DOI: https://doi.org//10.21747/16466195/ling17a3

Rhythm instruction to improve ESP students' fluency: A study of pausing

Leticia Quesada Vázquez
lquesada@nebrija.es
Universidad Nebrija

ABSTRACT

The present work is aimed at examining how efficient it is to teach language rhythm to enhance English for Specific Purposes (ESP) learners' fluency. 298 Spanish/Catalan speakers participated in the experiment, among which 42 students fulfilled treatment and were examined as subjects of the study. For ten weeks, participants took pronunciation instruction within their regular lessons. The pronunciation sessions lasted thirty minutes and followed a communicative framework. Students were divided into an experimental group, with explicit rhythm training, and a control group, without it. The total number of pauses and unfilled pauses were counted and compared to assess their frequency performing ANOVAs and t-tests. Descriptive analysis shows that the experimental group tends to pause less and make fewer unfilled pauses after training. However, statistical significance is only reached for the independent-samples t-test of the effect sizes of the number of unfilled pauses. Therefore, these findings suggest a greater command over pausing when rhythm is explicitly taught, and a consequent improvement of fluency in the second language, despite statistical results not always being significant.

KEYWORDS

Rhythm instruction; pronunciation teaching; pausing; fluency; pronunciation assessment.

1. Introduction

English for Specific Purposes (ESP) students' proficiency level when speaking varies considerably, and some of them experience rejections when trying to enroll at masters' programs or applying for international companies due to their low communicative competence in the target language (Räisänen & Fortanet-Gómez 2008); despite having the knowledge to address a given situation, these students do not manage to communicate effectively (Douglas 2000). Although ESP courses may contain much listening and speaking practice often implemented as group work and task-based activities, it does not mean that spoken interaction is taught effectively, as aspects such as

input and feedback do not tend to be closely examined (Dudley-Evans & St John 1998). For instance, many times ESP teachers do not know enough about the scientific content of a task so, on the one hand, sometimes the material brought to class is not updated and, on the other hand, the feedback provided might not respond to the final communicative purpose of the task, but rather to the grammatical correctness of the students' interactions.

Pronunciation does not tend to be explicitly taught in ESP context either, although it is essential for successful oral communication to happen. In fact, second language pronunciation teaching remains a hot topic nowadays. After long periods of neglect, a shift in focus towards communication and intelligibility has fostered new research approaches to pronunciation instruction, but these proposals do not usually filter down to the second language classroom (see Derwing & Munro 2015 for a review). Therefore, more class-based studies are needed in order to examine the applicability of the research advances on pronunciation teaching and guide practitioners on how to deal with pronunciation effectively in their classes to enhance students' oral competence.

1.1. Teaching rhythm to improve fluency

Introducing rhythm within the L2 classroom has been revealed to be beneficial to enhance the students' communicative skills (Anderson-Hsieh 1990; Chela-Flores 1997, 2003; Derwing & Munro 2015; Frost & Picavet 2014; Gilbert 2008; Goodwin 2013; Ordin & Polyanskaya 2015; Quené & Van Delf, 2010; Tajima *et al*. 1997; Tuan & An 2010; Wong 1987). However, whether rhythm should be at the forefront of the features to teach has been questioned. Jenkins (2000) did not encourage teaching how to produce an English-like rhythm for intelligibility purposes if English is working as a lingua franca, as nonnative speakers might adopt a different rhythm without jeopardizing communication (as cited in Levis 2018: 143). Levis (2018) argued, however, that in this globalized world L2 speakers are in contact with native Englishes, so it is important that they are aware of how native speakers produce rhythm. He also maintained that even when rhythm does not interfere with the correct interpretation of the message, it is

likely to ease its understanding.

Rhythm anticipates and delimitates crucial information for the successful comprehension of a message. Indeed, a misuse of the right rhythm in the target language can lead to serious communication issues that impede mutual understanding (Gilbert 2008). This is particularly true when the L1 and the L2 clearly differ in rhythmic patterns, such as the case of Spanish speakers learning English. Stress in English is characterized by considerably increasing the length, loudness and pitch of a syllable within words, or words within longer utterances. Knowing that is essential to understand and speak English because not only does it clarify lexical and grammatical misunderstandings, but it also points at the meaningful information of speech. By contrast, in spite of the fact that there are some well-defined rules on how to use syllable stress in Spanish, stressed syllables are not as long, loud and high in pitch as the English ones (Dauer 1987). Besides, sentence stress is not used to highlight important content; in Spanish a more flexible syntax allows pushing this information towards the end of a sentence so as to make it stand out (Chela de Rodríguez 1976).

These rhythmic cues inevitably affect timing: while the English rhythm is based on similar time intervals between stresses, Spanish rhythm is based on similar time intervals between syllables. Consequently, when Spanish speakers use English applying the Spanish rhythm, their speech sounds disjointed. Patel (2007) goes one step beyond and cognitively relates the organization of music with the organization of the discourse. As a piece of music can be arranged in beats, note values and bars, the speech is divided into different elements which are assembled according to perceptual boundaries that go from simple (i.e., the syllable) to more complex (i.e., the utterance) connections. In both cases, the boundaries are mainly established by the use of pausing. The hierarchical structure of the discourse helps us to understand the links among syllables within words, phrases, and full sentences, which constitute the complete meaning of speech. Although all these connections play a role, those that create thought groups (also known as intonational units) are considered key to build the rhythm of a language. Thought groups are semantically and syntactically coherent and meaningful segments into which an utterance is divided. Even though there are no strict rules on how to segment the discourse into intonational units

and the division can vary depending on the speed, people have an intuitive notion of when it is right to pause (Celce-Murcia *et al*. 1996). For example, if in a sentence like "the teacher told me to close the door", the speaker pauses like "the teacher told/ me to/close the/door", the sense, intention and meaning of the sentence are lost.

Although some researchers have studied the teachability of rhythm reporting encouraging results in terms of perception, error rate or intelligibility (Chela-Flores 1997; Couper 2006; Hahn 2004; Tsiartioni 2011), no such research has been conducted on how rhythm instruction might affect fluency. Pausing, phrasing or speed of delivery are features that alter the rhythm of the speech and play an important role in how smooth-spoken a speech is. Dudley-Evans and St John (1998) highlighted the importance of silence: controlling when and how to pause can make an impact on the audience, who uses that silent time to process the information given. If the pauses are either too long, wrongly placed, or made too often in a short period of time, speakers sound choppy and ineloquent. Cauldwell (2008) also supported the importance of pausing in his study on air traffic control, where he showed how long pauses can determine the level of fluency of the members of the International Civil Aviation Organization (ICAO). Hence, the transfer of the mother tongue's rhythm when speaking in a second language affects the frequency, location and general use of pauses, which may contribute to creating dysfluent speech. Practicing how and when to pause to adopt the proper language rhythm can enhance students' fluency in the target language.

1.2. Pausing: an indicator of fluency

In Wood's words, "the most complex and one of the most informative elements of fluency studied so far in empirical research involves pause phenomena" (2010: 23). After examining different research studies exploring second language speech, he resolves that, together with location, pausing frequency has been revealed as a crucial indicator of fluency. Lennon (1990) investigated the pause time of four German students doing an international stay in Reading (England). He concluded that the total unfilled pause time

decreased in 25% for three of the four participants while the total filled pause time remained the same. Riggenbach (1991) conducted a study with six Chinese speakers of English who had previously been classified as either fluent or dysfluent to see which linguistic features were fluency indicators. Among the features studied, the pausing analysis revealed that the number of unfilled pauses was a determining indicator of fluency. Besides, the difference in frequency of the total number of pauses between fluent and dysfluent speakers was significant. Derwing *et al*. (2004)'s research on the influence of different perceptual tasks on fluency ratings revealed that pausing stood for more than 60% of the variance in the mean fluency ratings for both the monologue and the picture description tasks. Therefore, analyzing the pauses second language speakers made shed light on how fluent they were in the target language.

