenever necessary. Figure 4 shows an auditory discrimination exercise during which users have to indicate how many times they hear the sound [ɛ]. It shows here that the user has failed to give the correct answer for an item and is given the option to listen to the sound again while reading the transcript of the item. This method of feedback also applies to listening comprehension exercises. The correct answers for failed items are shown along with the transcript of the audio.

Review management. The last condition for achieving learner autonomy is met through scheduling and enforcing revision sessions for users of the platform. Revisions are planned out along an SRS (Spaced Repetition System) which defines the two key features of the revision process. Based on the score achieved during the discovery phase and also on the score achieved during the revision phases, the SRS sets up a number of times the user will be presented anew with a given series of failed exercises, and also the duration of the increasing temporal spacing between two repetitions.

As for the activities for which no score can be established (e.g., listening, shadowing), it is the user himself who decides whether or not to add this activity to his revision list, and it is also he who decides the desired time spacing for this activity. It should be noted that the exercises given in revision mode are not timed. This way, users are given the option to consult external resources, such as the notes they took while taking their French class or a dictionary, to help them perform the exercise correctly. The reason why exercises are timed in the discovery phase is to allow the platform to make an evaluation of the user's prowess and, if necessary, to include the exercise in the user's revision list.

Fostering motivation

The platform was also designed to foster motivation in its users. Gamification artifacts play a great part toward achieving this objective. Here is how Kapp defines gamification within the context of a learning environment:

> On its surface, gamification is simply the use of game mechanics to make learning and instruction more fun. It seems 'fake' and artificial or like a short-cut. It's not. Underneath the surface is the idea of engagement, story, and meaning. Games give experiences meaning, they provide a set of boundaries within a "safe" environment to explore, think and 'try things out'. (Kapp, 2012, p. XXI)

On the platform, a number of artifacts not only allow users to compare each other's achievements, but also exceed their own potential. There are three sets of such artifacts:

Points. By doing exercises, users earn more or less points depending on how well they performed. The criteria used for assessing its performance are as follows: the ratio of correct answers they were able to achieve, the difficulty level of the exercise compared to the users' own learner's level (as assessed by the platform upon their first connection), the length of time under which they could perform the exercise (for some series of exercises only), and a "strike" bonus awarded to them by the platform when the exercise was perfectly done (i.e., when his correct answer ratio was 100%). Screenshot 2 in Figure 1 shows an example in the discovery phase during which a user obtained a score of 100% and completed the exercise 15 seconds before the time limit. This user was awarded 3.4 points. In the review phase, the user also earns points, again based on performance.

Groups and rankings. Upon registering on Furago, users are automatically and randomly assigned to a group of users (Screenshot 3 in Figure 1). This assignment process is repeated every month. If the teacher creates a specific class group however, this group assignment remains unchanged. Users can therefore rank themselves against the other users in their class, their real name however does not appear in the ranking. Appearing under a pseudonym in the class ranking allows, in the platform creator's opinion, emulation, but without the pressure and stress usually associated with competition. This is the "safe" environment that Kapp refers about.

Constraints. Any game has constraints that make it difficult, and therefore challenging. Players are consequently forced to develop a number of strategies to win the game. In the case of an e-learning platform, the constraints must first and foremost ensure the understanding and assimilation of the content taught, which are understandably the ultimate objectives of any teaching tool.

Levels. Users are only given access to exercises up to one level directly above their own. They are also given access to exercises of their own level or lower. To get access to content located two levels above their own, users must pass a test made up of exercises taken from the target level. This requirement to demonstrate one's level is a safeguard against the temptation users might feel to try to skip steps and miss out on some key learning opportunities.

Revision. Ensuring that users correctly assimilate the content they have worked on is crucial because it guarantees a progressive and structured learning process. However, students are often tempted to complete the series of exercises given as homework before, or without, doing the revisions scheduled by the SRS. In order to force users to go through the revisions, the platform forbids access to new exercises if the number of exercises available for revision is higher or equal to 15.

Characteristics of listening comprehension skills

As the current study aims at assessing how and if the use of Furago can help its users improve on their listening skills, an examination of what constitutes

progress in this basic language skill set is necessary. We will then appraise the added value the exercises and drills available on Furago can have for its users desirous of achieving that goal.

The cognitive model of listening comprehension as defined by Leclerq (Leclerq, 146-153) differentiates between "bottom-up processing," which derives meaning from linguistic forms, and "top-down processing," which processes meaning from contextual and social cues and other non-linguistic information. Bottom-up processing is broken down into different categories of cognitive processing steps which are ordered thus:

1. phonological decoding,
2. lexical processing,
3. syntactic analysis,
4. final discourse construction through elaboration of a situation model.

In this model, the higher categories encompass the lower categories. Phonological decoding is therefore mobilized all through bottom-up processing.

Let us now examine how the first three categories of bottom-up cognitive processing are mobilized on Furago when users perform exercises at levels A0 and A1, which are the first levels in the Common European Framework of Reference (CEFR), and also the levels of the participants in the present study.

Phonological decoding alone (speech extraction, abstraction of auditory input)

Auditory discrimination drills. Users hear a string of three syllables, each being made up of a different nasal vowel, such as /pɑ̃ pɛ̃ pɔ̃/. Users have then to choose the order in which they heard the nasal vowels by selecting an answer between, for example, these three groupings: /pɑ̃ pɛ̃ pɔ̃/, /pɑ̃ pɑ̃ pɔ̃/ and /pɛ̃ pɑ̃ pɔ̃/.

Phoneme identification drills. Users have to find how many times they hear a particular phoneme (/pɔ̃/ for example) in a string of syllables such as /pɛ̃ pe pɔ̃ pɑ̃ pɔ̃/.

Sound identification drills. Users have to identify meaningful words or segments of sentences extracted from a short conversation. The instruction for this type of exercise is "Listen to the dialogue and say if you hear the following elements."

Example of dialogue:

– Salut, ça va bien ?
– Oui, ça va, et toi ?
– Ça va pas mal, merci.

Examples of lexical items to be retrieved:

ça va bien ? [yes]
et vous ? [no → et toi ?]
ça va mal [no → ça va pas mal]
merci [yes]

Phonological decoding, lexical processing (word recognition) and syntactic analysis (semantic role recognition)

Word recognition exercises. Users must tell if a simple assertion is true or false. For example: “un chien est un animal,” “le ciel est bleu,” “un chat est une plante,” “un chat n’est pas une plante,” etc.

Classic listening comprehension exercises with a dialogue and its corresponding MCQ. Furago therefore clearly takes into account all three categories of the bottom-up processing steps that precede final discourse construction. Moreover, the platform specifically focuses on the most fundamental step for beginners: phonological decoding. Indeed, the platform offers about a hundred exercises and drills solely focused on phonemes discrimination. It should also be noted that these exercises, if not perfectly done during the discovery phase, will be automatically included into the users’ revision list and will therefore be submitted to them repeatedly until completion.

