Self- and Peer-Assessment of Student-Generated Podcasts to Improve Comprehensibility in Undergraduate EFL Students

Autoevaluación y evaluación por pares de *podcasts* generados por alumnos para mejorar la comprensibilidad de estudiantes del pregrado de inglés como lengua extranjera

Albedro Cadena-Aguilar
Claudia Patricia Álvarez-Ayure
Universidad de La Sabana, Chía, Colombia

This study reports on a mixed-methods research project into self- and peer-formative assessment of student-generated podcasts in a group of 18 undergraduate students. The aim was to determine whether there were any gains in the spoken comprehensibility of the participants while having them reflect on and adjust their use of suprasegmentals (thought groups, sentence stress, and intonation). Data were gathered from student logs, student-generated podcasts, and a questionnaire. Results unveiled the exhibition of self-regulated behaviours and gains in comprehensibility. This study highlights the importance of helping learners look critically and reflectively at their own oral production and of incorporating training on suprasegmentals within English as a foreign language courses to help learners communicate more effectively within a globalised context.

Keywords: English as a foreign language, formative assessment, higher education, podcasts, self-directed learning, spoken comprehensibility

Este estudio reporta una investigación de métodos mixtos sobre autoevaluación y evaluación formativa por pares de *podcasts* generados por 18 estudiantes de pregrado. El objetivo fue determinar si había algún aumento en la comprensión hablada de los participantes mientras se les hacía reflexionar y ajustar su uso de suprasegmentos (grupos de pensamiento, acentuación en oraciones y entonación). Los datos se obtuvieron de registros y *podcasts* generados por los estudiantes, y de un cuestionario. Los resultados muestran comportamientos autorregulados y aumento en la comprensión. Este estudio resalta la importancia de ayudar a los estudiantes a examinar crítica y reflexivamente su producción oral e incorporar capacitación en suprasegmentos en cursos de inglés como lengua extranjera para ayudarlos a comunicarse efectivamente en un contexto globalizado.

Palabras clave: inglés como lengua extranjera, aprendizaje autodirigido, comprensión del discurso hablado, educación superior, evaluación formativa, podcasts

Albedro Cadena-Aguilar https://orcid.org/0000-0002-0633-3856 · Email: albedro.cadena@unisabana.edu.co
Claudia Patricia Álvarez-Ayure https://orcid.org/0000-0001-7425-6437 · Email: claudiap.alvarez@unisabana.edu.co

Introduction

The incorporation of formative assessment practices into the English as a foreign language (EFL) classroom has been considered an essential element to promote learning and achievement (Harlen & Winter, 2004; Hattie & Timperley, 2007; van de Watering & van der Rijt, 2006). However, in higher education EFL contexts, assessment often remains dependent on teachers (Yorke, 2003), a situation that may hinder the development of self-regulation and lifelong learning skills (Nicol & Macfarlane-Dick, 2006) and deprive students from taking a more active role in their learning process.

At the same time, much English language instruction is still oriented towards classroom exchanges, with excessive emphasis on achieving native-like pronunciation (Derwing & Munro, 2015; Levis, 2018; Macías, 2010; Tahereen, 2015). This focus is out of step with processes of internationalisation in higher education; these seek to integrate international, inter-cultural, or global dimensions (Knight, 2003) with knowledge, attitudes, and values that help learners compete in global marketplaces (Green & Shoenberg, 2006). Given the status of English as an international language (Crystal, 2003; Jenkins, 2015; Jenkins et al., 2018), learners should focus more on achieving mutual comprehensibility with English speakers from different backgrounds than on achieving native-like pronunciation (Bøhn & Hansen, 2017; Jenkins, 2015; Tahereen, 2015; R. Walker, 2010).

Accordingly, the present study examined how self- and peer-assessment of student-generated podcasts influenced participants' comprehensibility. In this light, the study was guided by two research questions: (RQ1) Does the use of formative assessment practices help students improve their comprehensibility when recording a podcast? (RQ2) What is the effect of training students in the use of suprasegmental features seeking comprehensibility?

Theoretical Framework

Self- and Peer-Assessment

The shift from teacher-centred assessment to student self-assessment reports significant gains when the latter is used as a formative rather than a summative procedure and when students are guided properly (Ross, 2006; Sargeant, 2008). In this light, this study abides by the definition provided by Andrade and Du (2007), who explain self-assessment as a formative process through which "students collect information about their own performance and see how it matches their goals and/or the criteria for their work" (p. 160). Nevertheless, self-assessment must be attempted only upon prior and proper coaching where not only do students get acquainted with assessment instruments and criteria but also become involved in their construction along with training on their application within a safe learning environment.

Despite challenges with anxiety levels and concerns about the reliability of student feedback in peer-assessment (Topping, 1998; A. Walker, 2001), it has been found that peer feedback has positive effects on the development of higher-order thinking, adoption of positive attitudes towards lifelong learning, and enhancement of social attitudes and trust in others (Hamer et al., 2015; Noonan & Duncan, 2005). In this light, Zimmerman (1990) claims that students can become more effective learners as they self-regulate motivational, metacognitive, and behavioural aspects of their learning. Similarly, Chong (2016) notes benefits for both givers and receivers of peer feedback, observing that "those who assess their peers would use their strengths and would feel more confident giving feedback. On the other hand, those who are assessed would think that the peer assessors are more knowledgeable than they are in the area" (p. 22). The current study considers this mutual benefit as a key asset in consistent formative assessment practices (Boud & Molloy, 2013; Harlen,

2007) with evaluation instruments that are sufficiently flexible to adapt to students' strengths (Chong, 2016) and facilitate students' continuous involvement in and reflection on the learning process. The present study examined the participants' perceptions on self- and peer-assessment practices as they engaged in the recording of podcasts to improve spoken comprehensibility (Derwing & Munro, 2015).