The current study belongs to a larger project which investigates the effectiveness of explicit rhythm training to improve engineering undergraduates' comprehensibility and fluency in English. For this paper, only fluency will be examined. To this end, the following hypotheses are formulated:

H1. Students will make fewer pauses when language rhythm is explicitly taught and trained.

H2. Students will make fewer unfilled pauses when language rhythm is explicitly taught and trained.

2. Method

A classroom-based longitudinal study was conducted according to Derwing and Munro's (2015) framework. As explained in Quesada Vázquez & Romero (2020), a ten-week pronunciation module based on the instruction of rhythmic features was designed as part of a six-month technical English course taken at Rovira i Virgili University in 2017. This was a compulsory course for first-year undergraduates of all the engineering disciplines, so initially 298 students participated in the experiment, distributed into six class groups. These class groups were divided into two major treatment groups: rhythm was explicitly taught in three of them (experimental group), while

the other three were trained with conventional pronunciation explanations (control group). Students were recorded before and after instruction to assess their progress in comprehensibility and fluency depending on the instruction received.

2.1. Participants

Students were asked to fill in a consent form to participate in the experiment on the first day of class. Forty-two students (twenty-one experimental; twenty-one control) completed treatment and, hence, were eligible as final participants of the study out of the 298 students enrolled. As shown in Table 1, they were mainly 18-year-old Spanish/Catalan speakers with different initial levels of English.

TABLE 1 – Final participants' profiles (adapted from Quesada Vázquez & Romero, 2020)

	Age		First language			English level		
Group	18	18+	Sp/Cat bilinguals	Sp	Other	A2-B1	B1-B2	B2-C1
Experimental	21	0	20	1	0	10	6	5
Control	16	5	16	1	4	4	11	6

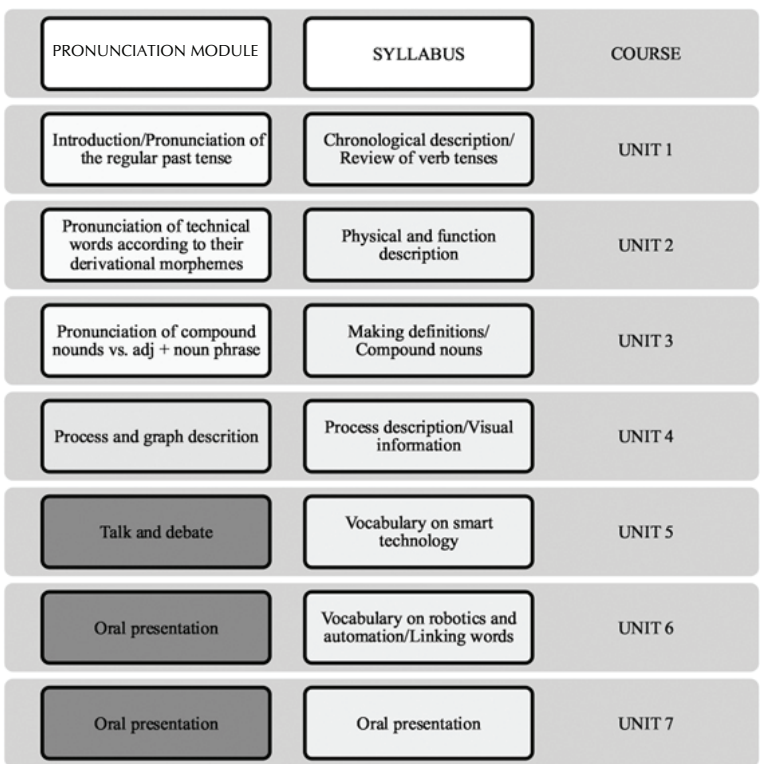
Note. Sp = Spanish. Cat = Catalan. English level is classified according to the Common European Framework of Reference for Languages (CEFR): A2-B1 = basic user. B1-B2 = independent user (intermediate). B2-C1 = independent/proficient user (upper-intermediate/advanced).

Students took a placement test as their first task submission to determine their language level, based on listening and grammar activities. The experimental group was made up of ten beginners, six intermediates, and five upper-intermediate/advanced students, while the control group consisted of four beginners, eleven intermediates, and six upper-intermediate/advanced. Regarding their mother tongue, students who were not Spanish/Catalan bilinguals reported a high command of both Spanish and Catalan, as all of them pursued primary and secondary education in Catalonia. None of them reported hearing problems.

2.2. The module and instruction

The pronunciation module was designed according to the course syllabus, using vocabulary and grammatical items previously seen and practiced in class, so as to guarantee students' motivation (Anderson-Hsieh, 1990) and ensure the applicability of the knowledge acquired to the students' direct reality, meeting the students' needs (Derwing & Munro 2015; Dudley-Evans & St John 1998). The course content was divided into seven units plus the pronunciation module. As displayed in Figure 1, each of the pronunciation sessions was related to one of the units:

FIGURE 1 – Distribution of the pronunciation module according to the course syllabus

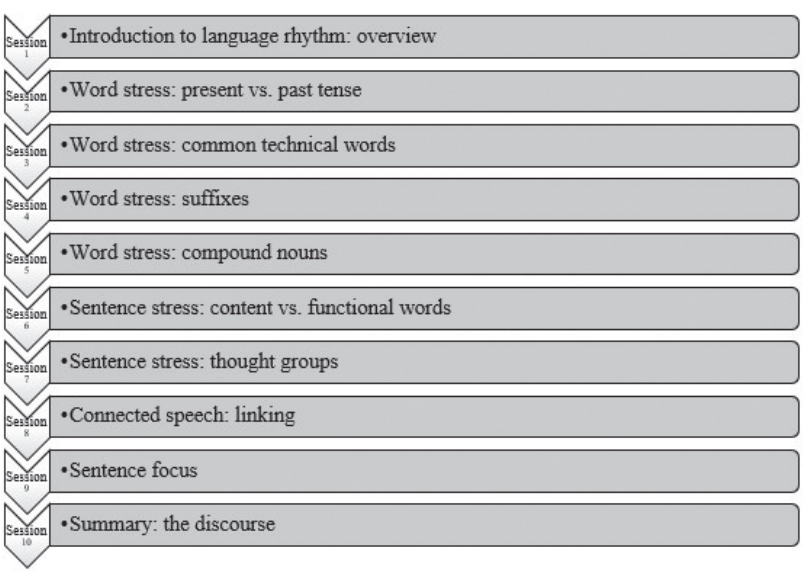


Colors represent the units of speech that were dealt with when working on each pronunciation item (i.e., white represents the word; light gray, the sentence; and dark gray, the discourse). Therefore, instruction was scaffolded.

The main points described at the pronunciation program are related to one or more of the course units. However, the module was divided into ten sessions, so some of the units, which tended to be the ones that took longer to teach, occupied more than one pronunciation session. For example, although sentence instruction was based just on Unit 4, there were two separate pronunciation sessions related to that unit: one worked on process description and the other one on graph description, concepts included within the same regular unit.

Each individual session followed Celce-Murcia *et al*. (1996) communicative framework, starting with an explanation, continuing with listening discrimination, and finishing with oral practice, from controlled to communicative exercises. The only difference between groups was rhythm instruction. Experimental training consisted of practicing word stress, sentence stress, linking and sentence focus.

