



Department of Education, University of Oxford

**A Systematic Review of Research on Listening Strategies
in Second Language Lecture Comprehension**

Atsushi Kanayama

Dissertation submitted to the University of Oxford in partial fulfillment of the degree

Master of Science in Applied Linguistics & Second Language Acquisition

Jesus College

Trinity Term 2020

Abstract

This study adopted a systematic research methodology to synthesize previous literature on listening strategies in second language lecture comprehension. After identifying 21 eligible studies, the review mapped out the current state of empirical research related to the topic. It also synthesized and assessed the quality of evidence pertaining to the three in-depth review questions:

1. What evidence is available on the deployment of listening strategies in second language lecture comprehension?
2. What evidence is available on the effectiveness of listening strategies in second language lecture comprehension?
3. What evidence is available on the development of listening strategies in second language lecture comprehension?

The results of this review revealed several research gaps in terms of research focus (e.g., participants, contexts) and methodology (e.g., study design, data collection, data analysis) and were discussed in relation to the above questions to make suggestions for future research.

Acknowledgement

Firstly, I would like to thank my supervisor, Heath Rose, for his great insights and patience during supervision. I also received significant help from Ziwei Huang; she kindly offered help for the screening of this review. Thank you, Rustem, for your friendship. You have become one of my best friends during the time at Oxford.

Natsuno, thank you for your encouragement while I studied here. I hope your DPhil thesis has been going well. Naosuke, you have treated me very well. Thank you so much for your kindness. Samuel, I appreciated your presentation on your systematic review and brilliant suggestions. Simon and Robert, I have truly appreciated your warm welcome and support. I sincerely appreciate Mr. Takahashi and others from the Nishinomiya Shukugawa Rotary club for giving me a chance to study for the current program.

I have been supported by many people outside Oxford. My endeavor in English education began when Mr. and Mrs. Nakamori decided to hire me as an English teacher at TRA Academy in Osaka. During the subsequent double degree program, I received academic supervision from Prof. Hiromi Ohtaka (Kwansei Gakuin University) and Dr. Anne Lawrie (the University of Stirling). Without their guidance, I

would not have been able to continue my study. Giacomo, thank you for supporting me during a difficult time. I would not have been able to survive this hardship without your psychological support. Emma, thank you for looking at my writing from time to time. I was very happy to attend Ms. Ishikawa's class at the University of Stirling last October. She has been a wonderful teacher. Pam and Dio, thank you for taking me to that café. I miss the time when we studied Japanese together. Dorothy and Archie, thank you so much for caring about me and asking me how I am from time to time. I felt very glad when I met Haydn, Rachel, and Lynn in Scotland and Eva in London again last year. Andy and Evelyn, I hope you have been well and look forward to seeing you soon. It was a huge encouragement for me. Alex, you have always been there for me during the difficult times. A Non, I miss smoking with you. I aim to stop smoking but I am willing to have some beer with you next time I see you. Nicholas, I have missed the strange helmet you always wore at a bar. Penelope, I was very glad to hear about your baby. Amanda and Esther, thank you for staying friends. I hope you have been well.

I have been blessed with great teachers throughout my academic life. Thank you so much for your heartwarming letters, Tabuchi Yoshie, who taught me when I was in elementary school. She is one of my oldest teachers; her love toward students has

had a huge impact on my attitude toward education. Prof. Ueda was my first academic supervisor when I studied business administration. I appreciated him for giving me a chance to pursue an academic career. I am also grateful for Prof. Kanai's patience and support during the study at Kobe University despite the two-year extension. I was very lucky to meet Prof. Umeda, Prof. Nagayama, and his wife, Michiko Nagayama, who have warmly watched over my growth for more than ten years. Prof. Hamada, I remember talking with you about my plan to contribute to children in the future without being clear about how I might actualize this idea. I never expected that I would end up studying child development next year. Prof. Nakagawa agreed to meet up with me and treated me delicious Japanese lunch before coming to Oxford. Thank you very much for your encouragement then. I would also like to appreciate Prof. Mizumoto for showing me a role model. I participated in his lecture before coming to Oxford, which gave me confidence in the knowledge of statistics. He also gave me advice when I had problems understanding some core concepts in statistics.

I have many friends in Japan that I would like to acknowledge here. Mr. Tanoue, I appreciated writing materials you gave me as a gift before coming to study at Oxford. Ma Liya, you have been one of my best friends since my undergraduate

years. I hope you and your family are doing well. Mr. Itani, thank you for your support and friendship during and even after my undergraduate study. Ken Hara, you have a beautiful heart and I have always admired that. Mr. Masui, I hope you and your family have been doing well. Mr. Tani, thank you for taking me around nice places and your warm encouragement. Ms. Nakai, thank you for trying to make time for me before studying abroad. I am sorry that we did not manage to meet, but I hope you and your son have been well. Ms. Mori, I hope you have been doing well in California. Ms. Sawamura, I have gratefully consumed all the Japanese food supply you gave me. Ms. Kitamura, thank you for helping me print out the poster for my poster presentation last year. I hope you have been able to accomplish your study successfully at the University of Stirling. Minoru, thank you for your friendship. We will visit our promised land together soon. Hinako, I was glad to hear that you became a teacher too. Ms. Morimoto, I hope you and your family have been safe and well. I hope you will come back and resume your study at your own timing. Yuyu, thank you for your warm friendship. Damion, you have been a great encouragement for me. Ms. Wada, I hope you will pass the national exam. Nate, I will never lose a pool game with you. Yuuya Okada, I believe you belong to academia; when are you coming back? Dr. Nishimura, I look forward to seeing you again in Japan. Ms. Abe, I hope your distant study has been proceeding well. Kaori

