

ARTICLE



Learning pronunciation through television series

Pawel Scheffler, Adam Mickiewicz University

Karolina Baranowska, Adam Mickiewicz University

Abstract

While there is ample research on the effect of exposure to foreign language (FL) video materials on developing vocabulary knowledge and listening skills, research on the impact of watching videos on acquiring pronunciation skills, especially in terms of both perception and production, is still in its infancy. This study investigates the effect of viewing an English language video in different subtitling conditions (L1, L2, and no subtitles) on learners' recognition of the correct pronunciation of words and on learners' ability to produce words intelligibly. The relationship between recognition, intelligibility, and comprehension of the video material is also examined. In the study, 54 Polish intermediate learners of English were divided into three experimental groups based on the subtitling condition. They completed pre-tests and post-tests on recognition and production of words, as well as a comprehension test. The results show that watching the video, either with or without subtitles, significantly improved the recognition of how words are pronounced. As for gains in production, L1 subtitles turned out to be the least effective; however, they led to highest comprehension. These results have practical implications for teachers' use of video material in pronunciation work in the classroom.

Keywords: *Pronunciation Learning, Television Series, Multimodal Input, Informal Learning*

Language(s) Learned in This Study: *English*

APA Citation: Scheffler, P., & Baranowska, K. (2023). Learning pronunciation through television series. *Language Learning & Technology*, 27(1), 1–16. <http://hdl.handle.net/10125/73520>

Introduction

The Use of Video in Informal and Classroom Learning

As Vanderplank (2020, p. 183) said, there is a general belief that “one *should* be able to learn a foreign or second language from watching TV programs, films, and other audiovisual material in a foreign language” (emphasis in original). This belief is supported by numerous stories about informal language learning recounted in the media, in which learners testify to a lasting impact of television on foreign and second language development. For example:

Every day for about five years, Israeli sisters Reut and Shoham Nistel ran home from school, made themselves sandwiches and plopped down on the couch to watch an Argentine telenovela with Hebrew subtitles. The girls became so proficient in Spanish that they started speaking it at home to keep secrets from their parents. “That’s how we learned English, too,” said Reut, now 26. “We had English class in school, but I never paid attention. All my English is from ‘Full House’ and ‘Family Matters.’” (Cohen, 2018)

Such stories are part of a large pool of anecdotal evidence pointing to the pervasiveness and perceived usefulness of audiovisual materials in informal foreign language learning. This evidence is reinforced by findings from studies of learners’ perceptions of the usefulness of audiovisual materials. For example, Sockett (2014), in a comprehensive overview and analysis of French learners’ online activities in English,

found evidence of increasing popularity of television series among French learners. Interestingly, among the benefits of viewing such materials, the participants mentioned progress in pronunciation and comprehension skills. In a more recent project, Trinder (2017) used a questionnaire to investigate, inter alia, how frequently 175 students at the Vienna University of Economics engage in online activity in English, and how they perceive the usefulness of various technologies in learning English. The findings demonstrated that students viewed English films, TV series, and video clips “daily or frequently” (p. 405) and that they believed them to be entertaining and useful in fostering the development of listening, pronunciation, speaking, and communicative competence.

The bulk of experimental research into the amount of learning that actually takes place when learners watch films and television programmes has focused on vocabulary. This research has found that engagement with audiovisual materials leads to significant gains (e.g., Arndt & Woore, 2018; Baranowska, 2020; Koolstra & Beentjes, 1999; Montero Perez et al., 2013; Pujadas & Muñoz, 2019). Despite the fact that pronunciation is often referred to by learners as an area in which improvement through viewing videos takes place, this aspect of foreign language ability has been addressed by very few experimental studies (Charles & Trenkic, 2015; Wisniewska & Mora, 2020). It is this gap that the present study aimed to address.

Learning from Multimodal Input

The use of multimodal learning materials draws on Dual Coding Theory (Paivio, 1986) and Cognitive Theory of Multimedia Learning (Mayer, 2001), which suggest that the integration of audio and video should be more effective for language acquisition, information processing, and comprehension than a single-modality presentation. In the Second Language Acquisition (SLA) field, research on multimodality has only gained momentum in recent years, as is evidenced by a number of recent publications on the impact of exposure to audiovisual input on different aspects of L2 learning (Montero Perez, 2020). Most notably, frequent exposure to audiovisual input has been found to be conducive to vocabulary development (e.g., De Wilde et al., 2020; Peters, 2018; Puimège & Peters, 2020). There is also evidence that multimodal material aids L2 comprehension more than single-modality material. For instance, Whiting and Granoff (2010) found that the combination of audio and video input resulted in better comprehension of a short story by L2 learners of English than audio input alone.

While multimodal input may facilitate information processing, neither the audio nor the visual channel should be overloaded as it will impair learning. This is in line with Cognitive Load Theory, which postulates that human working memory capacity is limited (Sweller, 1988; Sweller, 1994). Therefore, when additional text in the form of L1 and L2 subtitles (also referred to as captions) is added to animation and narration, it could potentially overload the visual channel in accordance with the redundancy principle (Mayer, 2001). It is also possible that viewers may find it difficult to integrate auditory input processing with caption reading and as a result, captions may be cognitively overloading. Such an overload may impair learning since cognitive resources are seriously depleted and cannot be allocated to acquiring new information. However, whether or not subtitles lead to cognitive overload depends on a variety of factors, such as the level of proficiency and the speed of subtitle presentation (Wisniewska & Mora, 2020, p. 4).