FIGURE 2 – Distribution of rhythm instruction per session



Rhythm instruction was mainly based on vowel lengthening: as not language-oriented learners, it was considered that length will be a simple concept to assimilate. Besides, vowel lengthening makes a notable difference between English and Spanish rhythms. The introduction of the rhythmic items in class was done progressively, from the simplest to the most complex ones. Its distribution within the sessions is summarized in Figure 2.

Explicit rhythm activities were mainly adapted from textbooks and research papers to suit the technical background of the course. An example of an activity that implied rhythm training is the following: When teaching the pronunciation of common technical words in session 4, the syllables of the word were replaced by *LA* when they carried the main stress and *la* for the rest of them (e.g. *electrician* became *la la LA la*, and *chemical* became *LA la la*), so students got familiarized and practiced the pronunciation of the words through their rhythmic patterns. Secondary stress was excluded in order to focus on the acquisition and practice of primary stress. This technique of using rhythmic patterns to explain word stress, adapted from Chela-Flores (1997: 127), was further used in other sessions to explain other items, such as the difference in pronunciation between compound nouns and adjective + noun combinations in session 5. Those other activities not focusing on rhythm explicitly were created from online resources, such as videos, website texts and images, or adapted from activities for academic purposes and pronunciation exercises.

Control groups did not take rhythm training, so pronunciation was explained and practiced with traditional methods, such as the pronunciation of technical words according to the suffix they carry (i.e., words ending in *–ian* stress the second from last syllable). For a full example of how sessions worked and differed depending on treatment, you can see Figure 2 in Quesada Vázquez & Romero (2020).

2.3. Recordings

Both sessions (before and after training) took place in the library of the School of Engineering. Three semi-sound-proof isolated rooms were booked to carry the recordings out with Two Sony PCM-M1O and a Zoom H4nsp

recorders. Students were recorded individually. The pretest and the posttest were the same in order to guarantee comparability. They consisted of four exercises, including both controlled and extemporaneous activities. The test started with a reading aloud exercise inspired on Couper's (2006) speaking test (p. 61). It was made up of ten sentences, both simple and complex and of an uncontrolled nature (rhythmic patterns were not considered for its usage) that contained technical vocabulary. The second exercise was a diagnostic passage proposed by Celce-Murcia *et al.*'s (1996: 398) to check learners' pronunciation, which students also had to read aloud. The third exercise consisted in introducing themselves, also inspired by Couper's (2006) test. Finally, the fourth exercise was about giving their opinions on social media. The present study only focuses on the assessment of the sentences (i.e., the first exercise).

2.4. Data analysis

The rhythm of the sentences was acoustically measured before and after the instruction. Two aspects were taken into account: VarcoV values (White & Mattys 2007, 2007b), based on vowel duration variability, and pauses. A total of 840 sentences were segmented and acoustically analyzed using PRAAT. An individual sound file and its corresponding text grid were created for each of the sentences. The segmentation was carried out manually following Ordin and Polyanskaya's criteria (2014: 5).

The text grid contained four different tiers: From top to bottom, the first tier showed the transcription of the sounds pronounced. Sound identification followed Olive *et al.* (1993) guidelines; the second tier displayed the syllable structure of each word; the third tier presented the different vocalic and consonant clusters; finally, the fourth tier showed the sentence to pronounce. Blanks represented both filled, i.e., "voiced *fillers* which do not normally contribute to additional lexical information" (Riggenbach, 1991:426), such as hesitations and repetitions, and unfilled pauses. The latter were mainly silent micropauses of less than .2 seconds as the study was carried out with short sentences, not full speeches (For an example of the text grid used, see Figure 3 in Quesada Vázquez & Romero

2020). For this study, the frequency of pauses was analyzed. Pauses were manually counted, and two analyses were performed: one took into account only the unfilled pauses while the other took into consideration the total number of pauses (both filled and unfilled).

2.5. Statistical analysis

Two different mixed repeated-measures ANOVAs were performed. First, the impact on the overall groups' performances depending on the instruction received was investigated by means of a mixed repeated-measures ANOVA with time and sentence as within-subjects factors, and group as a between-subjects factor. Then, in order to study the relevance of the difference in production before and after training, another mixed repeated-measures ANOVA was run with the difference between pretest-posttest pauses (either all of them or just unfilled ones) as the dependent variable, sentence as the within-subjects factor and group as the between-subjects factor. To further analyze the relevance of this difference and reinforce the ANOVAs' results, t-tests were also performed: two paired-samples t-tests comparing groups' learning progress were completed for each sentence. Finally, one more independent-samples t-test with the two groups' sentence effect sizes was also performed.

3. Results

As shown in Table 2, the means and standard deviations of the pauses for each sentence were estimated and distributed according to time (T1 = pretest; T2 = posttest) and group (experimental; control). The interest fell in examining whether students paused less after treatment. When that happened, values were highlighted in light gray. On average, both groups made approximately the same number of pauses before the instruction. However, despite the fact that both groups made fewer pauses after training, the experimental group's total mean decreased more than the one of the control group. In addition, when examining the sentence means, it was

observed that the experimental group decreased its means after instruction in 9 out of 10 sentences, whereas the control group's means fell on 7 out of 10. Therefore, a preliminary descriptive analysis shows that the experimental group reduced more the total amount of pauses after instruction than the control group.

TABLE 2 – Total number of pauses means (*M*) and standard deviations (*SD*) per sentence

	Control group				Experimental group			
	Time 1		Time 2		Time 1		Time 2	
Sentence	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.57	0.51	0.62	0.74	0.52	0.6	0.33	0.48
2	1.86	1.39	1.86	1.15	1.86	1.49	1.52	1.36
3	1.76	1.48	1.14	1.01	0.9	0.77	1.1	1.09
4	3.29	1.38	2.62	1.69	3	1.7	2.48	1.57
5	3.76	2.33	3	1.51	4.12	2.66	3.86	1.55
6	3.26	1.38	2.9	1.3	3.1	1.87	2.71	1.38
7	2.95	1.69	2.52	1.78	2.81	1.6	2.29	1.45
8	4.33	2.73	4.24	2.14	3.86	2.08	3.05	2.22
9	2.43	1.4	2.26	1.18	3.21	1.33	2.56	1.07
10	2.53	0.8	2.9	1.33	2.48	1.21	2.14	1.11
Total	**2.7**		**2.5**		**2.6**		**2.2**	

Appendix A contains the number of pauses made by each of the participants while reading aloud each of the sentences. Generally speaking, intermediate students from the control group and high-intermediate/ advanced students from the experimental group were the ones to show more improvement. Besides, although no clear pattern could be established because individual results varied considerably depending on the student and the sentence performed, participants' means showed that 16 out of 21 students from the experimental group paused less after instruction whereas in the control group they were 12 out of 21. Therefore, individual performances support a greater improvement for the experimental group.

In order to investigate the effect of instruction on group performance,

a mixed repeated-measures ANOVA was run. Time and sentence were the within-subjects factors and group was the between-subjects factor. The total number of pauses was the dependent variable. Results did not show significance for group, $F(1,40) = .427$, $p = .517$, but they did for time, $F(1,40) = 8.995$, $p < .005$, and sentence $F(9,32) = 46.743$, $p < .01$. However, no significance was found for the time*sentence interaction $F(9,32) = .578$, $p = .739$ (see Table 3).

TABLE 3 – Mixed repeated-measures ANOVA with the total number of pauses as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	8.995	.005**
	sentence	9	32	46.743	.01**
Interaction	time*sentence	9	32	.578	.739
Between-subjects	group	1	40	.427	.517

Note. ** $p < .01$.