Comprehensibility

The present study adopts Derwing and Munro's (2015) conceptualization of comprehensibility: "the ease or difficulty a listener experiences in understanding an utterance" (p. 5). This conceptualization attains the target language (L2) practitioners' purpose of helping learners achieve intelligible pronunciation which is paramount in their communicative success in day to day interaction with L2 speakers. Specifically, our research argues that the comprehensibility of oral discourse should be evaluated in terms of how easily both native English speakers (NESs) and non-native English speakers (NNESs) understand it. Similarly, Jenkins et al. (2018); Jenkins (2015); Levis (2018); Macías (2010); and Tahereen (2015) agree that students need not seek native-like pronunciation provided they are able to communicate with a wide range of English speakers. The present study argues that the enhancement of comprehensibility should be one of the priorities in L2 pronunciation lessons and that instruction on suprasegmentals (sentence stress intonation, thought groups) can be an effective way to achieve that objective.

Podcasts

Previous research (Campbell, 2005; Ng'ambi & Lombe, 2012; Popova & Edirisingha, 2010) has shown that using student-generated podcasts (SGPs) in combination with self- and peer-formative assessment can enhance engagement with learning materials and promote further learning beyond the classroom. These affordances are common in EFL and English as a second language context, where SGPs are frequently used to support the development of speaking and listening skills (Farangi et al., 2015), foster group work (Ng'ambi & Lombe, 2012), and promote interactive student-student environments (Rosell-Aguilar, 2013). However, such studies also show there has been little work on incorporating SGPs into formative assessment methodologies while training students to participate effectively in them. In the present study, podcasts were used as an instrument to help learners rehearse, practice, and evaluate their spoken production—practices necessary at early stages of EFL language development.

Method

Context and Participants

This study was conducted at a private university in Colombia. The participants consisted of 18 native Spanish speakers (11 male, seven female), aged 17 to 23, with an A2–B1 CEFR level (Council of Europe, 2011) previously determined by means of a Cambridge placement-L2 English test. All were full-time students in different degree programmes taking a 64-hour English-language course in which they met for four hours each week. However, analysis of diagnostic tests, classroom observations, and informal interviews with participants showed that, overall, their oral production was *basically incomprehensible* for their classmates and the course instructor mainly because their discourse presented salient issues in the formation of thought groups and in the use of sentence stress and intonation. This situation limited their chances for successful communication with native and non-native speakers in everyday informal exchanges.

Pedagogical Intervention

Towards the Enhancement of Comprehensibility

To support the participants' development of comprehensibility while generating their podcasts, we taught

using a blended-learning strategy over five months, both inside the classroom (20 hours) and in an online learning environment (20 hours). During that time, participants were guided to analyse models of podcasts about mishaps to: (a) become aware of their structure; (b) identify commonly used verbal tenses; (c) familiarize themselves with useful language; and (d) identify suprasegmental features.

To help participants focus on the analysis and use of suprasegmental features, they were first made aware of the rhythmic differences between their native language (Spanish) and the language of instruction (English). They realized Spanish is a syllable-timed language, in which all syllables have equal length, while English is stress-timed, with equal intervals between stressed syllables (Abercrombie, 1967; Celce-Murcia et al., 2010). Accordingly, we used mark-up as in the following example to indicate differences between the rhythms of both languages.

.   •   .   •   .   •

(1) The girl was looking for help

. • . • . . • . . • .

(2) La joven estaba buscando ayuda

Second, participants began work on suprasegmental features by identifying thought groups. Participants used scripts from model podcasts about mishaps to identify groups of words that represented semantically and grammatically coherent units of meaning (Celce-Murcia et al., 2010). Then, they listened to recordings to verify that the thought groups they had identified coincided with the pauses made by the speaker. Afterwards, to train participants in the use of sentence stress, their attention was directed to content and function words in sentences or utterances to identify those with prominent stress. To help participants become familiar with intonation, they listened to model podcasts while marking the scripts with upward arrows for rises in intonation and downward arrows for falls.

Students' Self- and Peer-Assessment Practices

After their training in suprasegmental features, participants engaged in a formative assessment process (Figure 1) consisting of self-and peer-assessment. Initially, participants recorded a podcast about a mishap using the models previously presented. Then, they became familiar with assessment criteria through a rubric (see Appendix A) designed for evaluating comprehensibility and self-assessing their own podcasts. Using the same rubric, they then peer-assessed their classmates' podcasts. The rubrics provided space for students to write comments on the aspects that went well and on those that needed improvement. This cooperative learning and the assessment process were intended to accustom the participants to providing and receiving feedback from different sources whilst "developing their capacities in monitoring and evaluating their own learning" (Boud & Molloy, 2013, p. 704). The process finalized when participants re-recorded their podcasts, seeking improvements in thought groups, sentence stress, and intonation by using their self and peer-assessments.

Instruments and Data Collection

Data needed for RQ1 were collected from student logs and a questionnaire (see Appendix B). The student logs, used throughout the pedagogical intervention and applied after each self-and peer-assessment activity, guided the participants to provide insights into their experiences (Cohen et al., 2007); their reflection was guided by a set of predesigned questions. The questionnaire, administered at the end of the intervention, focused on participants' viewpoints in the role of formative assessment in developing their comprehensibility using a five-point Likert scale.

To measure the effects of formative assessment practices on the participants' comprehensibility levels, all of

Figure 1. Formative Assessment Process
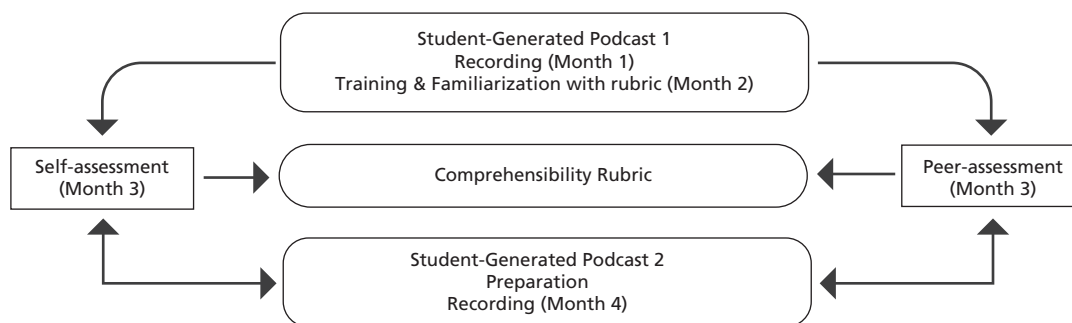
them were required to record a podcast about a mishap both at the beginning and end of the study. Afterwards, to address RQ2, the podcasts were transcribed to analyse the participants' discourse and observe any changes in their use of suprasegmental features. Additionally, these podcasts were subjected to the evaluation of comprehensibility by external evaluators, and the results were stored in a matrix for subsequent analysis.