Nevertheless, in L2 learning contexts, experimental research shows that subtitles are not redundant, and they in fact reinforce rather than hinder comprehension and learning. For example, in their research on the impact of subtitles on comprehension and vocabulary learning, Napikul et al. (2018) measured comprehension by means of a test based on Bloom’s taxonomy, which addresses different levels of comprehension. The results showed that L1 subtitles led to significantly greater comprehension than L2 subtitles and the absence of subtitles. Masrai (2020) focused specifically on the effect of L1 subtitles on listening comprehension skills and aural vocabulary knowledge through extensive exposure to foreign language (FL) videos. To measure listening comprehension skills, the researcher used two different versions of the listening sections of the International English Language Testing System (IELTS), one as the pre-test, and one as the post-test. After a 5-week period of out-of-class viewing of TV programs in the FL with native-language subtitles, the participants improved their overall listening comprehension skills. In a recent study, Baranowska (2020) directly compared the effectiveness of L1 and L2 subtitles and concluded that

the latter worked better for the acquisition of vocabulary by Polish secondary school learners of English. However, the L1 subtitle group had significantly higher scores in the comprehension test, concluding that there was a negative relationship between vocabulary acquisition and comprehension.

In contrast to vocabulary learning and listening comprehension, the impact of audiovisual material on foreign language speech ability has been researched much less extensively (Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015; Mitterer & McQueen, 2009). One question that has been addressed is whether subtitles facilitate the identification of words in a continuous stream of L2 speech. Mitterer and McQueen (2009) investigated the impact of L1, L2, and the absence of subtitles on accented speech perception in advanced learners of English. They found that L1 subtitles created lexical interference and inhibited speech perception when the participants were exposed to previously unheard utterances. By contrast, L2 subtitles were found to be conducive to developing speech perception skills in terms of old and new material.

Birulés-Muntané and Soto-Faraco (2016) replicated Mitterer and McQueen's (2009) study by exploring whether exposure to FL subtitled videos improves English speech perception, but in relation to intermediate learners whose L1 was a Romance language. The results were comparable in that L2 subtitles were found to be beneficial for speech perception skills. However, unlike Mitterer and McQueen's (2009) study, no detrimental effect of L1 subtitles was found. In their study on the effects of bimodal input on speech segmentation, Charles and Trenkic (2015) found that L2 subtitles improved the participants' ability to segment speech in a foreign language significantly more than the exposure to videos without subtitles. They also found that L2 subtitles help learners with speech segmentation in general, as exposure to subtitled videos led to the generalization of learning to previously unheard utterances.

The question of whether L2 subtitles influence the development of English as a Foreign Language (EFL) pronunciation skills was investigated in a recent study by Wisniewska and Mora (2020). The authors examined how different viewing modes (captioned in the L2 vs. uncaptioned) and task conditions (focus on phonetic form vs. focus on meaning) affected Spanish/Catalan adult EFL learners' speech processing and phonological accuracy in both speech perception and production. The treatment was extended over a period of eight weeks, with the goal of investigating the effect of regular exposure to bimodal input on speech processing and phonological accuracy. Phonological accuracy in perception was tested through a vowel contrast discrimination task, while speech production was examined by means of accentedness ratings of sentences pronounced by the participants.

The results showed that the learners improved their L2 speech processing skills regardless of the viewing mode. For example, they were able to segment English speech more efficiently. However, the benefits for phonological accuracy in perception were limited: it was only in the captioned meaning-focused condition that gains larger than in the control group were registered. As for production, there was improvement in two viewing modes: captions plus focus on meaning and no captions plus focus on pronunciation.

In the meaning focused condition, Wisniewska and Mora's (2020) participants were given comprehension questions to be answered while viewing the videos. The analysis of the responses revealed that the group with captions significantly outperformed the group without captions. However, since comprehension was tested only in the focus on meaning groups, it was not known to what extent a focus on pronunciation affected the overall comprehension of the clip (Wisniewska & Mora, 2020, p. 20).

Teaching and Learning Pronunciation

Approaches to English pronunciation teaching vary. In the past two decades or so, we have observed a shift from a traditional approach based on native-speaker accents to an intelligibility-based approach. The notion of intelligibility "refers to the extent to which an utterance is actually understood," whereas accentedness refers to "how strong the talker's foreign accent is perceived to be" (Munro & Derwing, 1995, p. 291). Yazan (2015, p. 202) points out that given the widespread use and acceptance of various non-native varieties of English, ensuring intelligibility is a very relevant goal for English teachers at present. However, what

teachers should perhaps strive for is the achievement by their learners of “comfortable intelligibility” (Kenworthy, 1987, p. 3), which means that learners’ utterances can be easily understood and their speech is not heavily accented. If this is not the case, then “speakers are judged to lack credibility and do not inspire confidence in either their knowledge or their persona” (Morley, 1994, p. 69).

Collins and Mees (2003) classified pronunciation errors into three categories. The first category encompasses errors that impede communication as they violate intelligibility (e.g., phonemic contrasts), the second category includes errors which do not prevent the speaker from being intelligible, but which are amusing or irritating (e.g., incorrect allophones), and the third category includes errors that are so insignificant that can be unnoticed (e.g., incorrect intonation). Thus, what can be considered so-called ‘correct pronunciation’ is pronunciation which is not necessarily native-like, but intelligible and free from global errors.

An important concept in pronunciation learning and in language learning in general is that of attending to input. As Schmidt (2001, p. 30) says, “[i]n order to acquire phonology, one must attend to the sounds of target language input.” However, different levels of attention have been distinguished in the SLA literature. The original concept developed by Schmidt (1990) involved the notion of *noticing*, which refers to learners consciously attending to input. In subsequent research (e.g., Tomlin & Villa, 1994), the concept of attention was extended to also cover detection without awareness. In his review of attentional mechanisms in SLA, Schmidt (2001) uses the term *noticing* to refer to the former and *registration* to describe the latter. It seems, then, that learners allocate different amounts of attentional resources to features of input, ranging from subconscious registration to conscious noticing and understanding (Housen & Pierrard, 2005, p. 7).

It is an open question how different levels of attention contribute to pronunciation learning. In general, the learning of new words occurs in stages which may require different levels of attention. In their review of research into incidental vocabulary learning from extensive reading, Waring and Nation (2004) stated that the first step in acquiring a new vocabulary item is the recognition of its form. Once a learner is familiar with the form, then the learning of other aspects of lexical knowledge can follow. In learning the spoken form, recognizing the correct pronunciation of a word, which entails the ability to perceive the relevant differences between sounds and suprasegmental features, may be considered the first step in learning how to produce it accurately.