Nevertheless, Figure 3 shows that the number of pauses made dropped more after treatment for the experimental group, agreeing with the tendency observed when comparing group means.

Another ANOVA was run so as to examine the magnitude of the difference between pretest and posttest performances. The dependent variable was the difference in the number of pauses for each sentence before and after treatment, sentence was the within-subjects factor and group was the between-subjects factor. Neither sentence, $F(9,32) = .827$, $p = .546$, nor group, $F(1,40) = 1.567$, $p = .218$, showed significance (see Table 4).

FIGURE 3 – Group progress on the total number of pauses according to treatment

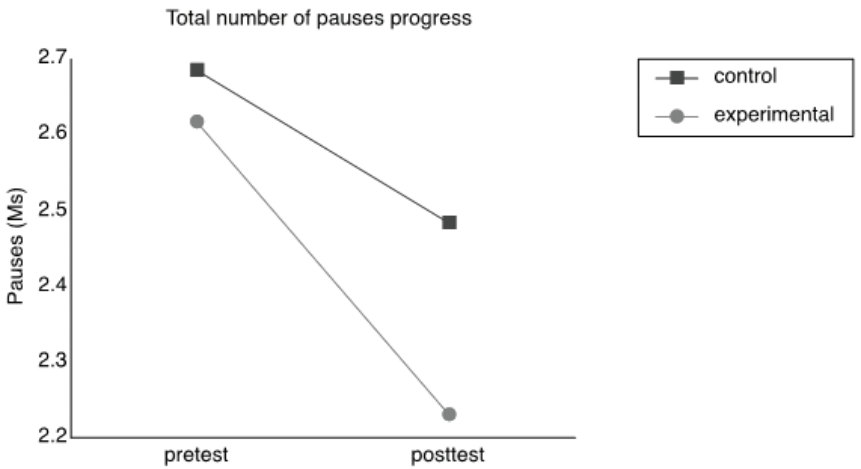


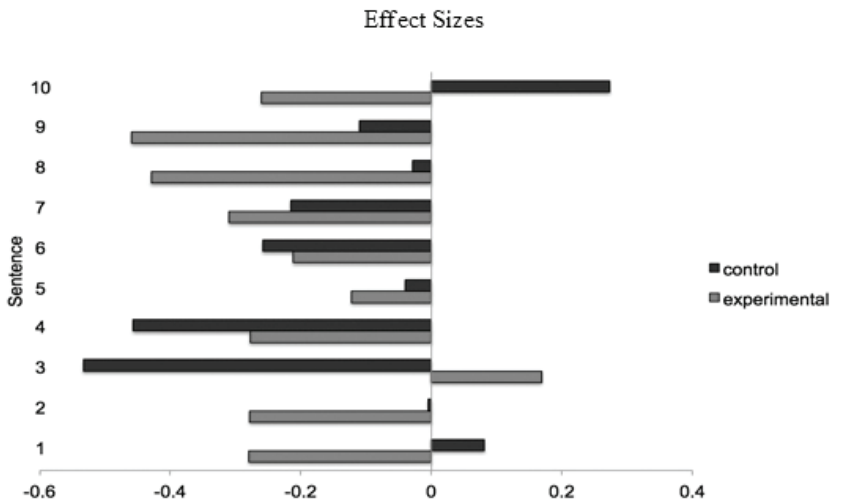
TABLE 4 – Mixed repeated-measures ANOVA with the total number of pauses difference over time as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	32	.827	.546
Between-subjects	group	1	40	1.567	.218

To further examine the relevance of the difference, two paired-samples (control T2 vs. control T1; experimental T2 vs. experimental T1) and two independent-samples t-tests (control T1 vs. experimental group T1; control T2 vs. experimental group T2) per sentence were calculated, which reached significance only in a few specific cases (see Appendix B). Independent-samples t-tests were significant when analyzing both groups before treatment in sentence 3, $T(40) = 2.355$, $p < .023$, and after treatment in sentence 10, $T(40) = 2.044$, $p < .048$. As for paired-samples t-tests, they showed significance for the control group performance in sentences 3, $T(20) = -2.444$, $p < .024$, and 4, $T(20) = -2.092$, $p < .049$, and for the experimental group performance in sentence 9, $T(20) = -2.103$, $p < .048$. A

final independent t-test comparing the effect sizes of all the paired-samples t-tests was also carried out to examine the relevance of the improvement. A difference between the two treatment groups' progress was not obvious, since both showed mostly negative effect sizes, which corresponded to a reduction of the number of pauses made (see Figure 4). Besides, results did not reveal significance $T(18) = 1.229$, $p = .235$.

FIGURE 4 – Effect sizes for the total number of pauses paired-sample t-tests



The same analytical process was followed to examine unfilled pauses. As shown in Table 5, the experimental group reduced the number of unfilled pauses in eight out of ten sentences, made the same amount in sentence 3, and made more in sentence 5. The control group, on the other hand, made fewer pauses in six out of ten sentences, the same amount in sentence 8, and more in sentences 1, 9 and 10. In addition, the total mean of pauses decreased more for the experimental group. Hence, the experimental group seemed to perform better after instruction.

TABLE 5 – Unfilled pauses means (*M*) and standard deviations (*SD*) per sentence

	Control group				Experimental group			
	Time 1		Time 2		Time 1		Time 2	
Sentence	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	0.52	0.51	0.57	0.75	0.48	0.51	0.33	0.48
2	1.95	1.47	1.81	1.03	1.86	1.49	1.52	1.36
3	1.43	1.03	1.1	0.89	0.9	0.77	0.9	0.83
4	2.76	1.04	2.52	1.63	2.81	1.66	2.29	1.62
5	3.43	1.8	3.33	1.35	3.52	1.97	3.79	0.89
6	2.72	0.79	2.71	1.1	2.81	1.54	2.43	1.16
7	2.57	1.36	2.19	1.5	2.29	1.19	1.86	1.06
8	4	2.41	4	2.05	3.71	2.03	2.86	1.93
9	2.19	1.08	2.24	1.37	3.05	1.16	2.33	1.02
10	2.33	1.06	2.71	1.1	2.38	1.07	2.1	1.04
Total	**2.4**		**2.3**		**2.4**		**2**	

Again, individual performance differs depending on the sentence examined (see Appendix C). However, none of the control beginners improved after instruction this time. As for the participants' individual means, 15 experimental and 10 control students decreased their number of unfilled pauses after training, which reinforces a better performance of the experimental group.

The results of first mixed repeated-measures ANOVA that was run (time and sentence = within-subjects factors; group = between-subjects factor; number of unfilled pauses = dependent variable) are displayed in Table 6. Results were not significant for group, $F(1,40) = .349$, $p = .558$, although they were for sentence, $F(9,32) = 56.725$, $p < .01$, and for time, $F(1,40) = 5.478$, $p < .024$. Nevertheless, the time*sentence interaction was, again, not significant, $F(9,32) = .773$, $p = .589$. Therefore, results seemed to correlate with the ones obtained when analyzing the total number of pauses.

TABLE 6 – Mixed repeated-measures ANOVA with unfilled pauses as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	time	1	40	5.478	.024*
	sentence	9	32	56.725	.01**
Interaction	time*sentence	9	32	.773	.589
Between-subjects	group	1	40	.349	.558

Note. * p = .05. ** *p* = .01.