The freemium pricing model

Furago operates along two business models. One is a freemium business model that allows novice users to use the platform without a lot of limitations at first, but soon leads them toward a paid subscription. The free account allows users to do a fixed number of new exercises and revision exercises every 23 hours. As soon as a freemium user has done more than 50 different exercises (i.e., not including revision exercises), the number of new exercises that can be done every 23 hours gradually decreases. Although it is still possible to complete all the exercises on the platform with a free account, it quickly becomes more convenient for users to switch to a paid subscription. This 23-hour period has been chosen so that freemium users can still study every day, provided they do it each day at roughly the same time, say, from 7 to 8 pm for instance.

Figure 1. Screenshots of the platform in smartphone mode

Table 1. Access to exercises given to Freemium users

Number of exercises already done in discovery mode (excluding review mode)	Number of new exercises that can be done in discovery mode per 23 hours	Number of new exercises that can be done in revision mode per 23 hours
<50	20	5
<90	15	4
<120	10	3
<160	7	3
<200	6	2
<300	5	2
≥300	4	1

Coefficient of regularity

One of the dashboard functions made available for Furago users allows them to see how consistent they are in the use of the platform. It is a coefficient that appears as a percentage. Logging in and using the platform on average 4 days a week gives out a regularity coefficient of 100%. This coefficient is calculated over a maximum period of 90 days.

Types of data collected from the platform

Each week a listening comprehension tests was given, a backup of the database was correspondingly made, making the following data available:

- The identifier for each user
- The coefficient of the user's regular usage
- The total number of points accumulated by each user (i.e., his or her total Furago score)
- Their Freemium or Premium user status (i.e., paid subscription vs. free account holders)
- The number of listening exercises completed by each user

Backups were made on May 10 and July 5, 2021. In the following sections, data collected on these dates will be respectively referred to as data collected in Week 7 and Week 13 of the first university term.

Listening skills evaluation and subjects profiles

Participants

The current study was initially meant to measure the impact of the French e-learning tool Furago on students' language skills. The idea of assessing all four basic language skills (listening, reading, writing and speaking) was however soon abandoned in favor of listening skills only. This choice was made for three reasons. First, because listening skills can greatly vary from one individual to

another, even at the very early stages of foreign language acquisition, which could not be said about vocabulary recognition for example. Beginners have been exposed to too little of the target language (L2) for that particular ability to have a differentiating effect between individual learners. Second, because contrary to speaking skills, very little research is to be found on the subject in second language acquisition literature (Zoghlami & Hilton, 2021). And third, because the e-learning platform Furago had a wide variety of listening exercises that students could perform at leisure, on their own time and at their own pace. It seemed therefore particularly fitting to evaluate what progress could be made in this basic language skill thanks to this learning tool, especially considering that very little classroom time is usually allotted to listening drills and exercises, as the French language instructors that composed the research team knew from personal experience.

To that end, language instructors teaching at three major universities located in Tokyo, namely: the University of Tokyo, Waseda University and Tokyo University of the Arts, were asked to give two listening skills tests during the course of the first university semester. Given that most Japanese universities operate within a thirteen- to fifteen-week long semester, the first of these tests was scheduled in Week 7, the other in Week 13, which is to say mid-May 2021 for the first test and the beginning of July 2021 for the second test.

These tests were aimed at absolute beginners in French. All participants were first-year students who had yet to choose a major and who had chosen French for their mandatory second foreign language course. These students were selected amongst those enrolled in what is often called *enshu* classes (演習) in Japanese universities. As is often the case, these classes were taught by native speakers of the target language, in this instance French, in 90- to 105-minute sessions once a week, but were paired with a grammar-oriented class taught by a Japanese professor. This mode of selection was meant to ensure that, despite the fact that students were enrolled in different universities at the time of the study, their overall learner profiles would be as similar as possible. Moreover, it enabled researchers to leave out students taking intensive French courses or those who had previously learned French, in high school for example.

Instruments

Test format and contents. Each of these tests was composed of three listening exercises of increasing difficulty, graded out of 5 points each, for a total of 15 points (Appendices A and B). The tests were designed by the first author and then proofread and approved by the other member of the research team (the second author). Drawing from their personal experience as foreign language instructors, both authors were confident that the tests would give an accurate assessment of the listening skills of ordinary Japanese university freshmen learning French for the first time.

All of these listening exercises were borrowed from *L'atelier A1* workbook, which none of the subjects ever used in their class prior to the study. The exercises featured however standard grammatical and lexical progressions

that mirrored as closely as possible the textbooks used in class by the students who took part in the current study, namely: *Interactions 1*, *Conversation et grammaire*, *Spirale Nouvelle édition* and *Rythmes et Communication*. Typically, the first exercise would focus on phoneme recognition or auditory discrimination, while the other two required actual understanding of key elements of the audio file. They were therefore, in that respect, very comparable to the listening drills and listening comprehension exercises offered by the Internet platform Furago. In truth, these tests were more designed to assess the overall phonological awareness (for definition, see Saiegh-Haddad, 2019) of the subjects than their actual understanding of the linguistic message conveyed by the audio segment. Indeed, listening, if understood in the broadest sense, is highly context-sensitive and involves a series of complex and redundant processes, such as lexical segmentation and syntax parsing, to name but a few (Zoghlami & Hilton, 2021). Limiting the scope of the subjects listening tasks was meant as a precaution against biases that could be caused by other metalinguistic factors.

Given the teaching constraints imposed by the Coronavirus pandemic at the time of the study, both tests were given through Google Form (see Appendices A and B). The fact that this means of evaluation greatly facilitated data collection for the following statistical analysis obviously also played a part in the research team's decision to forgo the more traditional pen-and-paper assessment technique. To ensure that all subjects took the tests under near identical conditions however, professors who had agreed to take part in the study were strongly advised to take time out of their class to give the tests and to only play each audio segment twice, and this was indeed the case, by the research team's estimate, for 80 to 85% of the respondents.

Worried that this condition might prove too demanding for some of the teaching staff involved in the study, the research team used an add-on timer function to the Google Form tests. This enabled instructors to only give out the Google Form link to their students and ask them to do the test in their spare time at home. The Google Form add-on timer was set at 12 minutes, which only allowed respondents to play each audio segment twice if they wanted to answer all the questions within the allocated time frame. Contrary to the in-class setting favored by the research team, students who took the tests at home could not be monitored by their teacher (either in person or remotely via Zoom with the camera setting on) to prevent them from getting outside help. The research team is of the opinion however that the time constraint, coupled with the fact that each respondent was only allowed one try per questionnaire (i.e., an email address only granted one login opportunity), made effective cheating very difficult. To further prevent bias in the study, it also should be noted that the same member of the research team graded tests on both occasions.

