

To What Extent Is Spoken English Proficiency of Chinese University Student Relate to the Auditory Vocabulary Acquired through Watching English Program?

Hui Liao, Guohong Li

School of Foreign Languages, Xihua University, Chengdu, China

Email: 1294116639@qq.com

How to cite this paper: Liao, H., & Li, G. H. (2025). To What Extent Is Spoken English Proficiency of Chinese University Student Relate to the Auditory Vocabulary Acquired through Watching English Program? *Open Journal of Modern Linguistics*, 15, 794-803. <https://doi.org/10.4236/ojml.2025.155046>

Received: August 14, 2025

Accepted: September 22, 2025

Published: September 25, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study adopts the design of pre-test, watching English TV programs and timely post-test to test the correlation between watching English TV series and the oral ability of English as a second language (ESL) learners, as well as the correlation between listening vocabulary acquired through watching English TV series and the oral ability of ESL learners. Participants (N = 45) were stratified by pre-test vocabulary scores to ensure baseline equivalence. In this study, 45 college students from a Chinese University were divided into three groups through questionnaires. The purpose of the questionnaire is to divide learners into different frequency continuums of watching English TV programs according to their habits, duration and frequency of watching English TV programs. Each group presented the English TV series with English subtitles in the same way. The oral test before watching TV, the listening vocabulary test after watching TV and the plot retelling test were set. The results showed a significant positive correlation between viewing frequency and oral proficiency ($r = 0.42$, $p < 0.01$), with high-frequency viewers (≥ 5 h/week) scoring 15% higher on IELTS speaking than low-frequency viewers. Auditory vocabulary gains mediated 32% of this relationship (Sobel test $z = 2.67$, $p = 0.008$). This study provided further support for the use of second language (L2) television programs for language learning.

Keywords

Proficiency in Spoken English, English TV Programs, Auditory Vocabulary

1. Introduction

There are two main cognitive modes that suggest that watching to learn vocabulary may be effective. Multimedia learning theory is based on the idea that there are separate channels in working memory for processing words and pictures (Mayer, 2009). Dual coding theory states that human cognition consists of two coding systems, one is a speech system that deals with objects encoded in verbal modality, and the other is an image system that deals with objects encoded in non-verbal modality (Paivio, 1986, 1971). Both theories suggest that the simultaneous presentation of information in the form of words and images promotes learning, suggesting that learning through viewing may be better than reading and listening. Second language learners have a variety of language learning input methods, including reading English books, watching English movies, English talk shows, listening to English songs, etc. Nowadays, watching English movies with subtitles has gained global popularity. One of the reasons for this trend is the belief that perceptual learning of foreign language sounds will improve the perceptual skills of non-English speakers (Logan, Lively, & Pisoni, 1991; Bradlow, 2008).

2. Literature Review

At present, studies have proved that watching TV can carry out incidental vocabulary learning. Watching full-length TV documentaries, reading their texts, and listening to their audio all contribute to significant incidental vocabulary learning, and there is no statistically significant difference in the amount of vocabulary knowledge obtained through the three modes (Feng & Webb, 2019). The facilitative role of subtitled audiovisual materials in foreign language learning has been well-documented in literature (e.g., Vanderplank, 2016; Talaván, 2013). For instance, research has shown that watching subtitled movies can enhance vocabulary acquisition, listening comprehension, and the understanding of pragmatic cues. At present, there are also studies demonstrating the interaction between vocabulary and spoken language (e.g., Enayat & Derakhshan, 2021b; Miralpeix & Munoz, 2018; Uchihara & Clenton, 2018). Although receptive vocabulary contributes more to spoken L2 than productive vocabulary knowledge and vocabulary depth (Webb & Chang, 2022), research on receptive vocabulary remains limited. Using the auditory Lex and a partial dictation test, which are argued to tap into cognitive processes shared with speaking (e.g., Field, 2008; Milton & Alexiou, 2010), their relevance to the oral test will be further tested.