The ability to recognize a word in L2 is, however, determined by a number of different factors. In some situations, phoneme discrimination is not sufficient, and the listener has to use context to compensate for the incomplete information at the phonetic level. For example, when the listener is exposed to homophones, they need to rely on the context to determine which word was used (Broersma & Cutler, 2008). Moreover, L2 listening involves the activation of different words which are not in the input. For example, the word “range” is likely to activate other words which are embedded in it, or which are pseudo-homophones (e.g., “ray,” “lay,” “rake,” and “lake”). This means that successful spoken word recognition depends on how much activation and competition occurs, since the more words are activated, the slower recognition becomes (Broersma & Cutler, 2008, pp. 24–25).

In order to facilitate the recognition of word boundaries in a stream of speech, listeners rely on different language-specific segmentation strategies. As Cutler (2000) explains, in stress-timed languages like English and Dutch, word boundaries are determined using a stress-based metrical segmentation strategy which relies on the contrast between strong and weak syllables. However, in French and Spanish, word boundaries are determined by means of syllable-based segmentation which rests on the assumption that syllable boundaries match word boundaries. Polish seems to be a mixed type in terms of its rhythmic structure: like syllable-timed languages it has no vowel reduction, and like stress-timed languages it incorporates complex syllable structures (Shoemaker & Rast, 2013). These properties of the Polish language may result in greater difficulty for Polish EFL learners than, for example, L1 Dutch learners in recognizing word boundaries in a stream of English speech.

In terms of measuring different levels of vocabulary knowledge, Waring and Nation (2004) also say that

different types of tests should be used. For example, sight recognition tests can be used to measure familiarity with a word's form, and translation tests to examine unassisted meaning recall. For research into the development of L2 pronunciation, this means that not only should learners' production be examined, but also that learners' ability to recognize the correct pronunciation of a given word should be assessed. In sum, pronunciation research should focus both on perception and production skills.

One recent empirical study which examines the effect of perception and production practice on pronunciation learning is that of Li and DeKeyser (2017). The authors set out to test whether the effects of practice are skill specific (i.e., whether perception practice mostly benefits perception skills and production practice mostly production skills). The feature that was investigated was Mandarin word tones. In the study, two groups of adult speakers of English went through explicit instruction concerning Mandarin tones followed by either perception or production practice sessions. When tested, each group outperformed the other in the skill they were taught, corroborating the skill specificity hypothesis.

The Study

Aims and Research Questions

As shown in the preceding sections, learners perceive viewing English language films and television series as an effective means of improving EFL pronunciation. However, research that aims to verify these claims is scarce. The work done in the area of spoken language processing (e.g., Birulés-Muntané & Soto-Faraco, 2016; Charles & Trenkic, 2015; Mitterer & McQueen, 2009), in general, demonstrates a positive impact of L2 subtitles on speech segmentation. The study by Wisniewska and Mora (2020), which is described as “a first attempt at exploring whether extended exposure to multimodal input through captioned video can benefit pronunciation development” (p. 5), offers some evidence that L2 subtitles can enhance the accuracy of sound production provided that learners' attention is not focused on phonetic form.

Some of the issues that have not been addressed in previous research concern the effect of L1 subtitles (rather than L2 subtitles/captions) on speech perception and production; the impact of L1 and L2 subtitles on intelligibility (rather than accentedness); the effect of a single (rather than extended) exposure to subtitled audiovisual material on pronunciation development; and the effect of L1/L2 subtitled material on pronunciation in relation to its comprehension. The main aim of the present study is to fill the abovementioned gaps by investigating whether viewing television series in different subtitling conditions can lead to recognizing how words are pronounced and whether such exposure can contribute to progress in producing English words intelligibly. An additional aim is to examine whether there is a relationship between pronunciation learning from TV material and the level of comprehension achieved. After all, as is evident from students' comments in Sockett's (2014) study, comprehension is a major concern for learners (and viewers in general) and ideally any learning that takes place while viewing TV material should not interfere with it.

In accordance with these aims, the following research questions were formulated:

1. What are the relative effects of viewing English language television series with L1, L2, or without subtitles on EFL learners' recognition of the pronunciation of words?
2. What are the relative effects of viewing English language television series with L1, L2, or without subtitles on the intelligibility of EFL learners' pronunciation?
3. Is there a relationship between (a) the level of EFL learners' recognition and intelligibility of English pronunciation resulting from viewing television series in different subtitling conditions and (b) their comprehension of TV material?

Methodology

Participants

The participants of the study were 54 Polish intermediate (B1) learners of English recruited from a

secondary school in Sompolno, a small town in the Greater Poland Voivodship. Their age ranged from 16 to 18 years old and they were all native speakers of Polish. Their level of proficiency was initially determined on the basis of the coursebook they were using and later by means of an online test of proficiency LexTale (Lemhöfer & Broersma, 2012) conducted just before the experiment. All the participants reported not having watched *Gilmore Girls* (the TV-series used in the experiment) before, meaning they had no prior knowledge of the experimental clip which could affect their performance.

Materials

Lexical Test for Advanced Learners of English (LexTale) was used to measure participants' level of proficiency as it is a quick and valid indicator of proficiency in English (Lemhöfer & Broersma, 2012). The audiovisual material used in the experiment was a clip adapted from the first episode of *Gilmore Girls*, an American television series. The clip was approximately 23 minutes long. The following words from the video clip were selected for the experiment: "dairy," "vacancy," "pursuit," "reenact," "actually," "immediately," "cliché," "unbelievable," "doubt," "acknowledgement," "knock," and "democracy." According to the Cambridge vocabulary list, the aforementioned words are mostly above B1 level of proficiency and thus the participants were unlikely to be familiar with the majority of them. Further, the pronunciation of these words could be expected to present difficulty for B1 learners of English as most of them are multisyllable words, contain silent letters, or differ from Polish in their stress patterns (e.g., "democracy"). Even if the word "actually" is an A2 level word, its pronunciation may pose a problem for Polish learners of English, which is the reason why it was also included in the experiment