Statistical tools and variables selection. All of the statistical analysis in the present study was conducted using R. R is a free software environment for statistical computing and graphics. It was chosen because it offered a wider variety of statistical tools and presented their results in more detail than Excel. Statistical analysis in the current study was carried out in two stages. The first

stage focused on correlations that could be drawn using four sets of variables[1]: two predictor variables that were binary: the use of Furago and the pricing model chosen by the subjects (**Freemium users** vs. **Premium users**); and three test variables that were continuous: the grades students obtained on the first then on the second listening test (**LCT1 Grade** and **LCT2 Grade**), points scored on the platform Furago by its users (**Furago Score**) and the coefficient of regularity generated by the platform for each of its users (**Regularity**). The pricing model chosen on the platform Furago differentiated therefore two sub-groups within the Furago users.

It should be noted that points scored on the platform Furago were collected twice: the first time in Week 7 when the subjects took LCT1 and the second time in Week 13 when the subjects took LCT2. This variable encompasses two separate data sets. In the following tables, charts and figures, relevant predictor variables will be mentioned under columns called **Week 7** and **Week 13**. The four sets of variables allowed for six different groupings whose means could then be compared two by two.

The second stage of the current study focused exclusively on Furago users. It aimed at drawing correlations between three sets of variables and the grades obtained by subjects on the first and second listening tests. The first two predictor variables were the same as in the first stage, that is to say: points scored on the platform Furago by its users up until testing week (**Furago Score Week 7** and **Furago Score Week 13**) and the coefficient of regularity generated by the platform for each of its users (**Regularity**). The third predictor variable referred to a series of listening drills and exercises performed by subjects on Furago up until the week of each test (**LCS**). To that end, simple linear regression models were adopted to assess the relationship between each of the three explanatory variables and **LCT1 Grade** and **LCT2 Grade**. A multiple linear regression model was also briefly considered but the predictor variables proved to be too intercorrelated to warrant this approach (See Appendix C). Whenever a correlation between a predictor variable and the test variable was deemed statistically significant, its coefficient was generated to foster further understanding of that particular parameter in the development of listening comprehension.

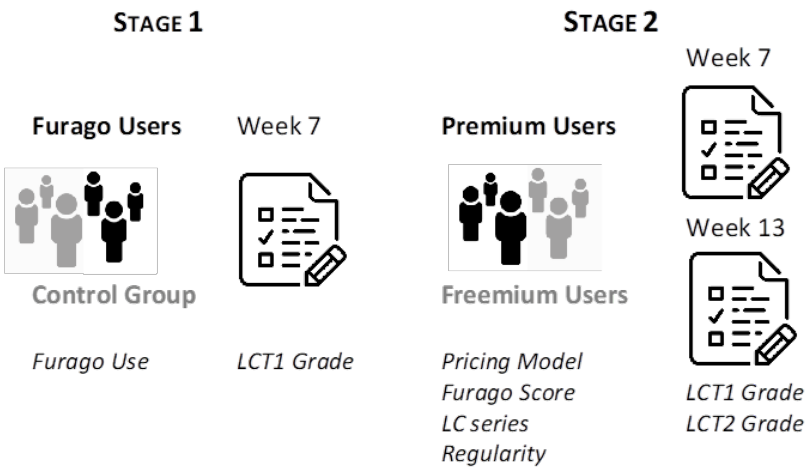


Figure 2. Experiment Design: Groupings and Variables

Data

Furago database

Table 2. Furago scores, LCT grade means & regularity depending on pricing model, Stage 2

	Week 7			Week 13		
	Furago score	**LCT1**	**Regularity**	**Furago score**	**LCT2**	**Regularity**
Freemium	516.129	8.838	0.335	1995.143	12.071	0.429
Premium	2176.565	10.516	0.535	5332.893	11.798	0.440

Google Form responses

Test group vs control group data. As shown in Figure 3, 206 students took LCT1, but only 173 took LCT2. This discrepancy is to be imputed to the fact that the later test was scheduled near the end of the university term. Not surprisingly, some of the language instructors who had initially agreed to participate in the study were pressed for time by their upcoming final examination. Therefore, they did not always have the time to either give the test during their class or remind their students to take the test at home. As will be discussed later in this article, this had the unfortunate consequence of depriving the researchers of a sufficient number of subjects in their comparison group for **LCT2**. Indeed, both LCT1 and 2 included a question on the use of Furago, and the Yes or No answer given to that question served as a binary and predictor variable in their first series of analysis. Grade means reported in Table 3 reflect this discrepancy further and show a higher mean for the control group in **LCT2**. Careful examination reveals however that the test group made up for more than 80% of the respondents for **LCT2** and therefore explains the near identical figures observed between the test group mean and the grand mean.

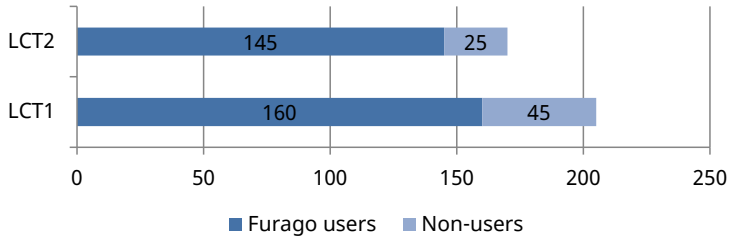


Figure 3. Total number of listening comprehension test participants, Stage 1

Table 3. Grade means for all listening comprehension test participants, Stage 1

	Furago users	Non-users	Grand mean
LCT1	8.99	7.82	8.74
LCT2	11.68	12.69	11.78

Linear regression model data. During the second stage of the current study, the research team experienced an unforeseen setback. When the time came to gather and compile the data generated by the platform Furago and the data that could be harvested from the Google Form responses, it soon became apparent that not all participants had used the same email address to set up their personal account on Furago and to sign in on Google Form for the listening tests. This difficulty was worsened further by the fact that the research team had offered the option to take LCT1 anonymously. It was indeed thought at the time that this alternative could increase participation amongst the comparison group. As a result, a significant portion of responses from both data sets could not be matched to individual subjects and ensure valid statistical analyses. With the exception of Figure 3 and Table 3 that show data used only for the first stage of the study, all figures recorded elsewhere in the article originated from a limited number of respondents. As shown in Figure 4 however, that still left enough subjects to conduct most of the intended statistical analysis, albeit with a somewhat reduced robustness.

Coupled with the fact that, as the university semester went on, more and more students switched to the Premium pricing model, the reduced number of respondents during stage 2, had also the unfortunate result of depleting the Freemium users' group. Since this ultimately only made up for 14 of the subjects (well under the threshold of 30 usually required for most statistical tests) in Week 13, the pricing model chosen for Furago by its users could no longer serve as a predictor variable when assessing other variables such as **LCT2 Grade**, **Furago Score** and **Regularity**. This variable was therefore only taken into account when analyzing data gathered in Week 7 (Tables 5 and 7).