Ikeuchi, without you, I would not have been able to continue my study. I would also like to thank the members of our study group, LCCE, for sharing your time, efforts, and knowledge: Akie Fukushima, Kaoru Iwasa, Kijima Sho, Suguru Ishida, and Ryuichi Suzuki. I also have other friends that I met at the University of Stirling: Tatsuru Arita, Takuma Ohba, Sakura Arai, Akira Kunimoto, and Kotone Arai. Thank you for your friendship. I am also thankful for Ms. Sato, for giving me your late husband's suitcase and gloves. I really miss you all, but I will stay for at least another half-year at Oxford. Please stay well.

I would also like to acknowledge Mizuho and Kouji for being a very good sister and a very good brother to me. Kousei, you have given me the meaning of my life.

Finally, I would like to dedicate my work to my parents, Naomi Kanayama and Tadahiro Kanayama.

Table of contents

<i>Abstract.....</i>	<i>i</i>
<i>Acknowledgement</i>	<i>ii</i>
<i>Table of contents</i>	<i>vii</i>
<i>List of tables</i>	<i>xii</i>
<i>List of figures.....</i>	<i>xiv</i>
<i>List of abbreviations</i>	<i>xv</i>
<i>Chapter 1 Introduction</i>	<i>1</i>
1.1 Background to the study	1
1.2 Rationale for a focus of L2 listening in academic contexts	3
1.3 Objectives.....	5
1.4 Overview	6
<i>Chapter 2 Literature review</i>	<i>7</i>
2.1 Second language lecture comprehension.....	7
2.1.1 Characteristics of authentic lectures	8
2.1.2 Models of comprehension	10
2.2 Listening strategies.....	17
2.2.1 Definitional and measurement issues	18
2.2.2 Effectiveness of listening strategies	22
2.2.3 Development of listening strategies	25
2.3 Review questions	28
<i>Chapter 3 Methodology</i>	<i>30</i>

3.1 Principles.....	30
3.2 Protocol	33
3.3 Eligibility criteria.....	33
3.4 Information sources	38
3.5 Search	40
3.6 Study selection	42
3.6.1 Title and abstract screening	43
3.6.2 Full-text screening	45
3.7 Data collection process	48
3.8 Data items.....	51
3.9 Risk of bias in individual studies	53
3.9.1 Overall quality assessment	54
3.9.2 Methodological quality assessment	62
3.10 Risk of bias across studies	64
3.11 Synthesis of results	66
<i>Chapter 4 Results.....</i>	69
4.1 Included studies	69
4.2 Characteristics of included studies	71
4.3 Overall quality of included studies.....	79
4.3.1 Deployment of listening strategies in second language lecture comprehension..	79
4.3.2 Effectiveness of listening strategies in second language lecture comprehension	88
4.3.3 Development of listening strategies in second language lecture comprehension	95
4.4 Methodological quality of included studies	100
<i>Chapter 5 Discussion and conclusion</i>	107

5.1 Summary of evidence	107
5.1.1 Overarching review question: what empirical research has been carried out on listening strategies for second language lecture comprehension between 1975 and 2020?	108
5.1.2 First in-depth review question: what evidence is available for the <i>deployment</i> of listening strategies in second language lecture comprehension?	109
5.1.3 Second in-depth review question: what evidence is available for the effectiveness of listening strategies in second language lecture comprehension?	112
5.1.4 Third in-depth review question: what evidence is available for the development of listening strategies in second language lecture comprehension?	115
5.2 Limitations.....	116
5.3 Conclusion	118
References	121
Appendices	159
Appendix A: Review protocol.....	159
Appendix B: Timeframe of this review	166
Appendix C: Search strategies	167
Appendix D: R script for generating random numbers using uniform distribution	168
Appendix E: Excel sheet for full-text screening	169
Appendix F: Data items.....	170
Appendix G: Evaluation of the quality of quantitative and qualitative findings	171
Appendix H: Further details of studies by task characteristics of listening activity.....	183

Appendix I: Further details of studies by text characteristics of listening activity.....	184
Appendix J: Summary of the studies related to the first in-depth review question	185
Appendix K: Summary of the studies related to the second in-depth review question	190
Appendix L: Summary of the studies related to the third in-depth review question	192
Appendix M: Annotated bibliography of included studies	193
1. Bidabadi and Yamat (2013)	193
2. Cai and Lee (2012)	196
3. Cai and Lee (2010)	200
4. Carrier (2003)	203
5. Chand (2007)	205
6. Chou (2015)	208
7. Du (2018)	212
8. Flowerdew and Miller (1992)	215
9. Fung and Macaro (2019)	217
10. Hosogoshi (2016)	220
11. Jin and Xu (2017)	222
12. Lee and Cai (2010)	224
13. O'Bryan and Hegelheimer (2009)	227
14. O'Malley et al. (1989)	231
15. O'Malley, Chamot, Stewner-Manzanares et al. (1985)	233
16. Olsen and Huckin (1990)	235
17. Rahimirad and Moini (2015)	237
18. Rukthong and Brunfaut (2020)	239