Evaluating Comprehensibility

Four evaluators—the course tutor and three other evaluators external to the course—with varied educational backgrounds, representing each of Kachru's (1985) circles and with C2 (CEFR) English proficiency levels, evaluated the levels of comprehensibility in the SGPs. The course tutor was a local non-native English teacher (NNEST1) from the expanding circle (i.e., countries where English has no historical or governmental role, but where it is used as a medium of international communication; Kachru, 1985) with a postgraduate degree in education. The first external evaluator was a native English speaker (NES) from the inner circle (i.e., countries with traditional bases of English and its speakers; Kachru, 1985). The second external evaluator was a non-native English-speaking teacher (NNEST2) from the outer circle (i.e., places where non-native official varieties of English are spoken because of their colonial history; Kachru, 1985) with a broad teaching background. The third external evaluator was a local

expert (EXP) and a language teacher trainer from the expanding circle.

The evaluators evaluated both SGP1 and SGP2 to determine their levels of comprehensibility using the criteria shown in Table 1. The analysis of the results from these evaluations allowed us to answer RQ2.

Data Analysis

In the analysis of data related to RQ1, the participants' logs were transferred to MAXQDA (VERBI Software, 2018) for better visualization and management. The grounded theory approach (Corbin & Strauss, 2015) was used to classify sets of data into patterns, determine subcategories, and establish a main category. Reliability of the coding was checked by a fellow qualitative researcher who also analysed blind-coded students' excerpts. This blind-coder and the authors coincided on 80% of the coded data and then agreed on the best coding for the remaining 20% of the information. The quantitative data obtained from the questionnaire were used to corroborate the qualitative data from the students' logs. We then triangulated these two sets of data. To this end, the participants' views on formative assessment reflected on the questionnaire as well as their insights and reflection throughout the learning process exhibited in students' logs were triangulated.

Concerning RQ2, we first transcribed SGP1 and SGP2 to analyse participants' uses of suprasegmentals. Analysis of the transcripts was carried out using a paired *t*-test.

Table 1. Levels of Comprehensibility

Comprehensibility level		Description
5. Fully comprehensible	FC+	**Speech is effortless to understand**
	FC−	Pitch variation may make the speech sound lively or engaging
4. Largely comprehensible	LC+	**Speech requires little effort to understand**
	LC−	Speech may be characterised by too many or too few variations in pitch, sounding disjointed or monotonous
3. Reasonably comprehensible	RC+	**Speech requires some effort to understand**
	RC−	Errors somewhat interfere with the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence)
2. Basically incomprehensible	BI+	**Speech is effortful to understand**
	BI−	Errors are detrimental to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence)
1. Largely incomprehensible	LI+	**Speech is painstakingly effortful to understand, or indecipherable**
	LI−	Errors are debilitating to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence)
0. Incomprehensible	I	**Speech cannot be rated**
		Speech is not assessable or is unresponsive to the task

Note. Adapted from "Developing a user-oriented second language comprehensibility scale for English medium universities" by T. Isaacs, P. Trofimovich, J. A. Foote, 2017, *Language Testing, 35*(2) (https://doi.org/10.1177/0265532217703433). The original rating scale is publicly available on the IRIS Digital Repository (https://www.iris-database.org/iris/app/home/detail?id=york:932362). CC BY-NC-SA 3.0.

This involved the generation of a matrix in which we counted instances in which the participants actually used suprasegmental as was evident in both the SGP1 and SGP2. Additionally, a Pearson correlation analysis (SAS 9.4) was performed on the three variables (thought groups, sentence stress, and intonation) in both podcasts. An analysis of variance (ANOVA) stratified by students was used to measure the effects of suprasegmental use on the participants' levels of comprehensibility. The evaluators' measurements of comprehensibility in SGP1 were contrasted with those in SGP2. Participants' data and time were used as a control factor for stratification. Therefore, the four evaluators evaluated each student's oral performance to determine any differences in the levels of comprehensibility within SGP1 and SGP2.

Results

Does the Use of Formative Assessment Practices Help Students Improve Their Comprehensibility When Recording a Podcast?

Analysis of the participants' views regarding their participation in formative assessment practices, as expressed through the student logs and the questionnaire, indicated they believed their oral production was improved by exercising self-regulated behaviours represented in metacognitive and behavioural aspects and the fostering of motivational beliefs. The student logs provided the examples quoted in the present study and were reproduced as written by the participants.

Exhibition of Self-Regulated Behaviours

Our analysis of the student logs and the questionnaire revealed metacognitive strategies and motivational beliefs that supported self-regulated behaviours (see Table 2).

Table 2. Participants' Self-Regulated Behaviours

	Aspects	Traits
Exhibition of self-regulated behaviours	Metacognitive	• monitoring learning • implementing speaking strategies • using assessment tools
	Behavioural	Changing attitudes towards formative assessment practices
	Fostering motivational beliefs	• sense of achievement • self-confidence

Metacognitive Aspects. One characteristic of self-regulated learners is their management of metacognitive strategies (Zimmerman, 1990). Metacognition is represented in what a student knows about themselves as a learner and has been conceptualised as a blend of knowledge and regulation (Flavell, 1979). Students' knowledge about learning strategies and the ways they monitor, plan, and evaluate their learning (Trujillo-Becerra et al., 2015) are also aspects of metacognition. The coding procedures revealed that participants managed their learning through three metacognitive strategies: monitoring learning, implementing speaking strategies, and using assessment methods.

Monitoring Learning. Analysis of data collected through student logs showed that participants recognized gradual benefits from practising self-assessment. In the following excerpt, the participant contrasts their difficulties when recording SGP1 and SGP2.