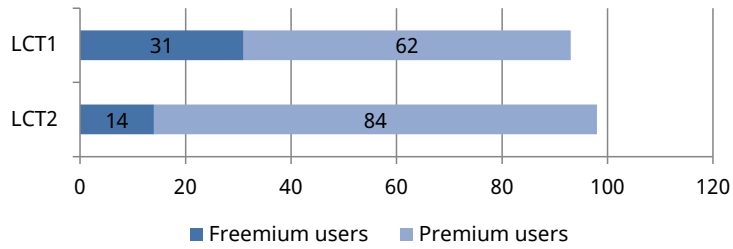


Figure 4. Listening comprehension tests participants, Stage 2

Table 4. All Furago users, Stage 2

	Furago scores	Grades	Regularity	LC series
LCT1	1623.09	9.96	0.47	10.02
LCT2	4856.10	11.84	0.44	23.13

Table 5. Freemium vs. Premium users, Stage 2

	Furago scores	LCT1 grades	Regularity	LC series
Premium	2176.56	10.52	0.54	12.87
Freemium	516.13	8.84	0.34	4.32

Analysis

Throughout the Test Group vs. Control Group analysis, T-tests were used whenever possible to determine whether the differences observed between means shown in the tables above held any statistical significance. However, when data was not found to be normally distributed, as was the case for the pricing model variable (Appendix D), Wilcoxon signed-rank tests had to be used instead. When correlations were deemed statistically significant (p<5%), they were also associated with a Pearson correlation coefficient to compare the force of the binary variables previously outlined between pair groups.

Stage 1: Test group vs. control group analysis

Results. As previously mentioned, stage 1 of the current study unfortunately produced very limited data. The control group was indeed too reduced in size to allow for valid statistical analysis.

Table 6. Test group vs. control group analysis, week 7, Stage 1

		Test variable
		LCT1 Grade
Predictor variables	Furago users	t-test p = 0.006 r = 0.193

Interpretation. Figures shown in Table 6 demonstrate beyond dispute that the use of Furago as a learning tool had a very clear and positive impact on the subjects' comprehension skills. Table 3 would, moreover, suggest that it accounts for a difference in mean of 1 point out of 15 on LCT1 (Cf. Table 3), which is to say a 6 to 7 points difference in a test graded out of 100. Though it would be tempting to recommend the use of Furago to all French language instructors on the basis of that margin of difference alone, it also should be noted that the correlation coefficient for this variable is very weak (< 0.2). We should therefore point out that it would be overly simplistic to assume that, ceteris paribus, an

average user of Furago is very likely to fare better than a non-user on listening comprehension tests. The following sections shall endeavor to bring forth a number of variables that also play a part on the grades achieved on LCT1 and 2 by Furago users.

Stage 2: Linear Regression Model

Predictor variable: Pricing model

Table 7. Premium users, week 7, Stage 2

		Test variables		
		Furago score	**LCT1 grade**	**Regularity**
Predictor variable	**Premium users**	Wilcoxon p = 0.000 r = 0.405	Wilcoxon p = 0.003 r = 0.323	Wilcoxon p = 0.001 r = 0.228

As shown in Table 7 above, the correlation between the pricing model and all three test variables is unquestionable. This is mostly due to the conditions under which the platform operates. Premium users can indeed use Furago a lot more regularly. They are consequently able to achieve higher scores than Freemium users within the same allotted time frame. Bearing in mind the commentary caution expressed in the previous section, we can safely say that Furago, if it should be used at all by French learners, shall be used by design as frequently as possible to bring satisfactory results. It should be remembered that Premium users' scores on Furago were almost four times higher than Freemium users' in Week 7 and that their grades on LCT1 were correspondingly more than 2 points higher (Table 5). If such a difference in grades is to be attributed to the use of Furago, as the following section will try to demonstrate further, this would make a very compelling argument for its use as a learning tool, provided that learners could benefit from the whole range of its features.

Predictor variables: Furago scores & listening comprehension series and LCT grades achieved in week 7 & 13.

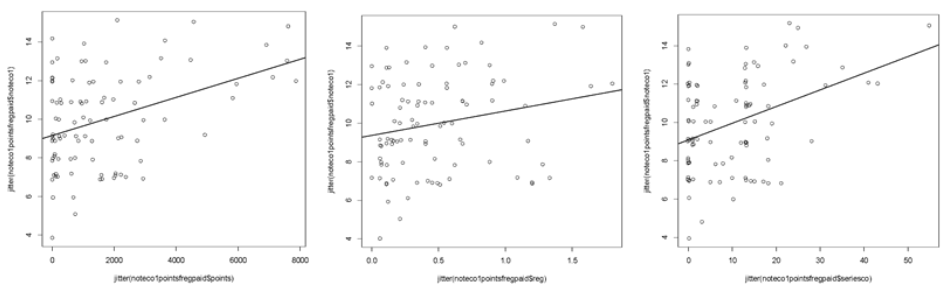


Figure 5. X axes: Furago score, regularity, listening comprehension series. Y axes: LCT1 grade

The scatter charts above were drawn from the data collected in Week 7 (the means of which can be seen in Table 2). They display the relationship between one variable, **LCT1 grade** (Y axes), and three explanatory variables which are, from left to right, **Furago Score**, **Regularity** and **Listening Comprehension Series** (X axes). They show, in other words, how subjects have fared on their first listening test, given their score, their consistency of use and the number of listening comprehension exercises they had achieved on the e-learning platform Furago by Week 7. Statistical analysis will be shown in the table below and subsequently discussed to determine whether the lack of clearly discernible patterns in these graphs (i.e., dots do not seem to cluster preferentially along the regression lines and, indeed, some of them could even be defined as outliers) is to be imputed to an absence of correlation between these variables or not.

Table 8. Linear model variables, week 7, Stage 2

		Test variables		
		LCT1 grade	**Furago score**	**Listening comprehension series**
Predictor variables	**Furago score**	p = 0.001 r = 0.390	NA	NA
	Listening comprehension series	p = 0.001 r = 0.385	p = 0.000 r = 0.912	NA
	Regularity	p = 0.040 r = 0.214	p = 0.000 r = 0.769	p = 0.000 r = 0.784

The table above calls for several explanatory remarks. Though **Furago Score** appears to be greatly correlated to the number of listening comprehension exercises and drills performed by its users, it is worth mentioning that teachers who require the use of Furago usually give their students two options. Students can either use the platform freely, take level checks and go through various exercises and drills as that are generated by the platform. Or they can do a specific series of exercises that teachers have selected for that particular class. Note that when students of the same class register on Furago, they also usually have to join the corresponding user group on the platform. This user group is created by their instructor at the beginning of the term and can be configured to foster practice on the specific series of exercises mentioned above. The fact that language instructors who agreed to take part in this study have all selected a series of exercises that focus on listening comprehension might therefore explain the strong correlation between **Furago Score** and **Listening Comprehension Series** (**LCS**). As shown in Figure 6 below, about a third of the exercises completed on the platform by the subjects of the study were indeed of that kind. Moreover, it is reasonable to assume that students who followed their teacher's advice in that respect were also very consistent in their use of

the platform. This in turn would explain the extremely strong correlation coefficient associated with **Regularity** and **Furago Score** and between **Regularity** and **LCS**.