19.	Schroeder (2016)	242
20.	Smidt and Hegelheimer (2004)	244
21.	Weinberg et al. (2011).....	246

List of tables

Table 1 Eligibility criteria	35
Table 2 Search terms.....	41
Table 3 Reasons for exclusion in title and abstract screening	45
Table 4 Clarification of criterion 3 and 4.....	46
Table 5 Reasons for exclusion in full-text screening	48
Table 6 WoE criteria and signaling questions.....	55
Table 7 Cross-tabulation: country by educational setting (n = 21)	72
Table 8 Studies by age (n = 21)	72
Table 9 Studies by first language (n = 21)	73
Table 10 Studies by second language (n = 21)	73
Table 11 Studies by second language proficiency (n = 21).....	74
Table 12 Studies by second language listening proficiency (n = 21)	74
Table 13 Studies by gender (n = 21)	75
Table 14 Studies by discipline (n = 21)	75
Table 15 Studies by strategy elicitation method (n = 26, not mutually exclusive).....	76
Table 16 Studies by task characteristics of listening activity (n = 21)	77
Table 17 Studies by text characteristics of listening activity (n = 21)	77
Table 18 Weight of evidence assessment for high overall quality studies in relation to the first in-depth review question.....	81
Table 19 Weight of evidence assessment for medium overall quality studies in relation to the first in-depth review question.....	83
Table 20 Weight of evidence assessment for low overall quality studies in relation to the first in-depth review question.....	86

Table 21 Weight of evidence assessment for medium overall quality studies in relation to the second in-depth review question.....	90
Table 22 Weight of evidence assessment for low overall quality studies in relation to the second in-depth review question.....	93
Table 23 Weight of evidence assessment for a high overall quality study in relation to the third in-depth review question	97
Table 24 Weight of evidence assessment for medium overall quality studies in relation to the third in-depth review question	98
Table 25 Weight of evidence assessment for low overall quality studies in relation to the third in-depth review question	100
Table 26 Methodological quality of qualitative studies.....	101
Table 27 Methodological quality of quantitative randomized controlled trials	103
Table 28 Methodological quality of quantitative non-randomized studies...	103
Table 29 Methodological quality of quantitative descriptive studies.....	104
Table 30 Methodological quality of mixed methods studies	106
Table 31 Four quantitative validity types and signaling questions	171
Table 32 Methods to test reliability	174
Table 33 Strategies to increase qualitative validity and reliability and signaling questions.....	176
Table 34 Issues concerning the validity and reliability of verbal protocols..	180

List of figures

Figure 1 Prisma flow diagram	70
Figure 2 Studies by year (n = 21)	71

List of abbreviations

ANOVA	Analysis of Variance
ASESS	the Academic Spoken English Strategies Survey
BEI	British Education Index
CEFR	the Common European Framework of Reference for Languages
CERQual	Confidence in the Evidence from Reviews of Qualitative research
DVD	Digital Video Disc
EEF	Education, Endowment, Foundation
EFL	English as a Foreign Language
EGAP	English for General Academic Purposes
ELT	English Language Teaching
EMI	English Medium Instruction
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ERIC	Education Resources Information Center
ESAP	English for Specific Academic Purposes
ESL	English as a Second Language
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HLK	High Linguistic Knowledge
HLP	High Listening Proficiency
IELTS	the International English Language Testing System
IRQ	In-depth Review Question
L1	First Language
L2	Second Language
LLBA	Language and Linguistics Behavioural Abstracts
LLK	Low Linguistic Knowledge
LLP	Low Listening Proficiency
MALQ	the Metacognitive Awareness Listening Questionnaire
MELAB	the Michigan English Language Assessment Battery
MMAT	the Mixed Methods Appraisal Tool

NNES	Non-Native-English-Speaking
OPT	the Oxford Placement Test
PRISMA	Preferred Reporting Items for Systematic Reviews
PRISMA-P	Preferred Reporting Items for Systematic review and Meta-Analysis Protocols
RCT	Randomized Controlled Trial
SCCI	Social Sciences Citation Index
TED	Technology, Entertainment, Design
TEFL	Teaching English as a Foreign Language
TESL	Teaching English as a Second Language
TESOL	Teachers of (or Teaching) English to Speakers of Other Languages
TOEFL iBT	the Test of English as a Foreign Language Internet-based Test
TOEFL PBT	the Test of English as a Foreign Language paper-based test
USA	the United States of America
WoE	Weight of Evidence

Chapter 1 Introduction

This chapter introduces the background to the study, its rationale, and the objectives of this study. In the background, a brief picture of the global context is presented to illustrate the importance of the current topic. Next, a rationale is provided for the current study's focus on L2 listening strategies in academic contexts. Then, the objectives of this study are outlined, before an overview of the dissertation is provided.

1.1 Background to the study

Teaching and learning academic content through a language other than a native tongue, or a second language (L2), dates back to at least two thousand years ago when children were educated in Greek and Latin to become successful in the ancient Roman world (V. A. Murphy, 2014). Since then, individuals have striven to acquire knowledge through an L2 due to a variety of personal, socio-cultural, economic, and political reasons. In the present day, some of the incentives may be driven by globalization and internationalization of business and commerce, the

advancement in telecommunication, the growth in voluntary immigration across nations, and the increased needs to preserve national and regional languages (Genesee, 2004). Educational institutions have played an important role in equipping students with the knowledge and skills to adapt to these global changes.