In order to determine whether the participants were capable of recognizing the pronunciation of the selected words in the input, a pre-test and a post-test were administered. The test contained a list of the 12 words. Each word was accompanied by two numbers: 1 and 2. The participants' task was to mark which pronunciation of the word, 1 or 2, was correct. The two pronunciations of each word were produced by the second author. The incorrect pronunciation deviated from the correct one in terms of stress, vowel quality, or the realization of silent letters. As for the pre-test and post-test on production, the participants were presented with a list of 12 sentences. Each sentence contained one word from the pool of words selected for the study. The sentences were taken from the *Oxford Advanced Learner's Dictionary* (Hornby, 2010). In the production tasks (both pre-test and post-test), the words were not presented in isolation, but embedded in sentences to divert the participants' attention and to prevent them from focusing on the pronunciation of the target words. In this way, the participants read the words in a natural way (i.e., they focused on the whole sentence rather than on a clear articulation of a specific word). Moreover, this procedure helped to make sure that while watching the video clip later, students would not focus on the pronunciation of the selected words. The test on comprehension consisted of ten open-ended questions about the content of the video clip, for example, "What happened to Lorelai when she was 16?," "Why did Lorelai call Miss Bell?," and so on. All the questions were in the participants' native language (i.e., Polish).

Procedure

Overall, the experiment took four consecutive 45-minute class periods for each experimental group. Due to time constraints, two groups participated in the experiment on the same day, and one group the following day. The participants were randomly assigned to three experimental groups, with 18 students in each of them. The only difference between the groups was the viewing condition—one group watched the video clip without subtitles, one with L1 (i.e., Polish), and one with L2 (i.e., English) subtitles. The following procedure was implemented in each group.

First, the participants completed the test to determine their level of proficiency. The students who did not qualify as B1 learners of English were excluded from the participation in the experiment. After the proficiency test, one of the researchers invited each participant individually to the classroom, where they were asked to read aloud the sentences containing the relevant words. This was recorded using a USB microphone connected to a laptop computer. Immediately after the session with individual recordings, the participants were gathered in one classroom and completed the pre-test on recognition. Then, the group

watched the video clip in a given subtitling condition. The participants watched the clip together on one screen, and they were supervised by their teacher and one of the researchers so as to prevent them from conversing and discussing the content of the video material while they were watching it. The students were instructed to watch the movie clip carefully and focus on meaning since after the viewing they would be asked comprehension questions. Immediately after the viewing session, the participants filled in the test on comprehension. Then, after all tests had been collected, each participant was invited individually for the post-test on production, which again involved the recording of 12 sentences read aloud by each participant. Finally, the participants were gathered together in the classroom and completed the post-test on recognition.

Data Scoring

In the recognition task, one point was awarded for identifying each correctly pronounced word. In the production task, each word was rated on a 3-point scale described in Thomson and Derwing (2016). The scale is a hybrid intelligibility-accent rating in which listeners judge whether they understand an utterance and whether it is a good or poor exemplar of a given category. Thus, no points were awarded for 'unintelligible' pronunciation, one point for 'intelligible but poor,' and two points for 'intelligible and good.' Additionally, a listener transcription technique was employed, meaning the raters were also asked to transcribe the words they heard (Munro & Derwing, 1995, p. 291).

The pronunciation of the selected words was evaluated by two independent raters who were experienced Polish academic teachers of English familiar with American accents. They evaluated the pronunciation of each word via the online e-learning platform Moodle. The percentage agreement between the two raters was 81.3%. If the first two rated the pronunciation differently, a third rater, who was also the second author of this article, evaluated the recordings. The pre-test and post-test performances were randomly presented to the raters, so that they were blinded to the time at which the tasks were completed (during the pre-test or post-test). This ensured a fair and valid judgement since the raters' expectations could not influence their evaluations. In addition, the raters could listen to individual words only once to prevent them from guessing the words from context as the aim was to assess the intelligibility of the words in isolation rather than the ability to infer their meaning on the basis of the surrounding context.

The target words were excised from the sentences for evaluation. Despite the fact that there may be a difference between naturally produced words in isolation and digitally excised words in isolation (e.g., Ernestus et al., 2002), this should not have been a problem in the current study. Since the target words were all content words, always produced in the full, not reduced form, this means that it was easy for the raters to understand them provided they were produced intelligibly. Also, the excised words were taken from every participant group, which means that this was not a confounding factor.

The recognition test as well as the comprehension test were graded by the second author since unlike the production test, these two tests had objective scoring keys (i.e., the answer was either correct or incorrect and there was no judgement involved). Still, the recognition and the comprehension tests were graded randomly by the second author so that she was blinded to the condition that a given participant was in.