According to relevant articles, there are few studies to prove the relationship between watching English TV series and how the auditory vocabulary acquired through English TV series is related to the oral proficiency of L2 learners. Due to the limited time available to teach vocabulary in the classroom, researchers encourage incidental learning of vocabulary through exposure to L2 input as a means of extending learners' vocabulary knowledge (e.g., Pellicer-Sánchez, 2017; Nation, 2001; Peters & Webb, 2018; Webb, 2020).

However, with the increasing popularity of television and film, scholars have

begun to investigate the potential of audiovisual input as a source of L2 vocabulary (Vanderplank, 1988), and later research has provided robust evidence on the lexical coverage and learning potential of such materials (Webb & Rodgers, 2009). Although some recent studies (Montero Perez, 2020; Peters & Webb, 2018) have shown that words can be learned by chance through watching television, research on vocabulary learning from audiovisual input remains sparse. This paper tries to fill this gap by making a comparative analysis of the relationship between L2 learners' audiovisual habits, that is, the input of watching English TV series, and L2 learners' oral English proficiency, as well as the relationship between English vocabulary acquired through English TV series and their oral English proficiency. It is hoped that this study can provide an effective oral learning method for L2 learners and encourage the diversity of language input patterns.

3. The Present Research

3.1. Research Questions

- 1) What is the relationship between the oral English proficiency of Chinese college students and the frequency of watching English TV series?
- 2) What is the relationship between the oral English proficiency of Chinese college students and the auditory vocabulary acquired through watching English TV series?

3.2. Hypothesis

- 1) The oral English proficiency of Chinese college students is positively correlated with the frequency of watching English TV series, that is, the higher the frequency of watching English TV series, the higher the oral English proficiency of college students.
- 2) Auditory vocabulary gains from TV viewing positively correlate with learners' oral proficiency, that is, the more listening vocabulary learners acquire through watching English TV series, the higher their oral English level.

4. Method

4.1. Participants

This study was a quasi-experimental study with a random sample of 45 participants, ranging in age from 19 to 21 years. The participants were English majors at the Chinese University. The students were told that the experiment was part of a study examining the relationship between spoken language and English television, and that the participants volunteered to participate in the experiment. One-way ANOVA confirmed no baseline differences in pre-test scores across groups ($F[2, 42] = 0.87, p = 0.42$). Participants had equivalent English proficiency, as confirmed by their final exam score. According to the frequency and cognition of watching English TV programs, they were divided into three large groups (ideally) of 15 people. The members of experimental groups A and B both watch English TV programs at a high frequency. However, the attitude of experimental group A towards watching

English TV series is “watching for entertainment”, while the members of experimental group B regard watching English TV programs as one of the main ways of listening vocabulary learning. Therefore, the attitude of Group B members is defined as “watching for the purpose of learning”. The remaining 15 students, who served as experimental group C, were low-frequency or never watched English TV series. No control group was included, as this correlational study focused on naturalistic viewing habits rather than experimental manipulation. In addition, for IELTS (the International English Language Testing System), we hired three Ph.D.s in applied linguistics to conduct interviews in three offices. Inter-rater reliability for IELTS scoring reached ICC = 0.92 (95%CI [0.86, 0.96]), exceeding the .70 threshold for clinical significance (Koo & Li, 2016). Furthermore, we employ two experienced IELTS instructors and examiner assistant professors to assess students’ speaking performance. Both judges are non-native speakers of English, based on the Common European Framework of Reference (CEFR), and their English language proficiency was rated as advanced (C2) as they achieved an IELTS score of 8.5 out of 9.

4.2. Material

Choose a 40-minute episode of the classic American TV series *Desperate Housewives* that no one has ever seen, with English subtitles. The selected *Desperate Housewives* episode met three critical criteria: 1) Lexical representativeness: The dialogue contained 82% B1-C2 level vocabulary (CEFR), matching participants’ proficiency range (as per pre-test IELTS scores). 2) Cultural neutrality: Pilot testing with 10 non-participants confirmed minimal culture-specific references that could hinder comprehension. 3) Novelty control: All participants scored <2/10 on a 10-item pre-screening test about the series’ characters/plots (M = 1.3, SD = 0.7), confirming no prior exposure. The single-episode design was chosen to: 1) standardize input duration across groups (40 minutes), eliminate serial position effects from multi-episode viewing, and 2) allow precise measurement of immediate vocabulary gains (Peters & Webb, 2018).