Recent technological development has led to the expansion of the resources, techniques, and modes of teaching and learning (Jones, 2007). In response, some scholars have questioned the value of lecturing to accommodate the diversified needs of learners (Biggs & Tang, 2011; DiPiro, 2009). Yet, others argue for its practical and pedagogical benefits (French & Kennedy, 2017). Despite the differences in viewpoints, didactic lectures are still pervasive in education (Zare & Keivanloo-Shahrestanaki, 2017) and remain “the central ritual” of the culture of teaching and learning (Benson, 1994, p. 181).

Another recent development that deserves attention concerns the medium of instruction. English has achieved the status of a global language (Crystal, 2012), rendering itself the most commonly used language in international academic and business communities. Accordingly, an increasing number of non-native-English-speaking (NNES) students are receiving lectures in English both in English-

speaking and non-English-speaking countries (Flowerdew, 1994; Macaro et al., 2018). Thus, listening to lectures, also referred to as academic listening (Richards, 1983), has become essential for the academic success of these students (Olsen & Huckin, 1990).

1.2 Rationale for a focus of L2 listening in academic contexts

Academic listening involves identifying the topic of a lecture and following topic development (Richards, 1983). It also has some unique characteristics, such as the requirement of a certain amount of subject knowledge, the ability to distinguish relevant information from irrelevant information, and unidirectionality with only a few opportunities for turn-taking (e.g., asking questions; Flowerdew, 1994; cf. Lynch, 2011). The distinctive features of a lecture may place special demands on L2 listeners (S. E. Thompson, 2003). Previous research has well documented the difficulties encountered during L2 lecture comprehension. For instance, NNES students may find it difficult to catch up with the speed of delivery or understand new vocabulary and concepts (Flowerdew & Miller, 1992; Graham, 2006; Hellekjær, 2017). Even with high L2 proficiency, learners may still lack the discoursal or cultural

knowledge to understand the conventions and cues that signal the main points made by a lecturer (Chaudron & Richards, 1986; Lebauer, 1984; Olsen & Huckin, 1990; Tauroza & Allison, 1994). Indeed, Bilbow (1989) once depicted the challenges faced by L2 listeners as being “entirely at the mercy of lecturers who may be oblivious to their problems” (p. 85). One way to approach the issue may be to investigate how lectures can be delivered to help L2 listeners understand the content better. However, this will only offer a partial solution to the problem. Even when lectures are well organized and presented, comprehension problems may still persist (Olsen & Huckin, 1990), which must be resolved on the part of L2 learners. Thus, L2 listeners’ strategy use, or conscious and purposeful deployment of mental actions, is another domain that seems worth exploring (Macaro, 2006; Oxford, 2017, 2018). The insights into learner strategies will help teachers and learners understand how to use and develop such strategies to facilitate comprehension in the face of difficulty in L2 lecture comprehension.

Listening remains the least understood and least researched L2 skill (Flowerdew & Miller, 2010; Morley, 1991; Santos et al., 2008; Siegel, 2015a; Vandergrift, 2007). Even less attention has been paid to academic listening (Lynch, 2011; Rowley-Jolivet, 2002). While recent years have witnessed a growing scholarly interest in L2

listening processes and strategies (Graham, 2017; cf. Plonsky, 2011; Vandergrift, 2007; Vanderplank, 2013), there is still a paucity of research specifically focusing on listening strategies in L2 lecture comprehension. For instance, in Macaro et al.'s (2005) systematic review of 84 studies on unidirectional listening, only two studies investigated listening strategies for academic lectures. Similarly, Macaro et al. (2007) evaluated the evidence from 44 empirical studies on listening strategies in depth, although the review did not focus on listening to lectures. While there have been reviews of academic listening (e.g., Lynch, 2011) or listening strategies in unidirectional listening (e.g., Graham, 2017) conducted recently, no systematic effort has been made to identify and evaluate the studies investigating listening strategies in L2 lecture comprehension.

1.3 Objectives

This study adopts a systematic research methodology to synthesize previous literature on listening strategies in L2 lecture comprehension. More specifically, the review aims to

- identify and map out the current state of empirical research related to listening strategies in L2 lecture comprehension
- synthesize the empirical evidence pertaining to the specific aspects of the research interests in this topic
- assess the quality of the included studies in terms of their contribution to this topic
- clarify where gaps exist in the literature and further research is necessary

The review questions are outlined following an initial review of the literature in Chapter 2.

1.4 Overview

This dissertation has four subsequent chapters. Chapter 2 contextualizes the review topic within wider literature on L2 lecture comprehension and listening strategies and presents review questions. Chapter 3 outlines the methodology of this review. The results are reported in Chapter 4 and discussed in Chapter 5.

Chapter 2 Literature review

This chapter situates the current topic in two fields of research and identifies the review questions. Firstly, research on L2 lecture comprehension is reviewed to illustrate how information in a lecture may be processed in the mind of L2 learners. Secondly, literature on L2 listening strategies is reviewed to highlight three key themes that are important to advance understanding within the field. Finally, based on this review of the literature, the review questions of the current study are stipulated.

2.1 Second language lecture comprehension

This section features important aspects in understanding L2 learners' mental processes in L2 lecture comprehension. It summarizes the text and task characteristics of authentic lectures. Then, various models of comprehension that previous L2 listening studies relied on are introduced.