> In the first diagnostic I was terrible, I had a bad pronunciation, I didn't put the stress of the words and I was very nervous. Now, I am still nervous, but when I listen my new recording, I got happy. I really improved my work, that's wonderful. [*sic*] (Log, Participant 7)

Implementation of Speaking Strategies. As the study progressed, participants showed initiative in using strategies that helped them produce more intelligible speech in the target language. Participant 8 referred to how organizing their speech whilst reinforcing it at a segmental level (sounds) and at a suprasegmental level (prose division) led to oral improvement: "With the time I learnt to create stories with the correct organization, pronunciation, grouping and the final recordings were better than the first" [*sic*].

Using Assessment Methods. On the questionnaire, all the participants rated the rubric used to self- and peer-assess their podcasts as useful (*n* = 5 *extremely useful*, *n* = 10 *useful*, *n* = 3 *somewhat useful*). In considering how the rubric helped determine the degree of task completion and the learning derived from self- and peer-evaluation, Participant 17 noted:

> Now I can decide when a work is well-done or need to be improved. It is thanks to the rubric. I also learn of my classmates as I learn of my own work, so I really found benefits with this experience. [*sic*] (Log)

However, using the rubric challenged some participants. For example, one claimed to have gained only partial understanding of the rubric criteria even by the end of the pedagogical intervention: "I think that the rubrics are still being difficult. But I interpret them better" [*sic*] (Log, Participant 18).

Behavioural Aspects. Zimmerman (1990) identifies the systematic use of behavioural strategies represented in "responsiveness to feedback regarding the effectiveness of their learning; and by their self-perceptions of academic accomplishment" (p. 14) as another trait common to self-regulated learners. Similarly, analysis of

the participants' perceptions about their participation in formative assessment practices revealed attitudes that supported more active involvement in the pursuit of higher achievement.

Changing Attitudes Towards Formative Assessment Practices. In this study, formative assessment consisted of self- and peer-assessment of SGP1 and SGP2. Concerning self-assessment, participants' abilities to self-assess evolved from limited or restricted participation to active involvement. At the beginning of the course, some participants were apprehensive about self- and peer-assessment and mainly awaited teacher assessment: "I guess that the person who have judgement to tell us our mistakes are the teachers" [*sic*] (Log, Participant 20).

By the end of the course, however, participants' perceptions of their role changed significantly:

> While I had been evaluating my work, I could see a lot of things that I was making in a wrong way and I correct them. Now, my evaluation is more accurate because I know what things I have to take into count when I do it. [*sic*] (Log, Participant 18)

Likewise, participants showed no inhibition about self-assessment, acknowledging it as a previously unexplored learning strategy: "I don't have any problem to reflect upon my podcast, finally we have to learn as best we can and also this is a good strategy" (Log, Participant 3).

In terms of peer-assessment, at the beginning of the course, participants were reluctant to face their classmates' judgement: "I do not like other people listen my audio because I consider that my pronunciation in different aspects is not appropriate" [*sic*] (Log, Participant 6).

As the course progressed, however, participants' views on peer assessment became more positive: "Evaluate my partners is good because I find the correct way of pronunciation some words" [*sic*] (Log, Participant 15).

Students also seemed to notice differences in their peers' production: "Before I couldn't understand some word that my partners said" [*sic*] (Log, Participant 4).

Such perceptions are paralleled by participants' questionnaire responses concerning the usefulness of peer-feedback versus teacher feedback. We found that most participants ($n = 16$) felt peer-feedback was *as useful* as teacher feedback, while a few ($n = 2$) agreed that it was at least *somehow useful*. Nevertheless, when asked about the usefulness of peer-feedback, only 20% believed it was *very useful*, while 50% thought that peer-feedback was *somewhat useful*, and the remaining 30% found it *just useful*.

Fostering Motivational Beliefs. Participants mostly regarded their own active participation in formative assessment practices as a positive experience, since it helped them increase their sense of achievement and build their self-confidence.

Sense of Achievement. When participants rated their *comprehensibility* at the end of the course, 90% ($n = 16$) considered themselves *reasonably comprehensible* while 10% ($n = 2$) perceived themselves as *largely comprehensible*. These perceptions are further paralleled in data from the student logs on how they perceived their progress: "The final version is too different to the original; it is more worked and absolutely better" (Log, Participant 10).

This excerpt exemplifies how participants recognised progress between the first and final versions of their podcasts, perhaps motivated by their own involvement whilst refining the product.

Self-Confidence. Data from the student logs also suggest the participants' enhanced self-confidence derived from a sense of achievement in a variety of domains. For example, one participant attributed their improved performance to lowered inhibitions and greater confidence as the course progressed: "But in the last day, my performance was better, I was more relax and for that reason I could express better my ideas" (Log, Participant 2).

Another participant who had exhibited considerable frustration at the beginning of the course reported enjoyment upon completing their round of podcasts. Their self-confidence increased as they undertook formative assessment throughout the course:

At the beginning of the course I really hated to listen my voice in English, I used to think that it sound terrible, but when the pass of the days I really started to enjoy listen my voice in English and know it feels so good. [*sic*] (Log, Participant 18)

What Is the Effect of Training Students on the Use of Suprasegmental Features Seeking Comprehensibility?

RQ2 was addressed by analysing transcripts of the SGP1 and SGP2, the subsequent matrix of participants' attempts to use suprasegmental features, and the levels of comprehensibility reported by evaluators.

Student-Generated Podcasts 1 (SGP1)

For SGP1 and SGP2, we analysed the transcripts to observe participants' uses of suprasegmentals before and after training. The mark-up used to identify the use of each suprasegmental is as follows: thought groups = /, sentence stress = <u>underline</u>, rising intonation = ↑ and falling intonation = ↓. Inaccurate pronunciation was signalled in *italics*. The following sample illustrates the basic use of suprasegmentals attempted by one of the participants in SGP1.