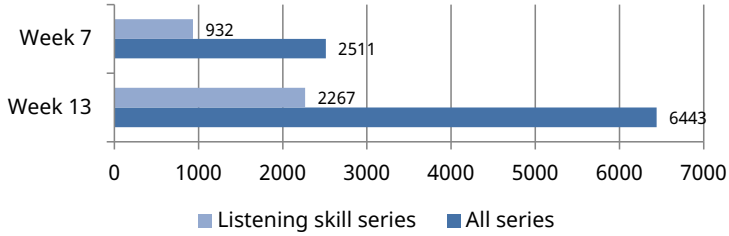


Figure 6. Total number of exercises by type, weeks 7 & 13

At this point in the reading of our study, it would probably be tempting for a language instructor to see Furago as the surest way to boost his students' listening skills and to consequently urge them to practice on this platform as often and as much as possible. Careful examination of the results provided by the linear model analysis conducted over R (Appendix E) shows however in the estimate column that, everything else being equal, a subject would need to collect as much as 2026 more points on Furago to raise his LCT1 Grade by one point. Yet, this is approximately the number of points that the most dedicated users of the platform (i.e., Premium users) would have reached, in average, by Week 7 (Table 5). The margin of progression is therefore a lot more limited than one would assume when looking at the table above.

As for **Regularity**, it is not surprising that it should be so clearly correlated to the score achieved by Furago users. It confirms what was already hinted at in the previous section regarding the consistency of use needed to achieve high scores on the platform.

What is more surprising however is the fact that, amongst all the variables considered, the score achieved on Furago seems to have the greatest impact on the grades received on LCT1. The correlation between these two variables is undoubtedly statistically significant ($p = 0.001$) and its coefficient is stronger than the other two correlations shown in Table 8. This would suggest that, contrary to a popular belief, held by a great number of language instructors, consistency might not matter as much as the overall amount of work students put into learning a given language. If confirmed, that finding could have a great many repercussions on the way foreign languages are being taught. As this type of research however does not fall within the scope of the current study and would require moreover statistical skills well beyond our present ability, we leave it to others to investigate this suggestion further empirically.

Table 9. Linear model variables, week 13, Stage 2

		Test variables		
		LCT2 grade	**Furago score**	**Listening comprehension series**
Predictor variables	**Furago score**	p = 0.007 r = 0.266	NA	NA
	Listening comprehension series	p = 0.042 r = 0.203	p = 0.000 r = 0.831	NA
	Regularity	p = 0.038 r = 0.211	p = 0.000 r = 0.711	p = 0.000 r = 0.621

Just as for Week 7, the correlations between predictor variables and test variables are statistically significant, thus confirming the interpretations made during the analysis conducted on data collected in Week 7. The estimate column provided by the linear regression model analysis (Appendix F) reveals however that, everything else being equal, an average user of Furago would need to score 4333 additional points on the platform to see a 1-point increase on his LCT2 Grade. Given that Furago users' score only amounted to an average of 4856 points in Week 13 (Table 4), that amount of dedication is unlikely to be seen in any but the most committed students.

There is however an overall trend that should be addressed. Correlations between the predictor variables and **LCT2 Grade** appear to be somewhat less statistically significant (albeit still within the acceptable range) and their respective coefficients are correspondingly not as strong as for Week 7. We hypothesize that this difference might be imputed to the fact that Furago, as far as listening skills are concerned, might have a greater impact during the earlier stages of foreign language acquisition. This interpretation obviously relates to the broader question of language learning curves and the corresponding challenge of modeling them accurately. Daller, Turlik, and Weir (2013) have noticeably reported on a body of literature pertaining to these issues as part of their work on vocabulary acquisition. They, too, observed that the rate of improvement of their subjects is reduced as their practice continues. In their article, the authors debate what kind of model would best describe language learning curves in general, namely what Fitts and Posner (1967) refer to a "power curve" or what Choi, Harring, and Hancock (2009) refer to an "S-shaped curve[2]." While our findings do seem to concur with the non-linear plateaued model, they advocate to account for language learners' growth rate in most education settings, our only two points of measurement (i.e., data collection in Week 7 and Week 13) are clearly insufficient to rule on one of the more specific models. Further study, drawing from more longitudinal data, would therefore be needed to better visualize the underlying process that governs our subject's learning curve. For the time being, we can only surmise that, as the semester progresses, students are likely to get more and more attuned to the

phonological features of the target language which, contrary to its lexical or its grammatical components, are finite. Achieving an (almost) complete mastery of a given language phonology might be therefore more within the beginners' grasp. And once subjects have acquired a sufficiently good "ear" for their target language, they might have plateaued, which would make improving further on their listening skills more time-consuming and/or challenging.

If that were the case, how could we then account for the fact that the LCT2 grand mean was higher by two points than LCT1's (Table 3)? The reason for these paradoxical results may lie in the fact that listening skills only partly rely on the subject's phonological awareness, and that phonological awareness itself may be improved on by other means than auditory discrimination or phoneme recognition drills. Indeed, practicing the other three basic language skills (i.e., reading, writing, and speaking) can benefit the fourth (Aydogan & Akbarov, 2014). It is reasonable to assume for example that, as the subjects' understanding of the target language vocabulary expands, the meaning of new words can be more and more easily inferred by context and/or by closely related words previously encountered (Aydogan & Akbarov, 2014). What is more, improved speaking skills and lexicon growth can be simultaneously beneficial to phonological awareness (Saiegh-Haddad, 2019).

Finally, this would explain, at least in part, why the correlation coefficient between **LCS** and **LCT Grades** is on both occasions slightly weaker than the correlation coefficient between **Furago Score** and **LCT Grades**. Even though **LCS** is highly correlated to **Furago Score**, we would expect the former to be a more important factor on the results of tests that focus solely on the listening skills of the subjects. If, as we surmised earlier, the type of exercises performed by students is not as strong a predictor of the type of language skills they develop as language instructors are intuitively tempted to believe, the lists of exercises that instructors recommend to their students need to be reconsidered in their contents. The exercises listed there should be chronologically ordered and focus less and less on listening and more and more on other language skills as the university semester proceeds. Expecting students to devote a third of their study time on listening exercises alone (Figure 4) might have the counterproductive effect of depriving them of more beneficial ways to improve their listening skills.