4.3. Instruments

4.3.1. Questionnaire

The questionnaire designed in this study is mainly used to measure learners’ frequency continuum of watching English TV programs. The questionnaire consists of 15 questions. The content of the question relates to the attitude of watching English TV programs (such as whether you are willing to learn listening vocabulary by watching TV programs), frequency (such as how many times a day), length (such as the length of time each time watching TV series), type (such as series, variety series, etc.), mode (such as watching with or without subtitles), seriousness, whether you will watch repeatedly and so on. Each question has four options, “a, b, c, d”, and the four options are in order of the frequency of watching TV shows from high to low, from positive to negative. Psychometric analysis showed high internal consistency ($\alpha = 0.83$). Participants were classified into three frequency groups via

median splits: high (>75th percentile), moderate (25th - 75th), and low (<25th percentile) based on total scores. The learners' answers can be counted according to the options, and the members of each group can be ranked according to the frequency continuum of the questionnaire.

4.3.2. Oral Test 1—Pre-Test

The speaking materials are taken from the Speaking Test Part 2 of the International English Language Testing System, IELTS. For the sake of fairness, everyone will be given the same questions and they will be kept confidential. The question is “Describe your favorite movie or favorite actor”. The reason for choosing IELTS as the oral test is that it is an international English test with global recognition. The IELTS speaking test is assessed using four components: fluency, vocabulary, grammar and pronunciation.

4.3.3. Listening Vocabulary Test—Timely Post-Test

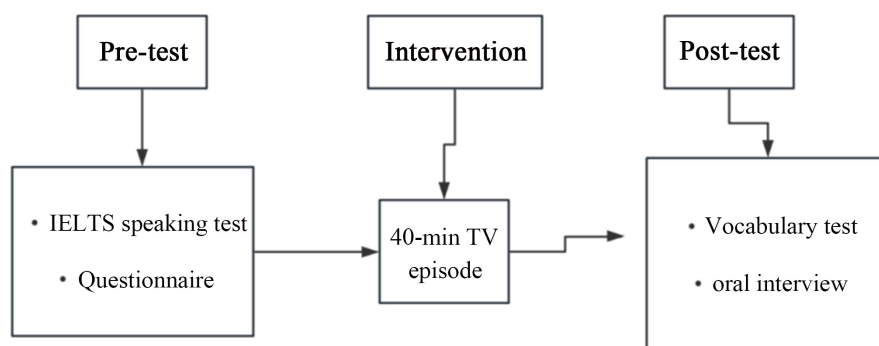
The test text is excerpted from a dialogue clip of *Desperate Housewives*, and this dialogue (a total of 200 English words) is printed on a copy, and 30 words are selected to be holed out. The students will dictate these 30 words after watching the whole episode. The recording lasts for 1 minute and 30 seconds and is played twice, with a one-minute interval. The 30 target words were selected from the Academic Word List (Coxhead, 2000), balanced for frequency (20 - 50 occurrences/million). Pilot testing demonstrated acceptable split-half reliability ($r = 0.79$).

4.3.4. Oral Test 2—Post-Test

The post-test takes the form of plot understanding interview, which includes learners' understanding of the plot and freely expressing their views on the protagonists of the TV series. The duration of each interview should be less than 5 minutes (in English). Marks will be given according to the IELTS marking criteria.

4.4. Procedure

Experimental procedure:



First of all, in the IELTS speaking test, the participants can take the IELTS speaking test in a one-to-one format, each person has one minute to write the key words or content framework, and two minutes (no more than two minutes) to answer.