However, Figure 5, which displays the progress of the different groups in regard of the number of unfilled pauses, shows a slightly different scenario from the one presented regarding the total number of pauses in Figure 4. Firstly, while the control group made more pauses than the experimental group on average at the beginning of the course, the number of unfilled pauses was almost the same for both groups. Secondly, when analyzing all the pauses, the progress of both groups was similar, since the control group made 0.2 pauses and the experimental group 0.4 pauses less than the first time they took the test. Nonetheless, when taking into account just the unfilled pauses, the impact of progress changed. The control group remained almost the same at 2.4 pauses, while the experimental group made 0.4 pauses less after instruction. Thus, the improvement of the control group seemed to result from making fewer hesitations and repetitions rather than silent pauses.

The second ANOVA was carried out with the difference in times as the dependent variable, sentence as the within-subjects factor and group as the between-subjects factor. As seen when analyzing the total number of pauses, results did not reach significance for any of the factors under study: sentence, $F(9,32) = .708$, $p = .643$; group, $F(1,40) = 2.470$, $p = .124$.

FIGURE 5 – Group progress on the number of unfilled pauses according to treatment

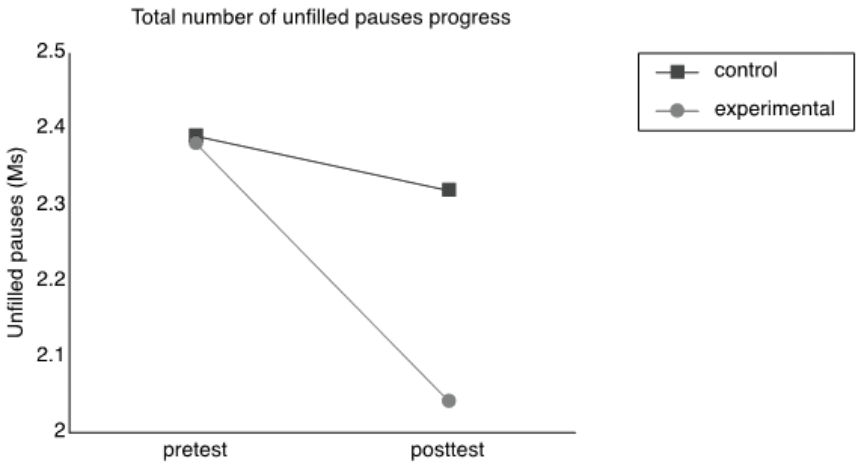
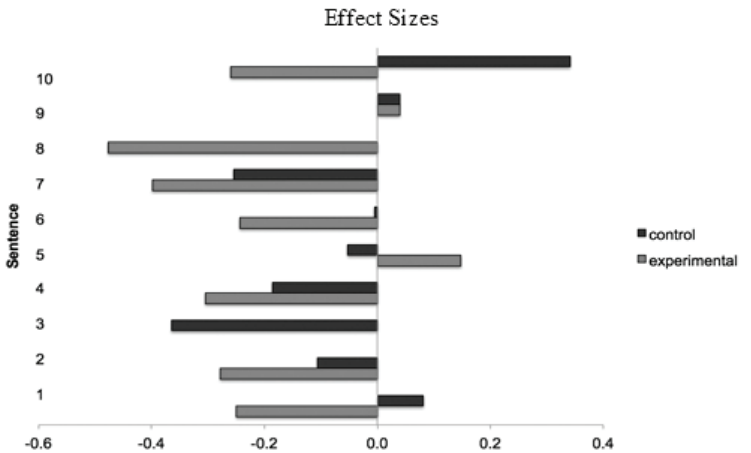


TABLE 7 – Mixed repeated-measures ANOVA with the unfilled pauses difference over time as the dependent variable

	Effect	df	Error df	*F*	*p*-value
Within-subjects	sentence	9	32	.708	.643
Between-subjects	group	1	40	2.470	.124

Regarding t-test analyses, these, again, reached significance in only some of the cases (see Appendix D). This time, only one independent-samples t-test showed significance: groups before treatment in sentence 9, $T(40) = -2.480$, $p < .017$. As for the paired-samples t-tests, results for the experimental group were found significant in sentences 8, $T(20) = -2.186$, $p < .041$, and 9, $T(20) = -2.855$, $p < .010$. However, this time the effect sizes of all the paired-samples t-tests showed a greater improvement for the experimental group performances in most of the cases (see Figure 6). What is more, the independent-samples t-test with these effect sizes did show significance $T(18) = 2.343$, $p < .031$.

FIGURE 6 – Effect sizes for the total number of unfilled pauses paired-sample t-tests



To sum up, the different analyses carried out for the two items under analysis (i.e., total number of pauses and unfilled pauses) suggest favorable effects on students' prosody when pronunciation was introduced in class, especially when rhythm was explicitly taught, even though results were not always statistically significant.

4. Discussion

For both unfilled and the total number of pauses, the descriptive analysis did not reveal a clear difference in the treatment received although the experimental group showed a slightly greater improvement in both cases (see Tables 2 and 5). On the one hand, both the control and the experimental group reduced their sentence means after training and, on the other hand, the overall mean per group was a bit lower for the experimental group after instruction. As for individual values, there were more students from the experimental group whose means decreased after training (see Appendixes A and C). However, no students improved in all the sentences and there is no sentence in which all of them improved. In fact, the sentence standard deviations oscillated between one and two in most of the cases, so a wide

variability among speakers seems to be confirmed.

As for the ANOVAs carried out with the total number of pauses and unfilled pauses as the dependent variable respectively, both the sentence and the time variables were found significant, but the time*sentence interaction was not, so no statistical connection could be established between the two variables. The statistical results are a bit more encouraging when examining the data for just unfilled pauses, though. As suggested in the previous section, the control group progress seems to be thanks to the fact that they made fewer hesitations and repetitions after instruction, which could be the consequence of the in-class practice of the pronunciation of certain words as part of the vocabulary of both the pronunciation module and the regular lectures. However, they made approximately the same number of silent pauses after instruction, which suggests that they kept pausing within thought groups. On the other hand, the experimental group reduced the number of both filled and unfilled pauses to the same extent. Besides, the analysis of the effect sizes of the paired-samples t-tests conducted for the unfilled pauses was significant ($p < .031$), while the one for the total number of pauses was not. These results suggest, hence, that the experimental group recognized thought groups better and avoided pausing in between them more than the control group, producing a more English-like rhythm.

Overall, the statistical analysis of the total number of pauses extracted from the sentences does not confirm hypothesis 1 while the analysis of the number of unfilled pauses provides more evidence of an improvement when rhythm instruction takes place, supporting hypothesis 2. Although the descriptive analysis of the total number of pauses shows that both groups improve after instruction to a certain extent and that the experimental group performed better after treatment, the statistical analyses conducted do not fully support these assumptions. As previously mentioned, individual variability could partially explain the results obtained, but there are other factors that could have also played a role, such as the limited number of sessions, or the dramatical decrease of students who fulfilled treatment and, hence, were taken as the final participants of the study. Nonetheless, the analysis of unfilled pauses seems to provide clearer results on fluency improvement despite a lack of robust statistical significance. The results of this study go along the lines of previous research such as Lennon (1990)

and Riggenbach (1991), who found unfilled pauses to be a good indicator of fluency. It is also interesting to point out that Wood claimed in his compilation of research studies of fluency that "the importance of pause times and frequencies tells us a great deal about speech fluency, particularly that related to the value of unfilled pauses. It appears that analyzing filled pauses yields mixed and inconclusive results (2010: 24)".