2.1.1 Characteristics of authentic lectures

Lectures can vary in terms of whether the lecture content is read from notes or processed in real-time (Flowerdew, 1994; Hansen & Jensen, 1994). On this point, Flowerdew (1994) argues that lectures are most commonly delivered in an informal and conversational way with reference to notes or handouts. Lectures share many features of written academic discourse, such as the use of specialist vocabulary. Nevertheless, lectures contain a more restricted range of lexical items when delivered to non-native listeners (Flowerdew, 1993) and exhibit those syntactic features that are typically found in spoken language (Flowerdew, 1994). In comparing an authentic lecture and a listening text in textbooks, Flowerdew and Miller (1997) discovered that authentic lectures were organized by tone units (instead of complete clauses), body language, and discourse markers and maintained the features of informal speech, such as false starts, redundancies, and repetitions. In addition, interpersonal strategies to establish rapport with students comprised an important feature of authentic lectures. In an authentic lecture, discourse markers were effectively used to highlight the macro-structures of lectures, such as cohesion and coherence. In contrast, the lecture extracts used in textbooks were often too short for L2 listeners to make use of such macro-markers.

Littlemore (2001) also showed that misunderstandings of the main points of the lecture can be caused by the misinterpretations of metaphors, which were prevalent among overseas students who attended lectures at a British university.

More recently, Wingrove (2017) has investigated the vocabulary use, lexical density (i.e., the percentage of lexical items by the total number of words), and speech rate in a lecture discourse corpus and a TED talk corpus. It was found that lectures contained more academic vocabulary than TED talks. Compared to TED talks, lexical density and speech rate in lectures were more consistent and negatively correlated with each other. Assessing the spoken texts used in listening sections of standardized tests, such as the IELTS, the MELAB, and the TOEFL iBT, Wagner (2016) argues that text characteristics of scripted lecture extracts in those tests are qualitatively different from real-world spoken language. For these reasons, researchers have often recommended the use of authentic listening materials, such as a video recording of a real lecture for the teaching of academic listening, although the use of other listening materials is also found beneficial to some extent (e.g., Field, 2011; Richards, 2003; Wingrove, 2017).

2.1.2 Models of comprehension

Within a cognitive framework, language comprehension is viewed as a general construct that is shared across modalities (J. R. Anderson, 1983, 2000; Færch & Kasper, 1986; Wolf et al., 2019). Therefore, the following views on L2 lecture comprehension have been largely built upon research into first language (L1) reading and listening, as well as L2 reading comprehension (Flowerdew, 1994; Flowerdew & Miller, 2010).

Listening can be conceptualized as a bundle of sub-processes (Lynch & Mendelsohn, 2019). The distinction between bottom-up and top-down processes has been fundamental in understanding comprehension (Vandergrift, 2011).

Bottom-up processing puts together pieces of linguistic input in a linear way to build and constrain the meaning of an utterance. Listeners act as a “tape recorder” (K. Anderson & Lynch, 1988, p.9) to record and store information as accurately as possible. In contrast, top-down processing makes use of prior knowledge to interpret a message. Prior knowledge is also known as “schema”, which can be further divided into linguistic, rhetorical, and content schema (An, 2013; Carrell, 1983, 1984, 1987). Linguistic schemata include lexical and syntactic knowledge, whereas

rhetorical schemata concern the knowledge of structure and organization of discourse genres, such as academic lecture (Field, 2004; Lynch & Mendelsohn, 2019). Content schemata are topical knowledge based on personal and second-hand experience as well as the information from the earlier text (i.e., co-text).

Schema theory posits that listeners formulate, confirm, and modify their hypotheses about the speaker's intended meaning based on their schemata. Therefore, K.

Anderson and Lynch (1988) characterized listeners using top-down processing as an "active model builder" (p. 11; cf. Tsui & Fullilove, 1998). It has been argued that the two types of processing primarily apply to unidirectional listening (Joyce, 2019), which makes this distinction relevant to L2 lecture comprehension.

The complexity resides in how bottom-up and top-down processes can be empirically distinguished (Flowerdew, 1994). In this regard, Field (2004) insists that the distinction should be made in terms of directions (e.g., bottom-up processing from smaller to larger units vs. top-down processing from larger to smaller units), whereas researchers commonly associate the two types with levels of processing (see Anderson's model of comprehension below; Vandergrift, 2011; Vandergrift & Goh, 2012). Moreover, the two types of processing operate in harmony in fluent listening (Field, 2004; McClelland et al., 1986; Vandergrift, 2004), the perspective

famously taken up by the Interactive-Compensatory Model of Reading (Stanovich, 1980, 1984). The model asserts that the degree of dependence on either processing type is determined by “the listener’s confidence as to the reliability of each” (Field, 2004, p. 367). This is exemplified by Lund (1991), who argued that L2 listeners depended on top-down processing more than L2 readers because L2 listeners had difficulties in decoding oral input. The effective use of both processes also involves other factors, such as the purpose of listening, text and task characteristics, and listener characteristics, such as L2 proficiency (Field, 2004; Vandergrift, 2011; Vandergrift & Goh, 2012).

Early studies claimed that skilled listeners employ top-down processing more often, whereas less-skilled listeners rely more on bottom-up processing (e.g., Hildyard & Olson, 1982; Shohamy & Inbar, 1991). However, there is ample evidence showing that skilled listeners are successful in employing both types of processing (e.g., Buck, 1991; Goh, 2000; Tsui & Fullilove, 1998; Tyler & Warren, 1987; Vandergrift, 1998). Taguchi (2001), for example, showed that proficient L2 listeners used a wider range of strategies including both bottom-up and top-down strategies. In Vandergrift’s (1998) study of French learners, the successful application of bottom-up and top-down processing was contingent upon their proficiency. Novice listeners

were forced to apply top-down processing due to the lack of decoding skills, whereas more proficient listeners used both bottom-up and top-down processing effectively. Similarly, Goh (2000) reported that lower-proficiency listeners found bottom-up processing problematic, indicating the need to develop skills related to both types of processing. The challenges faced by L2 learners to acquire bottom-up skills, such as word segmentation have been well documented (Field, 2008; Goh, 2000; Hulstijn, 2003; Michael, 2018; Swan & Walter, 2017b, 2017a; Vandergrift, 2011).