Conclusion

Before recapitulating the principal findings of the current study, several of its limitations should be reiterated. Regarding the first stage of the study, it is regrettable that the diminutive size of the control group did not allow for a follow-up analysis of the key variable **Furago Use** on the data gathered in Week 13. In retrospect, the research team feel that they should have more actively pursued the collaboration of other French instructors that did not use Furago. This could have ensured that, despite last minute withdrawals from the study, a minimum of 30 subjects were conserved in the control group. Regarding the second stage of the study, it was also very unfortunate that about a third of the

Google Form responses could not be matched to individual users listed on the Furago database because subjects very often used a different email address on each platform. That technical difficulty too, if foreseen, could have been easily avoided.

The results brought forth by the analysis that the available data allowed were however very encouraging. Correlations between predictor variables such as **Furago Use**, **Furago Score**, **Regularity**, and so forth, and the test variable **LCT1** and **2 Grades** were all found to be statistically significant, albeit with correlation coefficients that were rather weaker than expected (i.e., within a 0.2 to 0.4 range). Such results therefore confirmed the value of Furago as a means to improve students' listening skills, especially, it would seem, at the very beginning of foreign language acquisition. One might regret however that Furago users have a very narrow margin of progression, in terms of how much their listening test grades are likely to be improved by their extra work on the platform. The fact that this teaching device does not burden language instructors with any additional guidance, correcting or counselling duties however clearly compensates for that rather low return on investment on the part of the students. And since scores achieved on Furago can be seen as a clear indication of the users' listening skills, it can be made part of the overall evaluation process put in place by the institution as well, on the condition however that students all register as Premium users to fully benefit from the freedom of use this pricing model can offer.

Finally, this study yielded some interesting insights on the respective impacts of the overall quantity of effort students put in their study (reflected here by their scores on Furago) and the consistency rate (i.e., **Regularity**) that characterizes their study. As this was not the intended objective of the study however, the hypotheses briefly outlined on the subject are only to be taken as mere suggestions. Just as for the impact of the use of Furago on the other three basic language skills (reading, writing, and speaking), empirical studies that would encompass wider groups of subjects and take a more longitudinal approach are clearly needed. Such studies could moreover address the gamification parameter that is an inherent part of the concept under which Furago operates but that we did not discuss when interpreting the results of our study. By enabling its users to compete against each other for ranks and to level up after passing periodical checks, Furago does indeed try to foster the students' motivation and consequently ensure that they continue using it, despite the feeling of loneliness generally associated with e-learning tools used in unaccompanied settings.

Notes

1. When used in reference to a variable, items will appear in bold type. Ex: "LCT1" refers to the first listening comprehension test students took on week 7, while "**LCT1 Grade**" refers to the test variable.
2. Contrary to a "power curve," an "S-shaped curve" shows a slow growth at the beginning, which is then followed, as for the "power curve," by an

acceleration of the learning process, and then a flattening of the curve toward a plateau.

References

Aydogan H., & Akbarov A.A., (2014). The four basic language skills, whole language & integrated skill approach in mainstream university classrooms in Turkey, *Mediterranean Journal of Social Sciences*, *5*(9), 672-680.

Choi J., Harring J.R., & Hancock G.R., (2009). Latent growth modeling for logistic response functions, *Multivariate Behavioral Research, 44*(5), 620-645.

Cocton M-N., Pommier E., Ripaud D., & Rabin M. (2019). *L'atelier A1 : Cahier d'activités.* Paris: French and European Publications Inc.

Crépieux G., Massé O., & Rousse J-P. (2019). *Interactions* 1 (2nd ed.). Paris: Cle international.

Crépieux G., & Callens P. (2015). *Spirale nouvelle édition*. Paris: Hachette.

Derible A., Fife B., Terada T., & Blanchin Fujita J. (2017). *Rythmes et communication.* Tokyo: Asahi Press Ltd.

Daller M., Turlik J., & Weir I. (2013). Vocabulary acquisition and the learning curve. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 185-218). Amsterdam: John Benjamins.

Fitts P. M., & Posner I. M., (1967). *Human performance.* Belmont, CA: Brooks/ Cole Publishing.

Leclercq P., Edmonds A., & Sneed German E. (2021). *Introduction à l'acquisition des langues*. Louvain-la-Neuve, Belgium: De Boeck Supérieur.

Ohki M., Azra J.-L., & Vannieuwenhuyse B. (2007). *Conversation et Grammaire.* Kyoto: Alma Publishing.

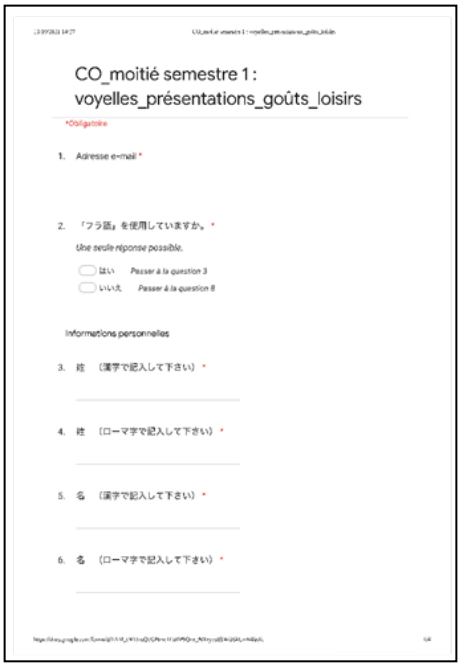
Saiegh-Haddad E. (2019). What is phonological awareness in L2? *Journal of Neurolinguistics, 50*, 17-27.

Kapp, K.M. (2012). *The gamification of learning instruction: game-based methods and strategies for training and education.* Hoboken, NJ: Pfeiffer.

Zoghlami, N. & Hilton, H. (2021). La compréhension de l'oral en langue étrangère. In P. Leclercq, A. Edmonds & E. Sneed German (Eds.), *Introduction à l'acquisition des langues etrangères.* (pp. 143-160). Louvain-la-Neuve : De Boek Supérieur.

Appendix A

Listening comprehension test 1, week 7



CO_moitié semestre 1 :
voyelles_présentations_goûts_loisirs

*Obligatoire

1. Adresse e-mail *

2. 「フラ語」を使用していますか。 *

Une seule réponse possible.