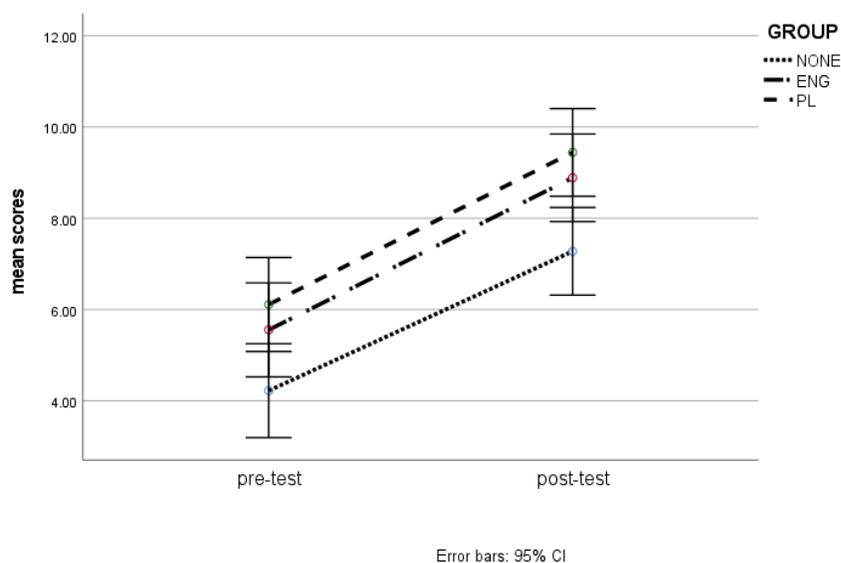
Results

Our first research question explored the impact of different viewing conditions on the recognition of correct pronunciation. [Table 1](#) provides descriptive statistics for the pre-test and post-test recognition scores. First, a one-way ANOVA was conducted to compare the scores of the three groups on the pre-test. The analysis uncovered a statistically significant difference ($F(2,51) = 3.585, p = .035.$), with the no-subtitles group being significantly different from the L1 group ($p = .032$), as indicated by the Tukey HSD test. Given that there was a statistically significant difference between two of the groups at the pre-test, a one-way ANCOVA was conducted to compare the impact of the three subtitling conditions on the recognition of the pronunciation of words while controlling for the participants' initial knowledge. The analysis found no significant difference in mean scores on the post-test between the groups [$F(2,50) = 2.61, p = .083$].

Table 1*Descriptive Statistics for the Pre-test and Post-test Recognition Scores*

	Group	<i>M</i>	<i>SD</i>
Pre-test	L1 subtitles	6.11	2.19
	L2 subtitles	5.55	2.57
	No subtitles	4.22	1.66
	Average	5.29	2.27
Post-test	L1 subtitles	9.44	2.20
	L2 subtitles	8.88	1.77
	No subtitles	7.27	2.08
	Average	8.53	2.19

The final analysis for RQ 1 compared the pre- and post-test means of the participants in the three viewing conditions. The paired samples *t*-test demonstrated that there was a significant difference in the number of correctly recognized words at the pre-test and at the post-test for each of the groups (L1 group: $t(17) = -5.279, p < .001$; L2 group: $t(17) = -5.831, p < .001$; No-subtitles group: $t(17) = -7.640, p < .001$). [Figure 1](#) illustrates the performance of the three groups.

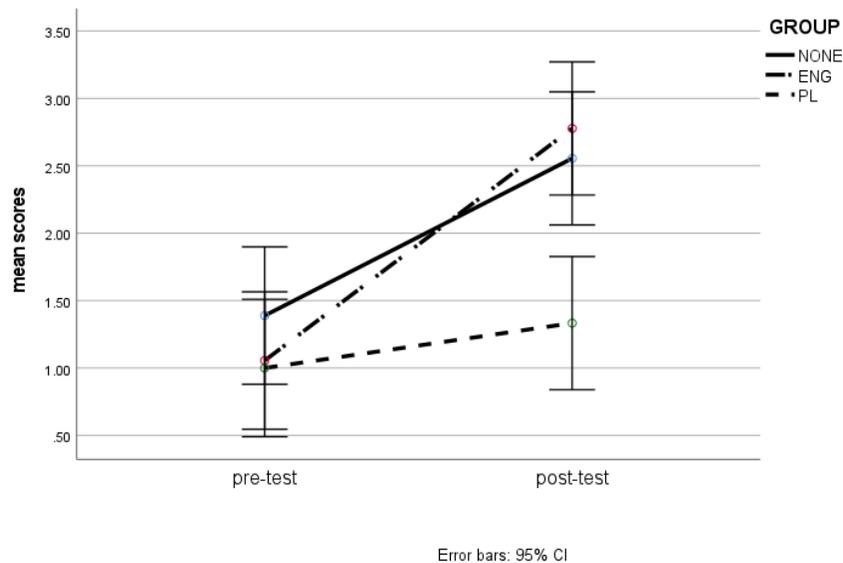
Figure 1*Results of Recognition Tests by Group and Time*

RQ 2 concerned differences between the groups on the production measure. Descriptive statistics for the pre-test and post-test production scores are presented in [Table 2](#). A one-way ANOVA was first run to check for any differences between the groups on the pre-test. The analysis showed that the three groups did not differ significantly ($F(2,51) = 0.686, p = .508$).

Table 2*Descriptive Statistics for the Pre-test and Post-test Production Scores*

	Group	Mean	Standard Deviation
Pre-test	L1 subtitles	1.00	0.97
	L2 subtitles	1.05	1.25
	No subtitles	1.38	0.97
	Average	1.14	1.07
Post-test	L1 subtitles	1.33	0.90
	L2 subtitles	2.77	0.87
	No subtitles	2.55	1.29
	Average	2.22	1.20

Since the groups had the same starting point, a mixed ANOVA was performed. The analysis revealed a statistically significant interaction between the time and group variables ($F(1,51) = 13.15$, $\eta^2 = 0.34$, $p < .001$). This means that for the participants, any changes in production between the pre-test and the post-test depended on membership in a particular group. In other words, the changes over time were not equivalent across the three groups. This is illustrated in [Figure 2](#) below.