J. R. Anderson's (1983, 2000) three-phase model of (first) language comprehension is perhaps one of the most frequently referenced models in the L2 listening literature. Other listening models, such as the four-phase model by Cutler and Clifton (1999) or the five-phase model by Field (2009, 2013, 2019) draw upon J. R. Anderson's model. J. R. Anderson (2000b) divided comprehension into perception, parsing, and utilization. In perception, listeners use bottom-up processing to recognize and decode incoming sounds into meaningful units. The decoded information is momentarily retained in echoic memory. During the parsing stage, the decoded speech is syntactically analyzed. Both bottom-up and top-down processes are involved to construct a literal meaning of what has been said. The third stage,

utilization, is primarily top-down processing, in which listeners make inferences to understand the speaker's message based on their prior knowledge. The newly produced propositions are stored in listeners' long-term memory, becoming part of their schemata. While it is assumed that the three stages are interconnected and recursive to some extent, the emphasis is placed upon comprehension taking place in the given order (J. R. Anderson, 2000). J. R. Anderson's model has been extensively applied by L2 listening researchers (e.g., Bacon, 1992b, 1992a; Goh, 2000; O'Malley et al., 1989). In the analysis of the verbal reports from English as a second language (ESL) learners, Goh (2000) was able to identify and categorize listening problems in various situations including listening to lectures experienced by high and low proficiency listeners according to J. R. Anderson's model. O'Malley et al. (1989) also utilized J. R. Anderson's model to understand ESL listeners' strategy use during lecture comprehension. This study is included in the systematic review and will be discussed in detail in Chapter 4.

Other models, such as the Structure Building Framework (Gernsbacher, 1997) or the Construction-Integration model (Kintsch, 1988, 1998), explain how comprehension succeeds or fails during the higher-level processing. Since there has been no indication of the empirical application of the Structure Building

Framework in L2 listening research, the Construction-Integration model is reviewed here. The model posits that comprehension begins with construction, which refers to the automatic activation of both relevant and irrelevant knowledge related to the content of the text. During the construction stage, the input is analyzed lexically and syntactically to create the mental representation of the surface structure of the text. Then, the propositions (or complete ideas) of the text are semantically analyzed to identify the microstructure (i.e., the interrelationships between propositions) and the macrostructure (i.e., the global topics and their interrelationships), whose mental representation is called textbase. In integration, textbase is integrated into prior knowledge to construct a situational model, which is a mental model of the situation described by the text. This is where inferencing occurs. How integration takes place depends on the amount of prior knowledge and the goals of processing the text. The three levels of mental representation, namely the surface structure, textbase, and situation model, have been used to understand learners' levels of comprehension in L1 and L2 reading research.

The empirical application of this model in L2 listening research is, however, scarce.

In Vandergrift (2003), think-aloud data collected from learners of French were analyzed to identify their use of listening strategies. The qualitative differences in the

use of inferencing and world elaboration strategies between more-skilled and less-skilled listeners were discussed in the light of the Construction-Integration model. More specifically, it was suggested that less-skilled listeners failed to suppress the activation of irrelevant information and could not reach a comprehensive mental representation of the text. Following the model, Roussel (2011) controlled text difficulty (to construct a situation model) by adjusting the saliency (i.e., position) of German compounds in a listening text. The study also revealed the relationship between listening strategies and the depth of listening comprehension, although strategies were not explicitly defined. The application of the model in L2 lecture comprehension research is found in Jeon (2007), in which the author operationalized two levels of understanding (i.e., text-based and situation model understandings) and investigated them in relation to content knowledge and L2 listening proficiency. The limitation of this study resides in the use of bridging inference questions, which prompt listeners to make inferences based on the information from the listening text (Kintsch, 1993), to measure situational model understanding. Nevertheless, the questions cannot tap into a deeper level of understanding that makes use of prior knowledge, which is highly relevant to L2 lecture comprehension (Jeon, 2007; Kintsch, 1998; Kintsch & van Dijk, 1978; Nassaji, 2002).

The above-mentioned comprehension models suggest that listening comprehension is “a highly complex problem-solving activity” (Byrnes, 1984, p. 318) rather than passive reception of speech code. While L1 listeners can do much of the comprehension processing (especially bottom-up processing) automatically, the effective coordination of these processes (e.g., bottom-up and top-down processes) often requires conscious efforts on the part of L2 listeners (Vandergrift & Goh, 2012). Furthermore, L2 listeners operate under the constraints of a lack of linguistic, rhetorical, and content knowledge (Brown, 1995; Call, 1985; Færch & Kasper, 1986; Goh, 2000). Considering the intensity of linguistic and cognitive demands during the L2 lecture comprehension, L2 listeners’ conscious use of strategies becomes of vital importance to compensate for comprehension problems or to more actively facilitate their comprehension (Goh, 2002a; Siegel, 2015b; Vandergrift, 2011).