(1) One day there was a group of friends/their name [*sic*] were Mateo Luis Camila

(2) and Oscar/*dead were* so happy and <u>*united*</u>/so they *decide* to travel around the

(3) <u>*mountain*</u> to find a new kind of mind/by connecting with the earth/but/they didn't

(4) have in mind which kind of problem they would have/so/they start to travel one

(5) Saturday at the morning/there was a sunny day/ but when they moved/they saw

(6) differents scennaries [*sic*] with snow/and specially↑ a <u>*mountain*</u> of color red/one

(7) day↓ when were when they were at this snow site↑/ one of the friends↑ Mateo/get

(8) sick/so/ they *start* to/search about the <u>*mountain*</u> a different kind of

(9) plants/to/to/have or get better to Mateo/so in <u>one</u> part of this place/they *<u>found</u>* a

(10) mystery plant . . .
(SGP1, Participant 12)

In this sample, the participant made early attempts to pause and form semantically and grammatically coherent segments of discourse or thought groups (Lines 1–4). However, some other thought groups were wrongly assembled (Lines 5–8). With regard to sentence stress, the discourse does not generally contain prominent words except for the case of the word "one" in Line 9. The researchers agreed that this was the participant's attempt to convey nuanced meaning by drawing the listeners' attention to the very first part of the statement. The use of rising and falling intonation in SGP1 is restricted to parenthetical expressions (Celce-Murcia et al., 2010), as observed in Lines 6 and 7, where the participant used adverbials (i.e., "specially," "one day") to utter non-finite asides, varying the intonation contour and, thereby, generating expectation of a main idea.

Student-Generated Podcasts 2 (SGP2)

For SGP2, we used the same mark-up as for SGP1; however, the analysis of SGP2 sought to identify changes in participant use of suprasegmentals. Accordingly, the following sample exemplifies a more sophisticated use of suprasegmentals by the same participant above.

(1) <u>Once</u> upon a time↓/a group of students from the Xxxxxx university/ they lived in

(2) Colombia↓/ <u>each</u> one with the respective family↓/ they *decided* to take a trip to the

(3) world's high <u>*mountains*</u>↓/ they began their *<u>journey</u>* through the *mountains*↓/ and

(4) every step that they walked/ they knew <u>more</u> and <u>more</u> <u>new</u> things↓/they had ever

(5) seen in their life↓/ they saw new animals↑/ Abandonate houses↑/ and

(6) <u>occasionally</u>↓/people <u>walking</u> by the same way↓/ after walking↓/ the *scouts*

(7) *decided* to take a <u>shortcut</u>↓/ to get <u>faster</u>/ where they began their *journey* through

(8) the *mountains*/ but↓/ between <u>many</u> <u>trees</u>↑/ they began to walk/ and they saw/that

(9) the way <u>didn't</u> have output↓/ so they *decided* to return/and always came to the

(10) same point↓/ <u>all</u> people felt concerning /because they <u>didn't</u> find a solution to the

(11) problem↓/ and the sky was <u>already</u> getting dark↓/ there was a time that one of the

(12) people ↑/ who had found↑/wanted to help↓/ so he *decided* to call the <u>police</u>/

(13) Finally↓/ they . . .

(sGP2, Participant 12)

In contrast with Participant 12's performance on their sGP1, their sGP2 sample demonstrates greater ability to pause logically during the stream of speech, as shown in Lines 2–8, where speech pauses marked with a slash (/) helped frame more coherent utterances. The participant's use of sentence stress improved also, giving more prominence to certain content words (e.g., "<u>more</u> and <u>more</u> <u>new</u> things," "<u>shortcut</u>," "<u>many</u> <u>trees</u>") and additional shades of meaning to their discourse (e.g., "the sky was <u>already</u> getting dark," "they were <u>very</u> happy"). The sample also shows that, in contrast with sGP1, the participant used intonation for different purposes; for example, the phrases "once upon a time"

(Line 1), "occasionally" and "after walking" (Line 6), and "finally" (Line 13) were uttered with an initial high pitch followed by a terminal fall. In Lines 8–9, there was fluctuation in intonation, changing from low, to intermediate, and then to high pitch before reaching the terminal fall. Through the use of such pitch variations, the participant was able to generate a sense of surprise; the researchers were easily able to infer the message *there was no way out* (Line 9), despite the participant's wrong choice of the word "output." In Line 10, the participant's stress on the words "all" and "didn't" helped the researchers clarify quantity and react to an unexpected result; these words were also uttered with a higher intonation contour, probably with the intention of drawing the listener's attention.

Matrix Analysis

The analysis of suprasegmentals in sGP1 and sGP2 using a paired *t*-test revealed a significant trend toward increased accuracy in the use of thought groups, sentence stress, and intonation (Table 3). The pair *t*-test applied to sGPs showed correlation within suprasegmentals before and after the sample population ($n = 18$) underwent training on the variables thought groups, sentence stress, and intonation to seek comprehensibility. These variables indicate correlation as the *p*-value in each of them is lower than 0.05 or 5%; thought groups (0.0004), sentence stress (0.0006), and intonation (0.0001). Lastly, intonation revealed the most significant correlation of all, followed by sentence stress and thought groups.

**Table 3.** Differences Between Scores Before and After the Pedagogical Intervention for Each Variable per Student

Variables	*N*	Mean	*SD*	Std. error mean	*T*	*Df*	*p*-value
Thought groups	18	0.2091	0.1993	0.0470	4.45	17	0.0004*
Sentence stress	18	0.3161	0.3205	0.0756	4.18	17	0.0006*
Intonation	18	0.2121	01539	0.0363	5.85	17	0.0001*

*$p < 0.05$ ns: Non-significant

Levels of Comprehensibility

In answering RQ2, the medians of grades reported by each of the evaluators (Figure 2) were also considered.