Figure 2*Results of Production Tests by Group and Time*

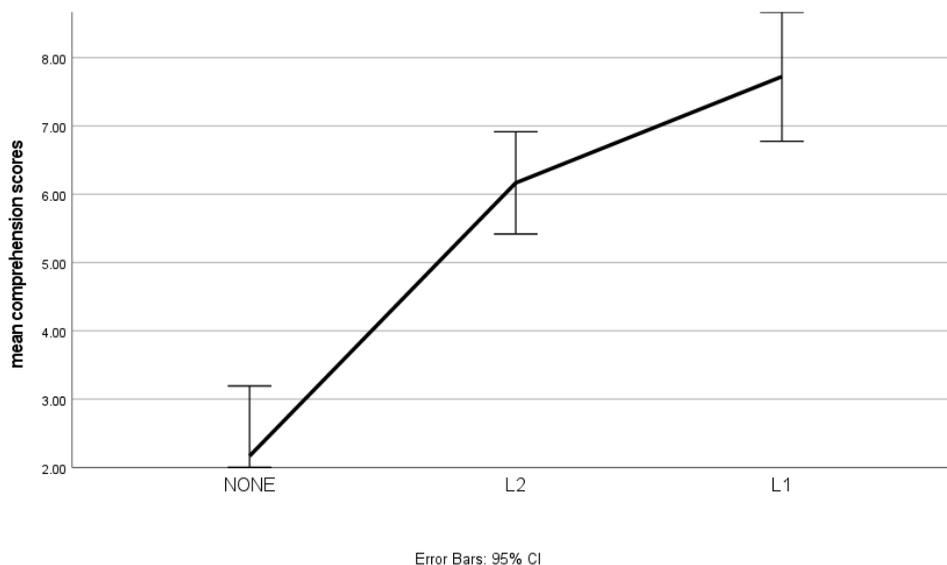
Pairwise comparisons were also run to specify which groups differed from each other and at which time point, as well as which groups improved significantly. The comparisons confirmed that there were no statistically significant differences between any of the groups in the pre-test and demonstrated that on the post-test the L1 subtitles group differed significantly from the other two ($p = .003$). As for the progress that was made by each group, significant differences were obtained for the L2 and the no subtitles group.

For RQ 3, first, the results of the comprehension test of the groups were compared. One-way ANOVA was

employed to determine the relationship between subtitling conditions and comprehension. Overall, the analysis yielded statistically significant differences between the groups ($F(2,51) = 43.69, p < .001$). The L1 subtitles group had the highest comprehension score ($M = 7.7$) and the no-subtitles group the lowest ($M = 2.16$). A Tukey post hoc test revealed that there was a statistically significant difference between the no-subtitles and the L2 subtitles group ($p < .001$), the no-subtitles and the L1 subtitles group ($p < .001$), and the L2 and L1 subtitles groups ($p = .037$). [Figure 3](#) shows mean comprehension scores obtained by each group.

Figure 3

Mean Comprehension Scores Obtained by Each Group



In assessing the relationship between learners' comprehension and pronunciation, only production scores were taken into account. The rationale behind this is that recognition scores differed significantly within groups, that is when the results of the pre-test and post-test were compared for each group, but did not differ significantly between the groups on the post-test. For comprehension and post-test production scores, the groups were ranked based on significant differences in their mean scores, ordered from highest to lowest. The procedure produced the orders displayed in [Table 3](#).

Table 3

Learners' Groups Ranked From Highest to Lowest according to Significant Differences in Mean Comprehension and Post-test Production Scores

Rank	Comprehension	Production
Highest	L1 subtitles	L2 subtitles/No subtitles
	L2 subtitles	L1 subtitles
Lowest	No subtitles	

Additionally, the differences in mean perception and production gain scores within each viewing condition were juxtaposed to compare learners' gains in these two different skills. As Table 4 shows, the differences in mean perception and production gain scores within each viewing condition produced effect sizes exceeding Cohen's (1988) convention for a large effect ($d = 0.80$).

Table 4

Differences in Mean Perception and Production Gain Scores Within Each Viewing Condition

	Mean Perception Gain	Mean Production Gain	Cohen's d
L1 group	+ 3.33	+ 0.33	1.60
L2 group	+ 3.33	+ 1.72	0.87
No-subtitles group	+ 3.05	+ 1.16	1.38

Discussion

The results of the tests demonstrated a differential effect of viewing conditions on the development of pronunciation ability in terms of recognition and production. The ability to recognize correct pronunciation of a word was not subject to influence by the choice of viewing condition: all the three groups exhibited significant improvement and when the pre-test differences between them were controlled, the differences on the post-test were not statistically significant. Furthermore, the gains in the recognition scores were not related to comprehension scores: each group improved regardless of their comprehension level. For teachers, the conclusion from these results can be that viewing videos enhances students' recognition of how words are pronounced regardless of subtitling, and that improvement in this area is independent of how well students understand the audiovisual material.

As far as production ability is concerned, the differences between the subtitling conditions turned out to be statistically significant. The L1 subtitles group had significantly lower scores than both other groups. Interestingly, L1 subtitling resulted in the highest comprehension scores. This indicates a negative relationship between comprehension and the acquisition of productive speech ability: the more comprehension, the less progress in production ability. This also means that two sources of input need not always confer significant advantages in terms of L2 acquisition (Dual Coding Theory; Paivio, 1986). It seems that at least in the context of viewing subtitled audiovisual material, advantages for production ability are significant if learners are exposed to bi-modal input in which both the aural and the written stimuli are encoded in learners' L2, provided learners are not overburdened by task focus. In Wisniewska and Mora's (2020) study, learners were engaged in a form-focused task. In the present study, learners were engaged in a comprehension task as they were asked to watch the video carefully in preparation for comprehension questions. In terms of sound production, the current study showed that, if L1 subtitles are used and if learners' attention is in some way directed to meaning, then L2 acquisition may be compromised. This result is in line with the results of Baranowska's (2020) experiment concerning vocabulary learning. This means a combination of focus on meaning and L1 subtitles results in learners neglecting the aural stimulus. This effect may further be strengthened by the fact that learners generally read textual input before its auditory presentation, as shown by Wisniewska and Mora's (2020) findings concerning text-sound synchronization.

The results of the LexTale test confirmed that participants were B1 level EFL learners. Therefore, the episode that they watched was probably hard for them to follow as without the support of subtitles: the mean comprehension level of the authentic English language material they were exposed to was only around 20%. Given that the participants were instructed to comprehend as much as possible, students in the L1

group may have focused on the native language subtitles to achieve this goal and as a result, they did not have enough cognitive resources left to attend to the aural input. Since attending to the sounds of target language input is seen as a precondition for the acquisition of phonology (Schmidt, 2001, p. 30), failure to attend to the relevant sounds may explain a lack of improvement in production.