2.2 Listening strategies

This section points to important theoretical and methodological issues in L2 listening strategy literature. The first subsection concerns the ambiguity in defining strategies

and how it has impacted on the identification of listening strategies. The second part discusses how researchers have attempted to determine the effectiveness of listening strategies. The third subsection looks at how previous studies explored the development of listening strategies.

2.2.1 Definitional and measurement issues

Research on learner strategies has proliferated since the influential studies on good language learners by Rubin (1975) and Stern (1975). Nevertheless, scholars have not yet come to terms with an agreed definition of strategies (Bialystok, 1983; Cohen, 2007, 2014; Dörnyei, 2005; Dörnyei & Skehan, 2003; Ellis, 1994; Macaro, 2006; Skehan, 1989). In Cohen (2007), a survey of experts did not reach a consensus as to the degree of consciousness or goal orientation of strategies, nor as to what strategies consist of. Recently, Oxford (2017) has updated her definition of strategies as thoughts and actions with reference to their complex, dynamic, teachable, and conscious nature based on the content analysis of 33 strategy definitions. While the definition may best represent our current understanding in the field (N. Thomas et al., 2019), the inclusion of overt behaviors in defining strategies

causes the “size-abstractness dilemma” (Stevick, 1990, p. 144), obscuring the size of and the relationship between individual strategies (Macaro, 2006; Sparks & Ganschow, 1993; Stevick, 1990). Instead, Macaro (2006) proposed to conceptualize strategies as purposeful mental actions. He argued that such a focus on the minimal unit of a cognitive operation serves theory development. Along the same lines, Briggs (2015) defined overt actions as “an opportunity for strategic behaviour to occur” (p. 299). Despite the disagreements on what encapsulates strategies, scholars’ views on strategies are commonly founded upon cognitive theory (Cohen, 2007).

From the cognitive perspective, listening strategies can be defined as “conscious plans to manage incoming speech” (Rost, 2002, p. 236). Such a view of listening strategies aligns with the dominant comprehension models, since they share an information processing paradigm (e.g., Graham, 2017). Listening strategies may need further clarification in comparison with other relevant concepts. Listening strategies are not equivalent to listening skills in that listening skills are ingrained competencies, whereas listening strategies compensate listening skills (Field, 1998). Likewise, listening strategies differ from teaching strategies (or techniques) in terms of the locus of control. The operation of the former is regulated by learners

themselves (Duffy, 2002; Horváthová, 2013), although it is possible to conceptualize the intermediate state of learner strategies regulated relatively by others, such as teachers (N. Thomas et al., 2019; N. Thomas & Rose, 2019). Moreover, the present review focuses on strategies for language use (i.e., to achieve comprehension in an L2 lecture situation) rather than strategies for language learning (Macaro, 2006).

Listening strategies have been elicited in a variety of ways including think-aloud protocols, retrospective protocols, interviews, questionnaires, and diaries (e.g., Goh, 1998, 2002a; Graham, 2003; Ma & Oxford, 2014; Macaro et al., 2007; O'Malley et al., 1989; Vandergrift, 2003b; Vandergrift et al., 2006). Most studies employed variants of a think-aloud protocol, although this method carries validity (whether verbalization accurately reflects thoughts) and reliability (whether data are collected in a consistent way) issues due to the complex cognitive activities involved in both listening and the think-aloud protocol (Field, 2011; Macaro et al., 2007; Russo et al., 1989). While verbal reports can describe the listening strategy use by a small number of participants in depth, questionnaires are suited for providing a snapshot of a comprehensive list of listening strategies with a much larger sample size. It has been argued that psychometrically validated questionnaires, such as the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006), provide

insights into an underlying trait of strategies (e.g., Dörnyei, 2005; Tseng et al., 2006) and enable the use of inferential statistics to generalize the findings. However, others have pointed out that such a focus on trait may undermine the original interest in a complex, dynamic state of strategic behaviors (Gao, 2007; Rose et al., 2018). Therefore, the triangulation of different data collection methods has been endorsed to increase methodological rigor and to arrive at a comprehensive understanding of listening strategies (Macaro et al., 2007).

Adopting an information processing perspective, researchers, such as O'Malley and Chamot (1990) and Oxford (1990), proposed classification schemes that grouped strategies into metacognitive, cognitive, and socio-affective categories.

Metacognitive strategies refer to the mental activities for regulating learning, whereas cognitive strategies are seen as the mental activities for processing the target language to accomplish a task (Vandergrift, 1997). Socio-affective strategies entail the activities involving interaction or affective control in learning (Vandergrift, 1997). Since the classification was provided, researchers have attempted to identify individual listening strategies. However, the lack of disagreement over what constitutes a strategy has led to different taxonomies of listening strategies (Santos et al., 2008). For instance, the strategy of "comprehension monitoring" included in

Vandergrift's (2003) taxonomy refers to “checking, verifying, or correcting one's understanding at the local level” (p.494). On the contrary, Goh (2002) differentiated the three kinds of mental operations under the strategy cluster of “comprehension monitoring”. Santos et al. (2008) also felt the need to create their own taxonomy grounded in data. In their study, “comprehension monitoring” was distinguished from “correcting one's understanding”; the latter was regarded as a different strategy type. Furthermore, the study identified different types of prediction according to the knowledge resources it made use of. Santos and her colleagues' approach highlights the significance of taking a grounded approach to producing a taxonomy of listening strategies that may differ depending on the listener, text, and task characteristics (Rubin, 1994).