5. Conclusion

Although further research should be conducted to obtain more conclusive results, the findings of this study are enlightening. On the one hand, pronunciation instruction, particularly explicit rhythm training, has been revealed to be helpful for ESP students to manage pausing when speaking in the target language and, thus, enhance their fluency in controlled oral production, such as reading sentences aloud. It is true, though, that this study only focuses on sentences that are read aloud and more research should be done on students' performances in the text reading and opinion tasks to examine longer speeches. In addition, it would be interesting to examine the exact location of the pauses made within the sentences to support the assumption that these were not made within thought groups. On the other hand, this study can become a model for those researchers and practitioners who are looking for approaches to introduce pronunciation effectively within ESP courses. By means of this study, it has been revealed that pronunciation instruction is possible within the ESP field adapting the lessons to the students' professional needs and, besides, can bring benefits to the students' communicative skills. However, more classroom-based studies are still necessary to fully support these hypotheses and help to bridge the gap between practice and research.

REFERENCES

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes*, *9*(3), 195-214. https://doi.org/10.1016/0889-4906(90)90013-3

Caudwell, R. (2008). Defining fluency for air traffic control. *Speech in Action*, *3*, 1-5. https://www.speechinaction.org/wp-content/uploads/2012/03/Defining-Fluency.pdf

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambrige University Press.

Chela-Flores, B. (1997). Rhythmic patterns as basic units in pronunciation teaching. *Onomázein*, (2), 111-134.

Chela-Flores, B. (2003). Optimizing the teaching of English suprasegmentals to Spanish speakers. *Lenguas Modernas*, (28–29), 255-273.

Chela de Rodríguez, B. (1976). Teaching English suprasegmentals to Spanish speakers. *Kansas Working Papers in Linguistics*, *1*, 63-82.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*(1), 46-66.

Dauer, R. M. (1987). *Phonetic and phonological components of language rhythm*. In T. Gamkrelidze (Ed.), Paper presented at the 11th International Congress of Phonetic Sciences (ICPhS) (pp. 447-450). Academy of Sciences of the Estonian S.S.R.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655-679.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge University Press.

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes. A multi-disciplinary approach*. Cambridge University Press.

Frost, D., & Picavet, F. (2014). Putting prosody first – Some practical solutions to a perennial problem: The Innovalangues project. *Research in Language*, *12*(3), 1-11. https://doi.org/10.2478/rela-2014-0002

Gilbert, J. B. (2008). *Teaching pronunciation: using the prosody pyramid*. Cambridge University Press.

Goodwin, J. (2013). Pronunciation teaching methods and techniques. In C. A. Chapelle

(Ed.), *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0970

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201-223.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387-417.

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambrige University Press.

Olive, J. P., Greenwod, A., & Coleman, J. (1993). *Acoustics of American English speech: A dynamic approach*. Springer-Verlag.

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*, 244-257. https://doi.org/10.1016/j.system.2013.12.004

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533-544. https://doi.org/10.1121/1.4923359

Patel, A. D. (2007). *Music, language, and the brain*. Oxford University Press USA.

Quené, H., & Van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication*, *52* (11-12), 911-918. https://doi.org/10.1016/j.specom.2010.03.005

Quesada Vázquez, L., & Romero, J. (2020). ESP within ESP: the design and implementation of a pronunciation module in a technical English course. *Onomázein*, (NE VI), 209–228. https://doi.org/10.7764/onomazein.ne6.11

Räisäenen, C., & Fortanet-Gómez, I. (2008). The state of ESP teaching and learning in Western European higher education after Bologna. In I. Fortanet-Gómez, & C. Räisänen (Eds.), *ESP in European higher education. Integrating language and content* (pp. 11-51). John Benjamins. https://doi.org/10.1075/aals.4.03rai

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*(4), 423-441. https://doi.org/10.1080/01638539109544795

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*(1), 1-24. https://doi.org/10.1006/jpho.1996.0031

Tsiartsioni, E. (2011). *Can pronunciation be taught? Teaching English speech rhythm to Greek students*. In E. Kitis, N. Lavidas, N. Tpointzi, & T. Tsangalidis (Eds.), *Selected Papers from the 19th International Symposium on Theoretical and Applied Linguistics,*

Thessaloniki 3-5 April 2009 (pp. 447-458). Aristotle University of Thessaloniki. https://doi.org/10.26262/istal.v19i0.5522

Tuan, L. T., & An, V. T. P. (2010). Teaching English rhythm by using songs. *Studies in Literature and Language*, *1*(2), 13-29.

White, L., & Mattys, S. L. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501-522. https://doi.org/10.1016/j.wocn.2007.02.003

White, L., & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaro, & M.-J. Sole (Eds.), *Segmental and prosodic issues in romance phonology (Current issues in linguistic theory series)* (pp. 237-257). John Benjamins.

Wong, R. (1987). *Teaching pronunciation: Focus on English rhythm and intonation*. Regents Prentice Hall.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum.