Furthermore, there may be a relevant distinction in this context between registration and noticing. As observed earlier, learners allocated different amounts of attentional resources to the input. Therefore, they might have registered or detected the features of input without awareness, or they might have consciously noticed its elements. This means if the learners in the L1 group had allocated their conscious attention to the native language subtitles—following the generally accepted view in psychology which explains that attention has a limited capacity (Sweller, 1988; Sweller, 1994)—they could only be able to subconsciously register most of the aural stimulus. However, while such registration was generally sufficient for the recognition of the target words' correct pronunciation in the post-test, it did not allow for developing intelligible production.

The students in the present study received information through the visual channel (picture and text) as well as the auditory channel (soundtrack). In the light of Cognitive Theory of Multimedia Learning (Mayer, 2001), dual modality input (audio and video) should benefit viewers and facilitate information processing and learning (e.g., De Wilde et al., 2020; Puimège & Peters, 2020; Whiting & Granoff, 2010). However, in accordance with Cognitive Load Theory (Sweller, 1988; Sweller, 1994), neither the audio nor the visual channel should be overloaded as this impedes learning given that working memory capacity is limited. Therefore, when there are two different forms of visual stimuli available (picture and text), this could potentially overload the visual channel and as a result, the viewer may not be able to successfully integrate the three sources of input (picture, text, and audio).

In the current study, the participants in all groups were given a comprehension task, so they sought to employ different techniques that would facilitate the completion of the task. The students in the L1 group, guided by the demands of the comprehension task, decided to ignore the audio input since the visual channel provided all the information they needed. Therefore, the information delivered via the auditory channel was simply redundant in relation to the task at hand for this group. In this way, they perhaps avoided cognitive overload which could have occurred if they had tried to focus on both native language text and foreign language audio. It is likely that more advanced English learners would not have needed to prioritize the visual channel in order to prepare for comprehension questions following this kind of viewing session, and they would have managed to process the information from both channels. On the other hand, for students in the L2 group, also B1 level learners of English, the English subtitles are likely to have posed a challenge and were not as appealing a comprehension aid as they were for students in the L1 group. So, in order to try to fully comprehend what was said, they relied on both the textual and the audio input. In more general terms, what follows from the above discussion is that cognitive overload may be conditioned by a complex set of factors which include presentation modality, task demands, and learners' language level.

In terms of comprehension, the findings of the present study are largely in line with those of previous research which indicated that both L1 and L2 subtitles facilitate comprehension (e.g., Baranowska, 2020). Even though L1 subtitles have generally been found to be the most beneficial for comprehension (e.g., Baranowska, 2020; Masrai, 2020; Napikul et al., 2018), the present study's results indicated that when the focus is on meaning, L2 subtitles can also significantly support comprehension, which was also found in Wisniewska and Mora's (2020) study. In the present study, the students in the L2 group used the English subtitles as an aid in understanding the storyline, but this did not consume enough attentional resources to block the noticing of the aural input. So, perhaps in L1 contexts similar to the one in this study, if the teacher's goal is to use audiovisual material to help B1 EFL learners improve the intelligibility of the words they produce while ensuring comprehension around the level of 60%, then, L2 subtitled videos can be used.

While there was statistically significant improvement in production scores in two of the groups, the scores were low. For two reasons, however, this should not come as a surprise. First, the students viewed the episode only once and there was only one instance of each of the selected words in it. Experiments in word

learning from extensive reading (Waring & Nation, 2004) tell us that to ensure effective word learning, numerous encounters with a given word are needed. Thus, one can assume that more substantial progress in intelligible word production would also require repeated viewing of audiovisual material. Further, as is the case with extensive reading (Waring & Nation, 2004), repeated viewing will also be needed to secure retention of the learned material over time. This was not examined in the present study, which is one of its limitations.

The relatively modest amount of progress in word production exhibited by the participants can also be related to research findings of studies comparing input and output practice within the framework of skill acquisition theory (e.g., for L2 morphosyntax, see DeKeyser, 1997; DeKeyser & Sokalski, 1996; Rodgers, 2011; for L2 pronunciation, see Li & DeKeyser, 2017). The general conclusion reached by these researchers is that the effects of input and output practice are skill specific, which is to say that input practice is more effective for reception skills and output practice results in significant advantages for production skills.

Moreover, in line with the skill specificity hypothesis, one should expect larger gains in this study's recognition task, resulting from participants' exposure to input which contained foreign language pronunciation models. The results of the study demonstrate that it was indeed the case since production gains were much lower than perception gains.

Conclusion

The present study sought to find empirical support for anecdotal claims that watching FL videos improves pronunciation skills, thus attempting to bridge the gap between research and real-world situations. Pronunciation skills were investigated in terms of both perception and production in relation to intelligibility rather than accentedness, reflecting the current pronunciation literature. The relationship between pronunciation and comprehension was also examined.

The significance of the present study lies in its contribution to addressing an important gap in existing research on the impact of exposure to subtitled audiovisual input on FL pronunciation learning. In very general terms, the study showed that subtitled programmes have a positive impact on developing L2 pronunciation. More specifically, however, the results revealed a complex pattern of interactions between the variables that were investigated. While the viewing condition did not interfere with the recognition of L2 sounds, it was relevant to their production. A possible explanation for this involves differences in how attentional resources were allocated in each condition. In the L1 group, learners directed most of their conscious attention to the native language subtitles in order to enhance comprehension. This resulted in insufficient resources being left for the acquisition of production skills in this group.

The general advantage for recognition over production, regardless of the viewing condition, may be related to the role of input and output practice as predicted by the skill specificity hypothesis (i.e., learners who practise production should have gains in production performance, and learners who practice perception should have gains in perception performance). The present study, which provided input but not output practice, showed that the scores in pronunciation recognition were significantly higher than production.

The present study can be taken as preliminary evidence that even a one-off viewing session without an explicit focus on L2 sounds can contribute to better L2 pronunciation. This is of course good news for learners and teachers, at least for those studying and working within a similar L1-L2 context. Future research should include other language pairs and perhaps also conditions with a direct phonetic focus.

Acknowledgements

We would like to thank the teachers and students from Zespół Szkół Ogólnokształcących i Technicznych in Sompolno for taking part in our study.