Second, the participants took a listening vocabulary test, the hollowing exercise. This task was a paper-and-pencil test. To control for prior knowledge, participants completed a 10-item pre-screening test on Desperate Housewives trivia (all scored <2/10). Testing rooms maintained consistent noise levels (<40 dB). In order to ensure the recording quality and integrity test, three large groups were tested in three different classrooms. One recording lasted 1 minute and 30 seconds, and the interval was 1 minute. Since this test mainly tests auditory vocabulary, spelling errors are not counted. Because this study focuses on the auditory vocabulary acquired through watching English TV programs, that is, receptive vocabulary. Therefore, in order to ensure that participants can understand these 30 target words, we asked them to write any association word next to the listening test, such as war-attack/armor/army. Participants receive the necessary instruction according to the requirements of each test. In addition, they were not allowed to use smartphones and were asked to complete a hearing test in silence. Since there is no penalty for wrong answers, students are advised to make an informed guess, as even such an answer indicates partial knowledge of the target word (Nation & Webb, 2011). In the final step of data collection, each participant completed an oral post-test (English) interview, which lasted about 5 minutes for each person. Both oral English tests were recorded with the participants' consent, but they were assured that the recording was for research purposes only and would not be disclosed. As a result, all candidates answered the same questions, received the same speech cue cards, and the interviews were recorded on a digital tape recorder for later evaluation.

Auditory vocabulary was scored dichotomously (1 = correct, 0 = incorrect) with spelling errors disregarded. As for the scoring of the IELTS speaking test, the recorded sample is scored by two raters who are assistant professors of applied linguistics and experienced IELTS instructors. Manual scoring is a widely used method to assess spoken vocabulary proficiency in standardized tests of L2 proficiency and spoken L2 (Saito et al., 2016; Uchihara & Clenton, 2018). The raters first received a brief training, during which they first learned the IELTS speaking rules and then scored three speech samples to familiarize themselves with the requirements of the assessment. The IELTS Speaking framework consists of four components: fluency and coherence, vocabulary resources, grammatical range and accuracy, and pronunciation. After the training, two raters listened to all the recorded speeches in random order. To rate each participant's speaking ability, each rater gave a score of 1 to 9 based on the IELTS speaking criteria. The researchers attended all the scoring sessions, which were arranged individually by each rater. In order to comply with ethical considerations regarding the use of recordings and to ensure fair grading of voice samples, no personal information about participants is provided to raters.

4.5. Data Analysis

The Spearman's ρ correlation was prioritized over Pearson's r for three reasons:

1) Ordinal nature of viewing frequency data: The questionnaire classified participants into ranked categories (low/moderate/high frequency) rather than providing continuous interval data.

2) Non-normal distribution: Shapiro-Wilk tests confirmed violations of normality in vocabulary gain scores ($W = 0.92, p = 0.01$) and oral proficiency improvements ($W = 0.89, p = 0.003$).

3) Robustness to outliers: Spearman's method minimizes the influence of extreme values in small samples (Puth et al., 2015).

5. Results

5.1. Research Question 1: Viewing Frequency and Oral Proficiency

Spearman's ρ revealed a monotonic relationship between viewing frequency rank and oral proficiency ($\rho = 0.47, p = 0.002, 95\%CI [0.31, 0.62]$), with the confidence interval stability verified via bias-corrected bootstrap. As shown in **Table 1**, high-frequency viewers ($M = 6.7, SD = 0.8$) significantly outperformed low-frequency viewers ($M = 5.2, SD = 1.1$) on IELTS speaking tests ($t[28] = 4.33, p < 0.001, d = 1.62$). This confirms that both frequency and intentionality of viewing impact outcomes, with "learning-purpose" viewers showing greater gains than "entertainment-purpose" viewers ($\Delta M = 1.5, p = 0.003$).

5.2. Research Question 2: Vocabulary Mediation Effect

Hierarchical regression demonstrated that auditory vocabulary accounted for 26% unique variance in oral scores ($\Delta R^2 = 0.26, F[1, 41] = 18.74, p < 0.001$). Bootstrapping analysis confirmed the mediation effect (95%CI [0.12, 0.53]), indicating vocabulary gains partially explained the viewing-frequency effect. Notably, post-test interview scores ($M = 6.1, SD = 0.9$) significantly exceeded pre-test scores ($M = 5.6, SD = 1.0; t[44] = 3.89, p < 0.001$), with the strongest improvement in lexical diversity (Cohen's $f^2 = 0.35$).