**Figure 2.** Evaluators' Scores for Comprehensibility Levels in SGP1 and SGP2
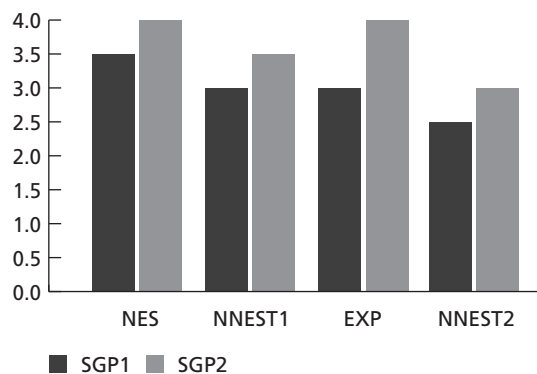
Note. SGP = Student-generated podcast, NES = Native English speaker, NNEST = Non-native English speaking teacher, EXP = Expert

The evaluation of SGP1 and SPG2 revealed differences in the levels of comprehensibility (Isaacs et al., 2017) of the average across the 18 participants in the study, as reported by each of the four evaluators. NES, NEST1, and EXP rated SGP1 as *reasonable comprehensible* while NNEST2 rated it as *basically incomprehensible*. NES rated SGP1 as *reasonably comprehensible* (RC+) with the highest score (3.5) in contrast with NNEST 2 who rated it as *basically incomprehensible* (BI+) with the lowest score (2.5). NNEST1 and EXP coincided in rating SGP1 as *reasonably comprehensible* (RC–). In SGP2, NES, NNEST1, and NNEST2 reported a slight gain in comprehensibility (0.5) when contrasting scores with SGP1 while EXP reported a more significant gain (1.0). NES and EXP coincided in rating SGP1 as *largely comprehensible* (LC–). NNEST2 reported that SGP2 barely achieved a reasonably comprehensible (RC–) level. None of the evaluators scored SGP2 as *fully comprehensible* (FC– or FC+). NES rated both podcasts with some level of comprehensibility, SGP1 as *reasonably comprehensible* (RC+) and SGP2 as *largely comprehensible* (LC–); he perceived a slight gain in comprehensibility between SGP1 and SGP2 (0.5). Evaluations of SGPs (Figure

3), contrasting the means of evaluators' grades in SGP1 ($M = 2.83$) and SGP2 ($M = 3.54$), reveal an overall gain in comprehensibility of 0.71 units. Thus, SGP1s were rated as *basically comprehensible*, while SGP2s were evaluated as *reasonably comprehensible*.

Figure 3. Comprehensibility Gains
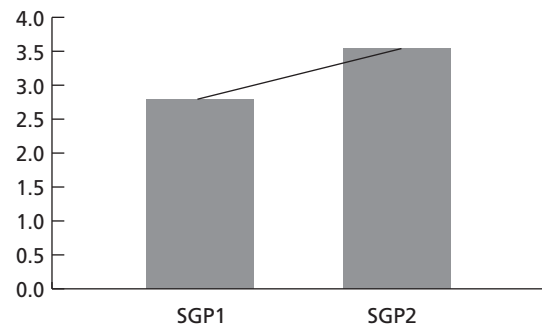
Discussion

This study examined how the use of formative assessment practices with student-generated podcasts influenced participants' spoken comprehensibility. For RQ1, the results show that participants' involvement in formative assessment practices, supported by self-regulated behaviours, contributed to their production of more comprehensible podcasts. We suspect a number of factors contributed to this result. First, participants had the goal of producing more comprehensible speech to meet their academic requirements despite their speaking flaws. Thus, producing a series of podcasts gave them opportunities to monitor their performances and use their speaking strategies while simultaneously benefiting from giving and receiving feedback (Chong, 2016). This aligns with Black and Wiliam's (1998) findings that students tend to react positively to frequent formative assessment. In their view, the use of formative assessment helps students reflect honestly on what they have learnt and also on what they still need to learn to achieve a particular learning objective. Second, engaging in self- and peer-assessment challenged students with an assessment task that, in their context, was more likely to be performed by teachers (Yorke, 2003). Although before

the intervention, students seemed reluctant about self- and peer-assessment and in favour of instructors being the only credible assessment entity, they clearly did not report any major differences in value between feedback received from peers and from the teacher. Third, the internal and external reference points (peer-assessment and rubric) and the students' central role in feedback processes (Nicol & Macfarlane-Dick, 2006) helped them observe their own progress with a sense of achievement and self-confidence. These findings align with Nicol and Macfarlane-Dick's (2006) views on the influence of self-generated feedback on cognitive, motivational, and behavioural levels. In their view, comparison of the progress achieved (in this case, the two versions of the podcasts) against the goals desired (comprehensible speech) can encourage students to reinterpret tasks and adopt new tactics and strategies, a characteristic of higher-performing students who tend to use more language-learning strategies (Griffiths, 2008). Such changed learner awareness could, in turn, have positive effects on the cognitive and motivational domains, as well as self-regulation. Similarly, Kawai (2018) found that participants who undertake rehearsal, planning, and monitoring experienced reduced fear of failure and, thereby, more successful performances.

For RQ2, we observed improved use of suprasegmental features in SGP2, with intonation as the feature showing the greatest improvement, followed by thought groups and sentence stress. This result was not unexpected because, as participants attended training sessions on the use of suprasegmentals and rehearsed while recording a second version of their podcasts, it became easier for them to group their ideas and decide how pitch should fluctuate (Celce-Murcia et al., 2010) to achieve more clarity in their discourse. In this regard, the greater levels of confidence and audience awareness achieved during the training process and formative assessment experience may have encouraged the participants to make new attempts to pitch changes in parenthetical expressions. As a result, these provided evaluators with better

hints as to the message being conveyed and enhanced participants' overall comprehensibility in SGP2. This suggests that explicit instruction on suprasegmental features in EFL could help learners achieve higher levels of comprehensibility (Gordon & Darcy, 2016). In contrast to the study of Bøhn and Hansen (2017), in which a group of teachers agreed on the importance of comprehensibility over native-like pronunciation but were unsure or not clear on the role of intonation when assessing pronunciation, the present study concludes that, after training on suprasegmentals, improvements on intonation were indeed the most significant aspects to help participants enhance their comprehensibility.

The present study also reports on the participants' changing levels of comprehensibility as observed by various qualified evaluators from different backgrounds. The results show that the evaluators perceived similar gains in comprehensibility, equal to 0.71 units with respect to the two evaluations performed (SGP1 and SGP2). Unlike the discourse observed in SGP1, in which only occasional words or phrases could be easily comprehended by listeners, the discourse exhibited in SGP2 contained longer stretches of language in which no significant effort from the listener was required for comprehension (Isaacs et al., 2017). Accordingly, we conclude that the level of comprehensibility in this group of learners generally evolved from *basically comprehensible* ($M = 2.83$) to *reasonably comprehensible* ($M = 3.54$).