We would also like to thank the three anonymous reviewers for their useful suggestions and comments on

earlier versions of this article.

Finally, we wish to thank our colleagues and researchers from Adam Mickiewicz University: Prof. Waldemar Wołyński for providing support with statistical analyses; Prof. Anna Dziemianko and Ewa Grzywaczewska-Stewart, M.A., for helping us with the auditory data; and Dr. Jarosław Weckwerth for technical support.

References

- Arndt, H. L., & Woore, R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning & Technology*, 22(1), 124–142. <https://doi.org/10.125/44660>
- Baranowska, K. (2020). Learning most with least effort: Subtitles and cognitive load. *ELT Journal*, 74(2), 105–115. <https://doi.org/10.1093/elt/ccz060>
- Birulés-Muntané, J., & Soto-Faraco, S (2016). Watching subtitled films can help learning foreign languages. *Plos One*, 11(6), e0158409. <https://doi.org/10.1371/journal.pone.0158409>
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36(1), 22–34. <https://doi.org/10.1016/j.system.2007.11.003>
- Charles, T. J., & Trenkic, D. (2015). Speech segmentation in a second language: The role of bi-modal input. In Y. Gambier, A. Caimi, and C. Mariotti (Eds.). *Subtitles and language learning: Principles, strategies and practical experiences* (173–178). Peter Lang.
- Cohen, E. (2018). The television trick to learning a new language. *CNN*. <https://edition.cnn.com/2018/03/19/health/learn-new-language-telenovela-trick/index.html>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Routledge Academic.
- Collins, B., & Mees, I. (2003). *The phonetics of English and Dutch*. Brill.
- Cutler, A. (2000). Listening to a second language through the ears of a first. *Interpreting*, 5(1), 1–23. <https://doi.org/10.1075/intp.5.1.02cut>
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 23(1), 171–185. <https://doi.org/10.1017/S1366728918001062>
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221. <https://doi.org/10.1017/S0272263197002040>
- DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46(4), 613–642. <https://doi.org/10.1111/j.1467-1770.1996.tb01354.x>
- Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1–3), 162–173. <https://doi.org/10.1006/brln.2001.2514>
- Housen, A., & Pierrard, M. (2005). Investigating instructed second language acquisition. In A. Housen and M. Pierrard (Eds.). *Investigations in instructed second language acquisition* (pp. 1–27). Mouton de Gruyter.
- Hornby, A. S. (2010). *Oxford advanced learner's dictionary*. Oxford University Press.
- Kenworthy, J. (1987). *Teaching English pronunciation*. Longman.

- Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitles television programmes at home. *Educational Technology Research and Development*, 47(1), 51–60. <https://doi.org/10.1007/BF02299476>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTale: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Li, M., & DeKeyser, R. M. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, 39(4), 593–620. <https://doi.org/10.1017/S0272263116000358>
- Masrai, A. (2020). Can L2 phonological vocabulary knowledge and listening comprehension be developed through extensive movie viewing? The case of Arab EFL learners. *International Journal of Listening*, 34(1), 54–69. <https://doi.org/10.1080/10904018.2019.1582346>
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *Plos One*, 4(11). <https://doi.org/10.1371/journal.pone.0007785>
- Montero Perez, M. (2020). Multimodal input in SLA research. *Studies in Second Language Acquisition*, 42(3), 653–663. <https://doi.org/10.1017/S0272263120000145>
- Montero Perez, M., Van Der Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.), *Pronunciation theory and pedagogy: New views, new directions* (pp. 64–91). TESOL.
- Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38(3), 289–306. <https://doi.org/10.1177/002383099503800305>
- Napikul, S., Cedar, P., & Roongrattanakool, D. (2018). The effects of film subtitles on English listening comprehension and vocabulary. *International Journal of Applied Linguistics & English Literature*, 7(6), 104–111. <http://dx.doi.org/10.7575/aiac.ijalel.v.7n.6p.104>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL—International Journal of Applied Linguistics*, 169(1), 142–168. <https://doi.org/10.1075/itl.00010.pet>
- Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 42(3), 525–549. <https://doi.org/10.1017/S027226311900055X>
- Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47(4), 479–496. <https://doi.org/10.1080/09571736.2019.1616806>
- Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *International Review of Applied Linguistics in Language Teaching*, 49(4), 295–319. <https://doi.org/10.1515/iral.2011.016>
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>

- Schmidt, T. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.
- Shoemaker, E., & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research*, 29(2), 165–183. <https://doi.org/10.1177/0267658313479360>
- Sockett, G. (2014). *The online informal learning of English*. Palgrave Macmillan.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Thomson, R. I., & Derwing, T.M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88–97). Iowa State University.
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183–203. <https://doi.org/10.1017/S0272263100012870>
- Trinder, R. (2017). Informal and deliberate learning with new technologies. *ELT Journal*, 71(4), 401–411. <https://doi.org/10.1093/elt/ccw117>
- Vanderplank, R. (2020). Video in informal language learning. In M. Dressman and R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 183–202). Wiley-Blackwell.
- Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World*, 4, 97–110.
- Whiting, J., & Granoff, S. (2010). The effects of multimedia input on comprehension of a short story. *The Electronic Journal for English as a Second Language*, 14(2), 1–10.
- Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, 42(3), 599–624. <https://doi.org/10.1017/S0272263120000029>
- Yazan, B. (2015). Intelligibility. *ELT Journal*, 69(2), 202–204. <https://doi.org/10.1093/elt/ccu073>

About the Authors

Paweł Scheffler is a researcher in the Faculty of English at Adam Mickiewicz University in Poznań. His research interests are second language acquisition, classroom foreign language instruction, and modern English grammar. He also writes language teaching materials for Polish learners of English.

E-mail: spawel@amu.edu.pl

Karolina Baranowska is a researcher in the Faculty of English, Adam Mickiewicz University in Poznań. Her research interests include second language acquisition, methods of teaching English as a foreign language, and cognitive load theory. She is also involved in teaching English as a foreign language at an academic level.

E-mail: karbar@amu.edu.pl