Appendix A. Total number of pauses per student

Total number of pauses – individual values and means for the control group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	1	2	2	5	5	4	4	5	1	4	**3.3**	1	4	2	5	6	4	5	8	4	4	**4.3**
2	1	1	2	5	4	3	3.52	3	3	3	3	**3.1**	1	3	4	1	3	5	1	5	4	2	**2.9**
3	1	1	1	2	4	1	5	3	1	4	3	**2.5**	2	2	2	4	4	6	3	5	3	3	**3.4**
4	1	1	1	3	3	5	4	4	6	2	2	**3.1**	0	3	1	2	4	3	4	5	2	5	**2.9**
5	2	0	3	0	3	2	3	2	3	5	3	**2.4**	0	3	0	4	6	4	6	3	2	3	**3.1**
6	2	1	1	2	3	3	3	4	2	2	3	**2.4**	1	1	2	2	4	3	1	4	1	4	**2.3**
7	2	1	2.14	3	2	9	2	3	6	2	4	**3.4**	1	2	1	1	4	3	1	5	2	3	**2.3**
8	2	1	2	1	4	4	5	4	3	2	3	**2.9**	0	0	1	2	4	4	1	2	1	4	**1.9**
9	2	0	4	5	5	5	2	5	4	4	1	**3.5**	0	1	1	2	6	2	5	4	3	3	**2.7**
10	2	1	5	2	4	5	7	6	4	6	2.57	**4.3**	2	3	2	6	6	3	5	8	2.52	4	**4.2**
11	2	0	0	1	2	2	2	0	2	1	1	**1.1**	0	0	0	2	1	1	0	2	0	3	**0.9**
12	2	1	2	2	4	4	2	4	12	3	3	**3.7**	1	2	2	3	5	3	1	5	3	1	**2.6**
13	2	1	2	3	4	10	3	3	6	2	2	**3.6**	1	2	1	4	4	3	3	3	2	3	**2.6**
14	2	0	3	2	7	5	4	3	5	3	3	**3.5**	0	3	2	6	3	3	2	8	2	6	**3.5**
15	2	1	1	1	2	5	2	2	2	1	3	**2.0**	0	1	1	2	3	2	5	6	2	3	**2.5**
16	3	0	2	0	2	2	2	2	0	2	2	**1.4**	1	0	1	1	3	2	1	2	2	2	**1.5**
17	3	0	0	2	3	2	2	5	7	0	2.57	**2.4**	0	1	0	1	1	0	2	0	0	1	**0.6**
18	3	0	0	0	2	3	2	0	8	2	2	**1.9**	0	2	0	1	4	3	1	5	2	2	**2.0**
19	3	0	0	0	3	1	3	4	6	2	2	**2.1**	0	1	0	0	4	2	2	2	4	2	**1.7**
20	3	1	4	1	2	3	5	1	2	2	2	**2.3**	2	3	1	3	2	2	2	3	4	2	**2.4**
21	3	0	2	0	1	2	3	0	4	2	2	**1.6**	0	2	0	3	2	3	2	4	2	1	**1.9**
Total												**2.7**											**2.5**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Total number of pauses – individual values and means for the experimental group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	0	0	0	3	4	4	2	2	2	0	**2.1**	0	0	4	1	2	3	2	0	2	2	**1.6**
2	1	1	1	1	4	10	2	6	4	1	1	**3.3**	1	1	2	2	4.19	3	6	2	4	3	**2.8**
3	1	0	0	0	1	0	5	0	1	2	0	**0.9**	0	0	0	1	4.19	2	2	1	1	0	**1.1**
4	1	1	2	2	3	4	4	1	5	3	1	**2.7**	0	3	1	0	2	4	2	2	4	3	**2.1**
5	1	0	0	0	3	0	2	2	2	1	0	**1.2**	0	0	1	3	3	4	4	1	1	1	**1.8**
6	1	0	3	0	1	8	3	4	6	3	0	**3.1**	1	4	0	3	5	3	2	3	3	3	**2.7**
7	1	0	3	2	3	4	5	2	4	3	0	**2.9**	0	2	3	3	3	2	2	5	3	2	**2.5**
8	1	1	2	2	1	4	8	2	7	4	1	**3.2**	0	2	3	3	7	4	1	8	4	4	**3.6**
9	1	0	1	0	1	2	4	4	3	3	0	**2.1**	0	1	0	0	4.19	2	2	2	3	1	**1.5**
10	1	0	0	1	3	4	3	4	4	3	0	**2.4**	0	1	1	3	3	3	0	1	1	1	**1.4**
11	2	0	0	1	5	4	1	4	2	6	0	**2.4**	1	2	1	3	4	3	2	2	4	3	**2.5**
12	2	0	1	1	6	5	2	2	3	3	0	**2.5**	0	0	1	2	6	2	2	2	4	2	**2.1**
13	2	1	4	2	3	6	3	3	7	3	1	**3.5**	1	1	1	3	6	4	3	6	2	3	**3.0**
14	2	1	3	1	5	7	3	4	8	3.48	1	**3.8**	0	2	1	5	5	6	3	7	2.76	3	**3.5**
15	2	0	1	0	0	0	1	0	3	2	0	**0.8**	0	0	0	1	2	1	1	5	2	1	**1.3**
16	2	1	4	1	6	4	1	3	4	3	1	**3.2**	1	4	1	4	4	2	3	4	2	4	**2.9**
17	3	1	2	0	3	5	2	6	3	6	1	**3.1**	0	2	1	1	3	1	3	4	3	3	**2.1**
18	3	0	1	1	4	7	0	2	1	4	0	**2.3**	0	0	0	1	5	0	1	1	1	1	**1.0**
19	3	1	4	1	1	3	6	3	7	4	1	**3.4**	1	1	1	4	6	4	5	2	2	2	**2.8**
20	3	1	4	2	4	8	3	2	2	5	1	**3.4**	1	4	1	6	6	3	2	5	3	2	**3.3**
21	3	2	3	1	3	4	3	3	3	3	2	**2.6**	0	2	0	3	2	1	0	1	2	1	**1.2**
Total												**2.6**											**2.2**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Appendix B: Comparison of total number of pauses – t-tests results

Comparison of total number of pauses across Time (T1 vs T2) and Group (Control vs Experimental) – t-tests results

Sentence	T-test		df	*T*	*p*-value	Cohen's *d*
		C T2-T1	20	.370	.715	.0814
	Paired	E T2-T1	20	-1.284	.214	-.279
		C T1-E T1	40	.277	.783	.090
1	Independent	C T2-E T2	40	1.482	.146	.464
		C T2-T1	20	-.022	.983	-.005
	Paired	E T2-T1	20	-1.276	.217	-.278
		C T1-E T1	40	.015	.988	.005
2	Independent	C T2-E T2	40	.855	.398	.269
		C T2-T1	20	-2.444	.024*	-.533
	Paired	E T2-T1	20	.777	.446	.169
		C T1-E T1	40	2.355	.023*	.729
3	Independent	C T2-E T2	40	.146	.884	.038
		C T2-T1	20	-2.092	.049*	-.457
	Paired	E T2-T1	20	-1.272	.218	-.278
		C T1-E T1	40	.597	.554	.187
4	Independent	C T2-E T2	40	.284	.778	.086
		C T2-T1	20	-.183	.857	-.040
	Paired	E T2-T1	20	-.561	.581	-.122
		C T1-E T1	40	-.741	.463	.228
5	Independent	C T2-E T2	40	-.764	.450	.236
		C T2-T1	20	-1.183	.251	-.258
	Paired	E T2-T1	20	-.969	.344	-.211
		C T1-E T1	40	.331	.742	.102
6	Independent	C T2-E T2	40	.460	.648	.141
		C T2-T1	20	-.987	.335	-.216
	Paired	E T2-T1	20	-1.419	.171	-.310
		C T1-E T1	40	.281	.780	.085
7	Independent	C T2-E T2	40	.475	.637	.142
		C T2-T1	20	-.133	.895	-.029
	Paired	E T2-T1	20	-1.966	.063	-.429
		C T1-E T1	40	.636	.528	.194
8	Independent	C T2-E T2	40	1.766	.085	.544
		C T2-T1	20	-.503	.620	-.110
	Paired	E T2-T1	20	-2.103	.048*	-.459
		C T1-E T1	40	-1.864	.070	.575
9	Independent	C T2-E T2	40	-.854	.398	.264
		C T2-T1	20	1.251	.225	.273
	Paired	E T2-T1	20	-1.195	.246	-.261
		C T1-E T1	40	.172	.865	.053
10	Independent	C T2-E T2	40	2.044	.048*	.629

Note. C = control group. E = experimental group. T = time. Paired = paired-samples t-test. Independent = Independent-samples t-test. Cohen's *d* = Effect size.
* $p < .05$.

Appendix C: Unfilled pauses per student

Unfilled pauses – individual values and means for the control group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	1	2	2	5	4	4	4	5	1	4	**3.2**	1	4	2	5	6	4	5	8	4	4	**4.3**
2	1	1	2	2	2	3	3.05	3	3	3	2	**2.4**	1	2	3	1	3	4	1	5	3	2	**2.5**
3	1	1	1	2	4	1	4	3	1	4	2	**2.3**	2	2	2	4	4	5	3	5	3	3	**3.3**
4	1	0	1	2	3	4	3	3	6	2	2	**2.6**	0	2	1	2	4	3	3	5	2	4	**2.6**
5	2	0	3	0	3	2	3	2	3	4	2	**2.2**	0	3	0	4	5	4	5	3	2	3	**2.9**
6	2	1	1	2	3	3	3	3	2	2	3	**2.3**	0	1	2	2	3	3	1	4	1	4	**2.1**
7	2	1	5	2	2	6	2	3	6	2	3	**3.2**	1	2	1	1	3	3	1	5	2	3	**2.2**
8	2	1	2	1	3	3	4	3	2	2	3	**2.4**	0	2	1	2	3	3	1	1	1	4	**1.8**
9	2	0	4	4	3	5	2	4	4	3	1	**3.0**	0	1	1	2	5	2	4	4	3	3	**2.5**
10	2	1	4	2	4	5	3.05	5	4	4	5	**3.7**	2	3	2	5	5	3	5	7	6	3	**4.1**
11	2	0	0	1	1	2	2	0	2	1	1	**1.0**	0	0	0	1	1	1	0	2	0	2	**0.7**
12	2	1	2	2	4	4	2	3	10	3	3	**3.4**	1	2	2	3	5	3	1	5	3	1	**2.6**
13	2	1	2	2	4	8	3	3	6	2	2	**3.3**	1	2	1	4	4	3	2	3	2	3	**2.5**
14	2	0	3	2	2.95	5	3	3	5	3	3	**3.0**	0	3	2	6	3	3	2	7	2	5	**3.3**
15	2	1	1	1	2	5	2	2	2	1	3	**2.0**	0	1	1	2	3	2	3	6	2	3	**2.3**
16	3	0	2	0	2	2	2	2	0	2	2	**1.4**	1	0	1	1	3	2	1	2	2	2	**1.5**
17	3	0	0	2	3	2	2	4	7	0	0	**2.0**	0	1	0	1	1	0	2	0	0	1	**0.6**
18	3	0	0	0	2	3	2	0	6	2	2	**1.7**	0	2	0	1	2	2	1	3	1	2	**1.4**
19	3	0	0	0	2	1	2	3	5	2	2	**1.7**	0	1	0	0	3	2	1	2	2	2	**1.3**
20	3	1	4	1	2	3	4	1	1	2	2	**2.1**	2	3	1	3	2	2	2	3	4	2	**2.4**
21	3	0	2	0	1	1	2	0	4	1	2	**1.3**	0	1	0	3	2	3	2	4	2	1	**1.8**
Total												**2.4**											**2.3**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Unfilled pauses – individual values and means for the experimental group (cleaned data)