- はい Passer à la question 3
- いいえ Passer à la question 8

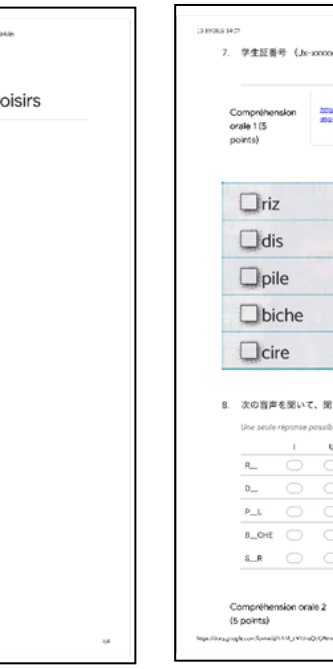
Informations personnelles

3. 姓 （漢字で記入して下さい） *

4. 姓 （ローマ字で記入して下さい） *

5. 名 （漢字で記入して下さい） *

6. 名 （ローマ字で記入して下さい） *



7. 学生証番号 （Je-xxxxxx の形式で記入してください） *

Compréhension orale 1 (5 points)

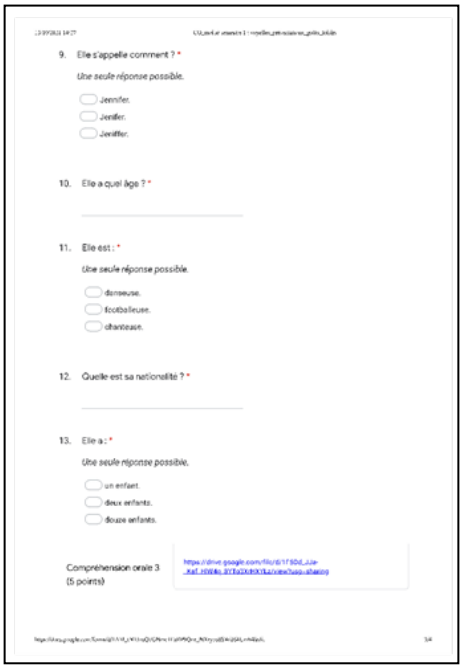
☐ riz	☐ rue	☐ roue
☐ dis	☐ du	☐ doux
☐ pile	☐ pull	☐ poule
☐ biche	☐ bûche	☐ bouche
☐ cire	☐ sûr	☐ sourd

8. 次の音声を聞いて、聞こえた方をチェックしましょう。 *

Une seule réponse possible par ligne.

	I	U	OU
R__	◯	◯	◯
D__	◯	◯	◯
P__L	◯	◯	◯
B__CHE	◯	◯	◯
S__R	◯	◯	◯

Compréhension orale 2 (5 points)



9. Elle s'appelle comment ? *

Une seule réponse possible.

- Jennifer.
- Jenifer.
- Jeniffer.

10. Elle a quel âge ? *

11. Elle est : *

Une seule réponse possible.

- danseuse.
- footballeuse.
- chanteuse.

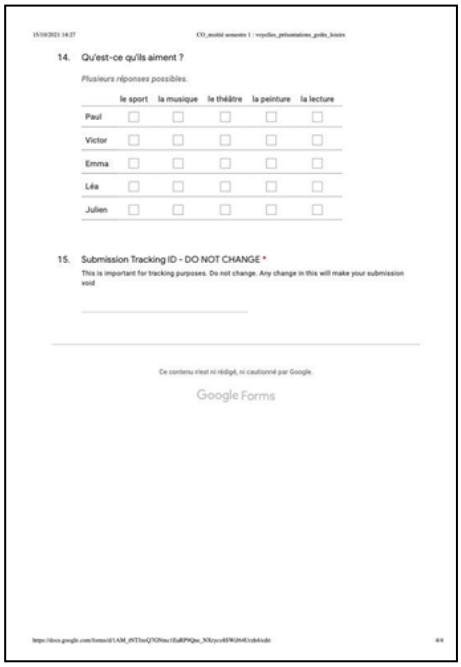
12. Quelle est sa nationalité ? *

13. Elle a : *

Une seule réponse possible.

- un enfant.
- deux enfants.
- douze enfants.

Compréhension orale 3 (5 points)



14. Qu'est-ce qu'ils aiment ?

Plusieurs réponses possibles.

	le sport	la musique	le théâtre	la peinture	la lecture
Paul	☐	☐	☐	☐	☐
Victor	☐	☐	☐	☐	☐
Emma	☐	☐	☐	☐	☐
Léa	☐	☐	☐	☐	☐
Julien	☐	☐	☐	☐	☐

15. Submission Tracking ID - DO NOT CHANGE *

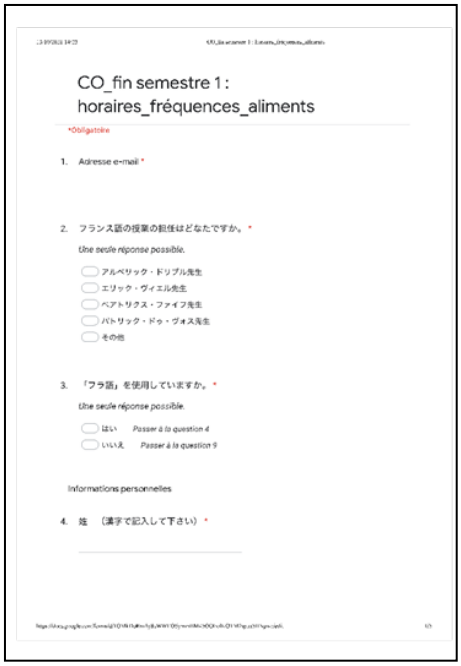
This is important for tracking purposes. Do not change. Any change in this will make your submission void

Ce contenu n'est ni rédigé, ni cautionné par Google.

Google Forms

Appendix B

Listening comprehension test 2, week 13



CO_fin semestre 1 :
horaires_fréquences_aliments

*Obligatoire

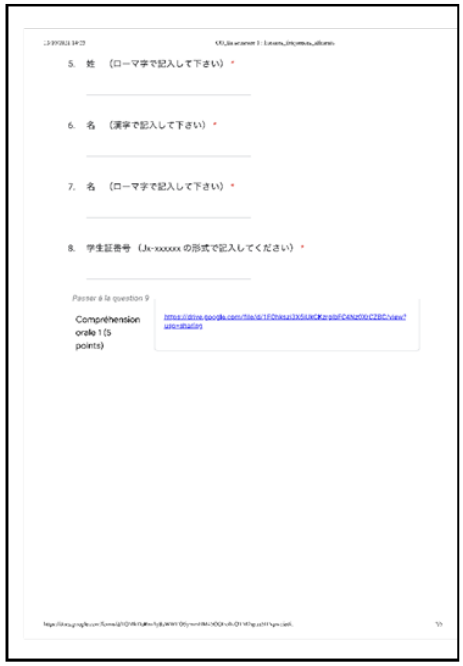
1. Adresse e-mail *

2. フランス語の授業の担任はどなたですか。 *
Une seule réponse possible.
- アルベリック・ドリブル先生
- エリック・ヴィエル先生
- ベアトリクス・ファイフ先生
- パトリック・ドゥ・ヴォス先生
- その他

3. 「フラ語」を使用していますか。 *
Une seule réponse possible.
- はい Passer à la question 4
- いいえ Passer à la question 9