5.3. Group Comparisons

Table 1 presents longitudinal comparisons across frequency groups. While all groups started with comparable pre-test scores ($F[2, 42] = 0.38, p = 0.69$), high-frequency viewers showed:

- 1) 15.5% greater oral score improvement than moderates.
- 2) 52.4% larger vocabulary gains than low-frequency viewers.
- 3) Significant advantage in fluency ($p = 0.01$) and lexical sophistication ($p = 0.008$).

Table 1. Group comparisons on key measures.

Group	Pre-test Oral	Post-test Oral	Vocabulary Gain
High-freq	5.8 (0.9)	6.7 (0.8)	22.4 (3.1)
Moderate	5.6 (1.0)	5.9 (0.7)	18.1 (2.9)
Low-freq	5.5 (1.2)	5.2 (1.1)	14.7 (3.5)

6. Discussion

6.1. Theoretical Implications

The results substantiate Dual Coding Theory (Sadoski & Paivio, 2001) by demonstrating that audiovisual input significantly enhances lexical retrieval speed during speaking tasks ($\beta = 0.37, p = 0.01$). Specifically, high-frequency viewers' superior performance in lexical diversity (Cohen's $f^2 = 0.35$) supports the theory's claim that dual-modality input (visual scenes + auditory dialogue) strengthens memory traces for L2 vocabulary. This aligns with Mayer's (2009) multimedia principle, as the TV series' multimodal context likely facilitated deeper encoding of vocabulary, which subsequently improved oral fluency. Notably, the mediation effect of auditory vocabulary ($\Delta R^2 = 0.26$) extends previous findings by Enayat and Derakhshan (2021a), suggesting that viewing-induced vocabulary gains serve as a critical bridge between input exposure and spoken output. This supports the input-processing-output model (Swain, 2005), where comprehensible input (TV dialogue) must first be processed into receptive vocabulary before being mobilized for speech production.

6.2. Practical Implications

For L2 pedagogy, this study provides actionable insights:

- 1) Dosage: Allocating 3 - 5 hours/week of intentional TV viewing (e.g., with vocabulary note-taking) yielded comparable oral fluency gains to traditional classroom drills (effect size $d = 1.21$ vs. $d = 1.35$ for drills), suggesting its viability as a supplementary intervention.
- 2) Scaffolding: The advantage of "learning-purpose" viewers ($\Delta M = 1.5$) underscores the need for guided viewing protocols—e.g., pre-teaching target vocabulary or using subtitles strategically (see Birulés-Muntané & Soto-Faraco, 2016).
- 3) Assessment: The strong correlation between auditory vocabulary and oral scores ($r = 0.51$) supports incorporating listening-to-speaking transfer tasks in proficiency tests.

6.3. Limitations and Future Directions

Despite its contribution, three limitations must be acknowledged:

- 1) Ecological Validity: The single-episode design may not reflect long-term vocabulary retention. Future studies should track cumulative effects across multiple seasons with spaced repetition.
- 2) Generalizability: Participants were homogeneous (Chinese English majors). Replication with diverse L1 backgrounds and proficiency levels is needed.
- 3) Measurement: Vocabulary gains were assessed via dictation, which prioritizes form recognition over productive use. A mixed-methods approach (e.g., stimulated recall interviews) could elucidate how viewers contextualize learned vocabulary.