		T 1											T 2										
St.	L	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	*M*
1	1	0	0	0	3	3	4	1	2	2	3	**1.8**	0	0	3	1	3.38	2	1	0	2	2	**1.4**
2	1	1	1	1	4	7	2	4	4	1	3	**2.8**	1	1	2	2	3.38	3	4	2	3	3	**2.4**
3	1	0	0	0	1	0	4	0	1	2	0	**0.8**	0	0	0	1	3.38	2	1	1	1	0	**0.9**
4	1	1	2	2	3	4	3	1	5	3	2	**2.6**	0	3	1	0	3.38	3	2	2	3	3	**2.0**
5	1	0	0	0	2	0	1	2	1	1	2	**0.9**	0	0	0	2	3.38	3	2	1	1	1	**1.3**
6	1	0	3	0	1	6	3	3	6	3	3	**2.8**	1	4	0	3	5	2	2	3	3	3	**2.6**
7	1	0	3	2	3	4	5	2	4	3	3	**2.9**	0	2	2	3	3	2	2	5	3	2	**2.4**
8	1	1	2	2	1	3	6	2	5	3	1	**2.6**	0	2	2	2	5	3	1	5	3	3	**2.6**
9	1	0	1	0	1	2	4	3	3	3	3	**2.0**	0	1	0	0	3.38	2	1	2	3	1	**1.3**
10	1	0	0	1	3	3	3	3	4	3	2	**2.2**	0	1	1	3	3	3	0	1	1	1	**1.4**
11	2	0	0	1	4	4	1	2	2	5	1	**2.0**	1	2	1	3	4	3	2	2	3	3	**2.4**
12	2	0	1	1	6	3	2	2	3	3	2	**2.3**	0	0	1	2	3	2	2	2	3	2	**1.7**
13	2	1	4	2	3	6	3	3	7	3	3	**3.5**	1	1	1	3	5	4	3	6	2	3	**2.9**
14	2	1	3	1	5	5	3	3	8	6	3	**3.8**	0	2	1	5	5	5	3	7	5	3	**3.6**
15	2	0	1	0	0	0	1	0	3	2	1	**0.8**	0	0	0	0	3.38	1	1	4	2	1	**1.2**
16	2	1	4	1	6	4	1	3	4	3	4	**3.1**	1	4	1	4	4	2	3	4	2	4	**2.9**
17	3	1	2	0	3	5	2	5	3	4	3	**2.8**	0	2	1	1	3	1	3	4	2	3	**2.0**
18	3	0	1	1	3	3	0	2	1	4	3	**1.8**	0	0	0	1	4	0	1	1	1	1	**0.9**
19	3	1	4	1	1	3	5	3	7	4	4	**3.3**	1	1	1	4	5	4	3	2	2	2	**2.5**
20	3	1	4	2	4	6	3	2	2	3	3	**3.0**	1	4	1	6	5	3	2	5	3	2	**3.2**
21	3	1	3	1	2	3	3	2	3	3	1	**2.2**	0	2	0	2	2	1	0	1	1	1	**1.0**
Total												**2.4**											**2**

Note. T1 = Time 1. T2 = Time 2. St = Student. L = Level. S = Sentence. *M* = Mean. Light gray shadow = Improvement after treatment in the individual scores. Dark gray shadow = Improvement after treatment in means.

Appendix D: Comparison of unfilled pauses – t-test results

Comparison of number of pauses across Time (T1 vs T2) and Group (Control vs Experimental) – t-tests results

Sentence	T-test		df	*T*	*p*-value	Cohen's *d*
		C T2-T1	20	.370	.715	.0814
				-		-.250
	Paired	E T2-T1	20	1.142	.267	
		C T1-E T1	40	.302	.765	.078
1	Independent	C T2-E T2	40	1.227	.227	.382
		C T2-T1	20	-.484	.634	-.106
				-		-.278
	Paired	E T2-T1	20	1.276	.217	
		C T1-E T1	40	.209	.836	.061
2	Indepndent	C T2-E T2	40	.766	.448	.240
				-		-.365
		C T2-T1	20	1.673	.110	
	Paired	E T2-T1	20	.000	1	.000
		C T1-E T1	40	1.870	.069	.584
3	Independent	C T2-E T2	40	.717	.477	.232
		C T2-T1	20	-.852	.404	-.186
				-		-.304
	Paired	E T2-T1	20	1.395	.178	
		C T1-E T1	40	-.117	.908	.036
4	Indepedent	C T2-E T2	40	.475	.637	.142
		C T2-T1	20	-.241	.812	-.052
	Paired	E T2-T1	20	.678	.505	.148
		C T1-E T1	40	-.164	.871	.048
				-		.402
5	Independent	C T2-E T2	40	1.302	.201	
		C T2-T1	20	-.024	.981	-.005
				-		-.244
	Paired	E T2-T1	20	1.116	.278	
		C T1-E T1	40	-.240	.811	.074
6	Independent	C T2-E T2	40	.816	.419	.247
				-		-.254
		C T2-T1	20	1.164	.258	
				-		-.400
	Paired	E T2-T1	20	1.826	.083	
		C T1-E T1	40	.724	.473	.219
7	Independent	C T2-E T2	40	.830	.412	.253
		C T2-T1	20	.000	1	.000
				-		-.477
	Paired	E T2-T1	20	2.186	.041*	
		C T1-E T1	40	.416	.680	.130
8	Independent	C T2-E T2	40	1.860	.070	.573

		C T2-T1	20	.181	.858	.040
				-		-.623
	Paired	E T2-T1	20	2.855	.010**	
				-		.768
		C T1-E T1	40	2.480	.017*	
9	Independent	C T2-E T2	40	-.255	.800	.074
		C T2-T1	20	1.563	.134	.341
				-		-.260
	Paired	E T2-T1	20	1.188	.249	
		C T1-E T1	40	-.144	.886	.047
10	Independent	C T2-E T2	40	1.869	.069	.568

Note. C = control group. E = experimental group. T = time. Paired = paired-samples t-test. Independent = Independent-samples t-test. Cohen's *d* = Effect size.
* $p < .05$.