Informations personnelles

4. 姓 （漢字で記入して下さい） *



5. 姓 （ローマ字で記入して下さい） *

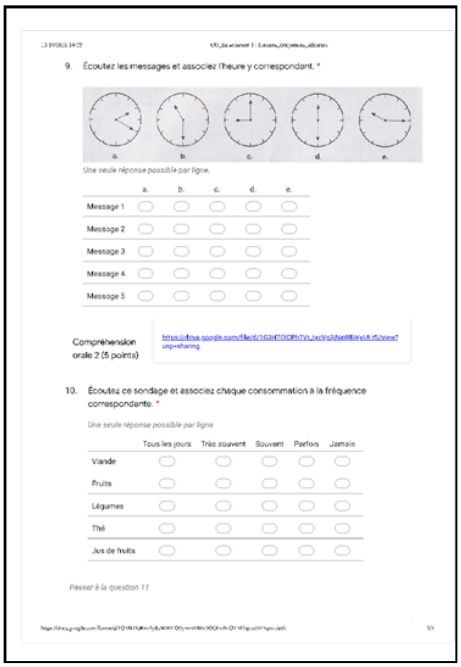
6. 名 （漢字で記入して下さい） *

7. 名 （ローマ字で記入して下さい） *

8. 学生証番号 （Jx-xxxxxx の形式で記入してください） *

Passer à la question 9

Compréhension orale 1 (5 points)



9. Écoutez les messages et associez l'heure y correspondant. *

Une seule réponse possible par ligne.

	a.	b.	c.	d.	e.
Message 1					
Message 2					
Message 3					
Message 4					
Message 5					

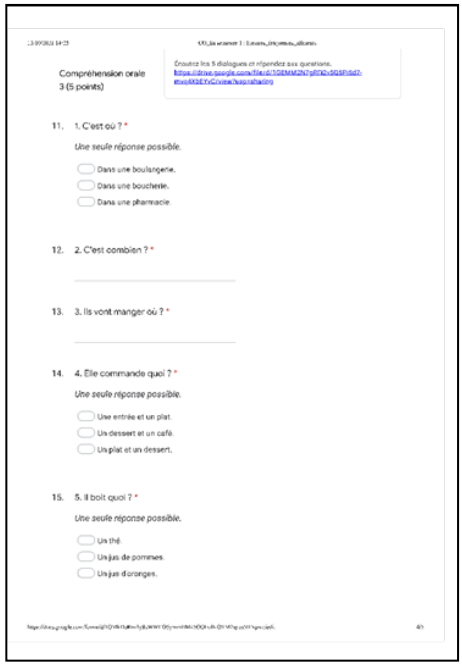
Compréhension orale 2 (5 points)

10. Écoutez ce sondage et associez chaque consommation à la fréquence correspondante. *

Une seule réponse possible par ligne.

	Tous les jours	Très souvent	Souvent	Parfois	Jamais
Viande					
Fruits					
Légumes					
Thé					
Jus de fruits					

Passer à la question 11



Compréhension orale 3 (5 points)

11. 1. C'est où ? *
Une seule réponse possible.
- Dans une boulangerie.
- Dans une boucherie.
- Dans une pharmacie.

12. 2. C'est combien ? *

13. 3. Ils vont manger où ? *

14. 4. Elle commande quoi ? *
Une seule réponse possible.
- Une entrée et un plat.
- Un dessert et un café.
- Un plat et un dessert.

15. 5. Il boit quoi ? *
Une seule réponse possible.
- Un thé.
- Un jus de pommes.
- Un jus d'oranges.

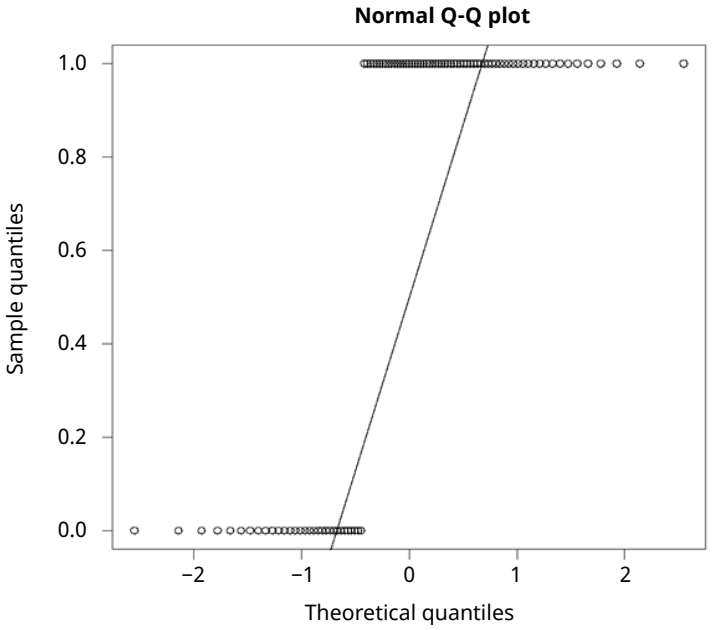
Appendix C

Intercorrelation testing

```
Call:
lm(formula = noteco1 ~ points + ttseries + seriesco + regularity *
    points + ttseries + seriesco + regularity, data = noteco1pointsfregpaid)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3668 -1.6790 -0.2808  1.8536  4.7355

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       9.4942317  0.4295285  22.104   <2e-16 ***
points            0.0009575  0.0007010   1.366    0.176
ttseries         -0.0580233  0.0562285  -1.032    0.305
seriesco          0.1307069  0.0827748   1.579    0.118
regularity       -2.1237405  1.0917560  -1.945    0.055 .
points:regularity  0.0001170  0.0002454   0.477    0.635
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.267 on 87 degrees of freedom
Multiple R-squared:  0.1959,  Adjusted R-squared:  0.1497
F-statistic: 4.239 on 5 and 87 DF,  p-value: 0.001698
```

Appendix D

Normal probability plot for the pricing model variable



Appendix E

Linear regression model analysis, LCT1, Stage 2

```
Call:
lm(formula = noteco1 ~ points, data = noteco1pointsf)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1558 -1.5103 -0.1677  1.6996  4.8422

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.1558224  0.3080757  29.719  < 2e-16 ***
points      0.0004936  0.0001220   4.044  0.00011 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.275 on 91 degrees of freedom
Multiple R-squared:  0.1524,  Adjusted R-squared:  0.143
F-statistic: 16.36 on 1 and 91 DF,  p-value: 0.0001097
```

* 1 point on the Furago score accounts for 0.0004936 point on LCT1 Grade

Appendix F

Linear regression model analysis, LCT2, Stage 2

```
Call:
lm(formula = noteco2 ~ points, data = noteco2pointsf)

Residuals:
    Min      1Q  Median      3Q     Max
-6.9872 -1.3650  0.5706  1.9845  3.6574

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.074e+01  4.765e-01  22.545  < 2e-16 ***
points      2.308e-04  8.405e-05   2.746  0.00716 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.539 on 99 degrees of freedom
Multiple R-squared:  0.0708,  Adjusted R-squared:  0.06141
F-statistic: 7.543 on 1 and 99 DF,  p-value: 0.007158
```

* 1 point on the Furago score accounts for 0.0002308 point on LCT2 Grade