Nevertheless, although evaluators reported similar gains in comprehensibility, it is interesting that NES' evaluation was more forgiving of grammar and word-level irregularities than the one provided by the other evaluators, who were more attentive to accuracy. This difference may have been influenced by their educational backgrounds: NEST1, NEST2, and EXP possess university degrees in English language teaching, whereas the NES has a university degree in a different area and was in the process of learning Spanish. These conditions probably led NES to focus more on communication, accepting the accuracy and pronunciation flaws often presented by

users of English as an international language (Jenkins, 2015; R. Walker, 2010). In contrast, NEST1 and NEST2 tended to be more concerned about accuracy, perhaps obeying rooted older paradigms of English language teaching despite the trends of globalisation on contemporary English usage (Crystal, 2003; Tahereen, 2015).

To conclude, because the participants in this study were experiencing difficulties when attempting a clear discourse in English, a pedagogical approach using explicit instruction, enhanced by student-driven assessment on the use of suprasegmentals—thought groups, sentence stress, and intonation—was incorporated into their regular EFL course. A promising finding was that participants progressed from producing *basically comprehensible* to *reasonably comprehensible* discourse for an audience of both native and non-native speakers. Moreover, the present study adds to the body of empirical evidence concerning student-content generation mediated by podcasts, demonstrating that students can go far beyond content generation when student-generated podcasts are intertwined within a self- and peer-formative assessment framework that allows them to rehearse, practice, and revise gaps in oral production. Such practices can help learners not only achieve higher levels of comprehensibility but, more importantly, exert control over their own learning as they experience senses of achievement and self-confidence while both giving and receiving feedback. Overall, self- and peer-formative assessment practices create a potential for change in classroom practices by getting students involved in their learning achievement and, consequently, empowering them to develop self-regulated behaviours that, in turn, contribute to more autonomous and effective lifelong learning for a globalised world.

Limitations and Further Research

There were a number of limitations in the present study that should be acknowledged. Due to time and availability constraints, some students decided not to take part in the study; having a larger group of participants could have illuminated more issues. Also, an additional round of self- and peer-assessment of podcasts would have probably helped students to reflect more confidently on what they had already learnt, what they were in the process of learning, and what they had not yet learnt. Finally, the short duration of the present study represents another limitation; a lengthier study could provide further opportunities to strengthen the training on suprasegmental features. Further research can target the role of teachers in formative assessment practices, more specifically on the provision of models and demonstrations of self- and peer-assessment procedures that help students to enact these practices confidently. Research on the role of external evaluators in determining the degree of comprehensibility of the students' oral discourse can help instructors better understand how to tailor their lessons to help their students communicate effectively with a broader range of English speakers.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32(2), 159–181. https://doi.org/10.1080/02602930600801928

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

Bøhn, H., & Hansen, T. (2017). Assessing pronunciation in an EFL context: Teachers' orientations towards nativeness and intelligibility. *Language Assessment Quarterly*, 14(1), 54–68. https://doi.org/10.1080/15434303.2016.1256407

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Campbell, G. (2005). There's something in the air: Podcasting in education. *Educause Review*, *40*(6), 32–47.

Celce-Murcia, M., Brinton, D. M., Goodwin, J., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Chong, I. (2016). How students' ability levels influence the relevance and accuracy of their feedback to peers: A case study. *Assessing Writing*, *31*, 13–23. https://doi.org/10.1016/j.asw.2016.07.002

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge. https://doi.org/10.4324/9780203029053

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Sage Publications.

Council of Europe. (2011). *Common European framework of reference for languages: Learning, teaching, and assessment*. Cambridge University Press.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.

Derwing, T., & Munro, M. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.42

Farangi, M. R., Nejadghanbar, H., Askary, F., & Ghorbani, A. (2015). The effects of podcasting on EFL upper-intermediate learners' speaking skills. *CALL-EJ Online*, *16*(2), 1–18. Retrieved from http://callej.org/journal/16-2/Faranji_Nejadghanbar_Askary_Ghorbani2015.pdf

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56–92. https://doi.org/10.1075/jslp.2.1.03gor

Green, M. F., & Shoenberg, R. (2006). *Where faculty live: Internationalizing the disciplines*. American Council on Education.

Griffiths, C. (2008). Strategies and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 83–98). Cambridge University Press. https://doi.org/10.1017/CBO9780511497667.009

Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education*, *40*(1), 151–164. https://doi.org/10.1080/02602938.2014.893418

Harlen, W. (2007). *Assessment of learning* (1st ed.). Sage Publications.

Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing*, *21*(3), 390–408. https://doi.org/10.1191/0265532204lt2890a

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Isaacs, T., Trofimovich, P., & Foote, J. (2017). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice*, *2*(3), 49–85. https://doi.org/10.1515/eip-2015-0003

Jenkins, J., Baker, W., & Dewey, M. (Eds.). (2018). *The Routledge handbook of English as a lingua franca*. Routledge. https://doi.org/10.4324/9781315717173

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the World: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.

Kawai, Y. (2018). Speaking and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 218–230). Cambridge University Press. https://doi.org/10.1017/CBO9780511497667.020

Knight, J. (2003). Updated definition of internationalization. *International Higher Education*, *33*, 2–3. https://doi.org/10.6017/ihe.2003.33.7391

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Macías, D. (2010). Considering new perspectives in ELT in Colombia: From EFL to ELF. *How*, *17*, 181–194.