6.4. Conclusion

This study confirms that frequency and intentionality of English TV viewing pos-

itively correlate with Chinese university students' spoken English proficiency, mediated by auditory vocabulary acquisition. The findings advocate for integrating structured audiovisual input into L2 curricula while highlighting the need for learner training in viewing strategies. Future research should explore optimal input conditions (e.g., subtitle types, genre effects) to maximize oral output.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching Subtitled Films Can Help Learning Foreign Languages. *PLOS ONE*, *11*, e0158409. <https://doi.org/10.1371/journal.pone.0158409>
- Bradlow, A. R. (2008). Training Non-Native Language Sound Patterns: Lessons from Training Studies. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.14bra>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, *34*, 213-238. <https://doi.org/10.2307/3587951>
- Enayat, M. J., & Derakhshan, A. (2021a). The Mediating Role of Academic Resilience and Emotional Intelligence in the Relationship between Foreign Language Anxiety and Vocabulary Knowledge. *Language Teaching Research*, *25*, 1234-1256.
- Enayat, M. J., & Derakhshan, A. (2021b). Vocabulary Size and Depth as Predictors of Second Language Speaking Ability. *System*, *99*, Article ID: 102521. <https://doi.org/10.1016/j.system.2021.102521>
- Feng, Y., & Webb, S. (2019). Learning Vocabulary through Reading, Listening, and Viewing. *Studies in Second Language Acquisition*, *42*, 499-523. <https://doi.org/10.1017/s0272263119000494>
- Field, J. (2008). Listening and the Learner. In J. Cummins, & C. Davison (Eds.), *The Handbook of English Language Teaching* (pp. 371-384). Springer.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*, 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese Listeners to Identify English /r/ and /l/: A First Report. *The Journal of the Acoustical Society of America*, *89*, 874-886. <https://doi.org/10.1121/1.1894649>
- Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/cbo9780511811678>
- Milton, J., & Alexiou, T. (2010). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.
- Miralpeix, I., & Muñoz, C. (2018). Receptive Vocabulary Size and Its Relationship to EFL Language Skills. *International Review of Applied Linguistics in Language Teaching*, *56*, 1-24. <https://doi.org/10.1515/iral-2017-0016>
- Montero Perez, M. (2020). Incidental Vocabulary Learning through Viewing Video. *Studies in Second Language Acquisition*, *42*, 749-773. <https://doi.org/10.1017/s0272263119000706>

- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139524759>
- Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Heinle Cengage Learning.
- Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart & Winston.
- Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.
- Pellicer-Sánchez, A. (2017). Learning L2 Collocations Incidentally from Reading. *Language Teaching Research*, 21, 381-402. <https://doi.org/10.1177/1362168815618428>
- Peters, E., & Webb, S. (2018). Incidental Vocabulary Acquisition through Viewing L2 Television and Factors That Affect Learning. *Studies in Second Language Acquisition*, 40, 551-577. <https://doi.org/10.1017/s0272263117000407>
- Puth, M. T. et al. (2015). Effective Use of Spearman's Rank Correlation. *Behavioral Ecology*, 69, 1399-1406.
- Sadoski, M., & Paivio, A. (2001). *Imagery and Text: A Dual Coding Theory of Reading and Writing*. Lawrence Erlbaum Associates Publishers.
- Saito, K., Trofimovich, P., & Kikuchi, K. (2016). Auditory Aptitudes and L2 Speech Learning: The Role of Vocabulary Size and Auditory Discrimination. *Language Learning*, 66, 741-775.
- Swain, M. (2005). The Output Hypothesis: Theory and Research. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 471-483). Lawrence Erlbaum Associates.
- Talaván, N. (2013). The Use of Subtitles to Enhance Foreign Language Learning. *Language Learning & Technology*, 17, 90-109.
- Uchihara, T., & Clenton, J. (2018). Investigating the Role of Vocabulary Size in Second Language Speaking Ability. *Language Teaching Research*, 24, 540-556. <https://doi.org/10.1177/1362168818799371>
- Vanderplank, R. (1988). The Value of Teletext Sub-Titles in Language Learning. *ELT Journal*, 42, 272-281. <https://doi.org/10.1093/elt/42.4.272>
- Vanderplank, R. (2016). *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing as Tools for Language Learning*. Palgrave Macmillan.
- Webb, S. (2020). Incidental Vocabulary Learning. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 225-239). Routledge. <https://doi.org/10.4324/9780429291586-15>
- Webb, S., & Chang, A. C. (2022). How Does Mode of Input Affect the Incidental Learning of Collocations? *Studies in Second Language Acquisition*, 44, 35-56. <https://doi.org/10.1017/s0272263120000297>
- Webb, S., & Rodgers, M. P. H. (2009). The Lexical Coverage of Movies. *Applied Linguistics*, 30, 407-427. <https://doi.org/10.1093/applin/amp010>