Ng'ambi, D., & Lombe, A. (2012). Using podcasting to facilitate student learning: A constructivist perspective. *Educational Technology and Society*, *15*(4), 181–192.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research and Evaluation*, *10*, 1–8. https://doi.org/10.7275/a166-vm41

Popova, A., & Edirisingha, P. (2010). How can podcasts support engaging students in learning activities? *Procedia: Social and Behavioral Sciences*, *2*(2), 5034–5038. https://doi.org/10.1016/j.sbspro.2010.03.816

Rosell-Aguilar, F. (2013). Podcasting for language learning through iTunes U: The learner's view. *Language Learning & Technology*, *17*(3), 74–93.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation*, *11*. https://doi.org/10.7275/9wph-vv65

Sargeant, J. (2008). Toward a common understanding of self-assessment. *Journal of Continuing Education in The Health Professions*, *28*(1), 1–4. https://doi.org/10.1002/chp.148

Tahereen, T. (2015). Challenges in teaching pronunciation at tertiary level in Bangladesh. *International Journal of English Language & Translation Studies*, *3*(1), 9–20.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.2307/1170598

Trujillo-Becerra, C. L., Álvarez-Ayure, C. P., Zamudio-Ordoñez, M. N., & Morales-Bohórquez, G. (2015). Facilitating vocabulary learning through metacognitive strategy training and learning journals. *Colombian Applied Linguistics Journal*, *17*(2), 246–259. https://doi.org/10.14483/udistrital.jour.calj.2015.2.a05

van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, *1*(2), 133–147. https://doi.org/10.1016/j.edurev.2006.05.001

Walker, A. (2001). British psychology students' perceptions of group-work and peer assessment. *Psychology Learning & Teaching*, *1*(1), 28–36. https://doi.org/10.2304/plat.2001.1.1.28

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford University Press.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477–501. https://doi.org/10.1023/A:1023967026413

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

About the Authors

Albedro Cadena-Aguilar is a lecturer and researcher in the Languages Department at the Universidad de La Sabana. He is also a coordinator and a teacher in the English proficiency programme for the upper-intermediate level. His areas of research lie within self-directed learning, assessment, and teacher training.

Claudia Patricia Álvarez-Ayure works as a lecturer in the on-campus and online master's programmes for ELT in the Faculty of Education at the Universidad de La Sabana. Her research ranges over online learning environments, formative assessment, and teacher education.

Appendix A: Rubric for Self-and Peer-Assessment

Category	Meets expectations	Meets most expectations	Shows consistent progress	Needs improvement	Shows minimal progress
Understanding the task and organization	• Narrates a personal experience mishap using a logical and coherent sequence: setting, beginning, climax, and finalizes with an amusing conclusion. • Shows high knowledge and command of the subject/topic. • Makes excellent use of connective devices as well as discourse and sequence markers. • Manages time adequately (2 minutes).	Some traits of 5 and some traits of 3.	• The mishap structure does not include two or more of the elements required. • Shows some knowledge and fair command of the subject/topic. • Connective devices and or discourse markers are missing and or not always used correctly. • Time limit is slightly exceeded or is barely reached.	Some traits of 3 and some traits of 1.	• The mishap is minimally developed. • Shows limited or no knowledge of the subject/topic. • Connective devices and/or discourse markers are not used/are incorrectly used. • Time is managed inappropriately.
Narrating actions in a mishap	• Expresses ideas with ease in well-controlled sentence structure. • Uses appropriately a range of narrative verb tenses (simple, continuous, and perfect past tenses)	Some traits of 5 and some traits of 3.	• Expresses ideas and responses with enough control of grammatical forms and proper sentence structure. • Uses a limited range of narrative verb tenses with a fair degree of accuracy.	Some traits of 3 and some traits of 1.	• Frequently expresses ideas and responses with inconsistent and fragmented sentence structure. • Shows several inconsistencies with verb tense usage.
Using the right words	• Makes consistent and appropriate word choice as regards the topic and context. • Uses correct word inflections to form nouns, adjectives, adverbs.	Some traits of 5 and some traits of 3.	• Makes fair word choice with occasional flaws or unnecessary repetition. • Makes fair attempts to word inflections although some are not appropriate.	Some traits of 3 and some traits of 1.	• Shows limited vocabulary that is little related and often repetitive. • Lacks awareness of word inflection.

Making myself clear	• Delivers speech smoothly and shows little hesitation. • Grouping of words helps to understand the story better. • Uses intonation appropriately most of the time. • Places sentence stress accordingly. • Articulates individual sounds clearly most of the time.	Some traits of 5 and some traits of 3.	• Delivers speech clearly enough despite some hesitation. • Attempts to use grouping, but sometimes inappropriately. • Uses intonation with fair appropriacy. • Attempts to place sentence stress although inappropriately in some occasions. • Mispronounces some individual sounds that do not interfere with comprehension.	Some traits of 3 and some traits of 1.	• Delivers speech with some distortion and frequent hesitation. • Makes wrong grouping and/or distracts the listener. • Lacks awareness of intonation differences / uses wrong intonation. • Uses stress inaccurately. • Mispronunciation of individual sounds often causes misunderstanding.
Comments	Good things				
	Things to improve				

____ I certify I have made a thorough and objective analysis of the podcasts assigned to me.

Appendix B: Students' Assessment Questionnaire

Dear students,

The responses to this survey will assist us in determining your perception of self- and peer-assessment practices seeking comprehensibility.

Select the option that best describes you.

Feedback instruments

	Very useful	Useful	Somewhat useful	Not useful	Not very useful
1. The implementation of rubrics as an instrument to evaluate my peers' work is...					
2. The guideline questions in the student logs used to reflect on self and peer assessment were…					

Level of success

	Strongly agree	Agree	Somewhat agree	Disagree	Strongly disagree
3. I believe my own feedback on podcasts helped me exceed/achieve my speaking goals for this course.					
4. I believe peer recording and listening to podcasts was a valuable strategy to become a more fluent speaker.					

Personal views on peer feedback

5. I see peer feedback as an opportunity to reflect and learn from others.					
6. To me, peer feedback is as useful as teacher's feedback					

Usefulness of feedback

	Very useful	Useful	Somewhat useful	Not useful	Not very useful
7. In my opinion, the feedback I gave my peers on the stories told for the purposes of this course was...					
8. The feedback my peers gave on the stories I told for the purposes of this course was...					

Future benefits of feedback

	Strongly agree	Agree	Somewhat agree	Disagree	Strongly disagree
9. In my view, students who take this course and record several versions of podcasts incorporating feedback from peers, would improve comprehensibility of storytelling podcasts.					

	Very likely	Likely	Somewhat likely	Unlikely	Very unlikely
10. How likely would you be to recommend peer feedback on podcasts as a strategy to improve comprehensibility in a second language?					