2.2.2 Effectiveness of listening strategies

Listening strategies have been investigated in association with other individual differences, such as gender, learning style, motivation, self-efficacy, language proficiency, listening ability, and prior knowledge (e.g., Bacon, 1992b; Goh, 2002b; Graham, 2011; Vandergrift, 1997, 2003b, 2005). Of particular interest has been the relationship between listening strategies and successful listening comprehension.

Despite the initial focus on a frequency or a variety of strategies (e.g., J. M. Murphy, 1985; O'Malley, Chamot, Stewner-Manzanares, et al., 1985), subsequent research discovered that effective and less effective L2 listeners deploy similar strategies but in different ways (e.g., Goh, 2002a; Graham, 1997; Vann & Abraham, 1990).

Because of this, strategy researchers now believe that success in a specific task is associated with an effective orchestration or self-regulation of strategies in response to the task requirement (Ehrman et al., 2003; Macaro, 2006; Oxford, 2017; Oxford et al., 2004). In addition, previous research generally supports the notion that successful listeners use effective metacognitive strategies (Goh, 1997, 1998; Graham & Santos, 2015; Jurkovič, 2010; Osada, 2001; Rahimirad & Zare-ee, 2015; Rost & Ross, 1991; Vandergrift, 1997, 1998, 2003a), which is theorized to be a manifestation of metacognitive knowledge concerning learners' beliefs and knowledge about learning (Vandergrift et al., 2006).

Macaro et al. (2007), however, argue that the evidence for the effectiveness of listening strategies is weak due to the lack of consensus on the measurement of strategies as well as outcome measures of listening success. The authors draw our attention to the need to control for linguistic knowledge (i.e., vocabulary and grammatical knowledge) to attribute the difference in listening comprehension to

listening strategies. In one of the few studies which controlled for linguistic knowledge, Santos et al. (2008) found that L2 listeners with higher linguistic knowledge were able to employ metacognitive strategies and cognitive strategies involving prior knowledge more effectively than learners with lower linguistic knowledge. However, some of the learners with higher linguistic knowledge showed the strategy use that was similar to that of the learners with lower linguistic knowledge. The relationship between listening strategies and knowledge sources was further investigated in Graham et al. (2010), in which both linguistic knowledge and prior (or non-linguistic) knowledge were controlled for. The findings showed that the facilitative use of metacognitive strategies and prior knowledge requires sufficient linguistic knowledge. Nevertheless, higher linguistic knowledge did not guarantee the effective use of strategies. These studies suggest linguistic knowledge and prior knowledge need to be controlled for to investigate the effectiveness of listening strategies. Moreover, this research indicates the importance of developing listening strategies even when L2 listeners possess high linguistic and prior knowledge.

2.2.3 Development of listening strategies

One strand of research that informs the development of listening strategies relates to the studies investigating the effects of listening strategies. There is some evidence that suggests that effective listening strategy use can be taught and learned to foster L2 listening success (e.g., Graham & Macaro, 2008; O'Malley, Chamot, Stewner-Manzanares, et al., 1985; Seo, 2000; I. Thompson & Rubin, 1996; Vandergrift & Tafaghodtari, 2010; Yeldham & Gruba, 2014). The studies, such as I. Thompson and Rubin (1996), Kohler (2002), and Graham and Macaro (2008), for instance, compared learners with and without strategy instruction and reported significant gains in listening comprehension in the intervention group. In light of the findings, researchers have claimed that strategy instruction should 1) be based on needs analysis, 2) develop metacognition, 3) provide opportunities to practice both top-down and bottom-up strategies in a balanced fashion, and 4) be motivating and clear about learning outcomes (Graham & Macaro, 2008; Yeldham & Gruba, 2014). However, study design issues were pervasive among these studies, such as the lack of a pretest or a delayed posttest and a lack of clarity about participant assignment procedures (cf. Graham & Macaro, 2008). It has also been argued that the mixed results produced by relatively few studies (Macaro et al., 2007; Plonsky,

2011) and little consensus concerning the principle of listening strategy instruction (Graham & Macaro, 2008) make it difficult to make generalizations or meaningful comparisons across studies. Additionally, listening strategy development has been implied (indirectly) by the increase in outcome measures related to listening success in many of these studies. Only a few studies collected data on listening strategies at multiple points in time to observe the development of listening strategy use (e.g., Seo, 2000; Vandergrift & Tafaghodtari, 2010; Yeldham & Gruba, 2014). The inclusion of data on listening strategies is especially important to understand what and how strategies can be developed by the provision of strategy instruction.

Other studies investigating the effectiveness of strategy instruction were mixed-method studies, relying largely on qualitative data analyses (e.g., Chen, 2009; Mareschal, 2007). Mareschal (2007) explored five low-proficiency and three high-proficiency learners of French who received strategy instruction for eight weeks. Using various data collection methods (e.g., think-aloud protocols, stimulated recall, questionnaires, and diaries), it was found that the learners increasingly used both top-down (e.g., inferencing) and bottom-up (e.g., identifying keywords) cognitive strategies more effectively. She also found that less-proficient learners responded to the intervention more favorably than proficient learners. In Chen (2009), 31

university students in Taiwan developed their use of listening strategies throughout 14-week strategy instruction. The data collected through reflective journals showed that the learners increased metacognitive strategies and top-down strategies (e.g., listening for gist, elaboration, prediction, and visualization), whereas the use of bottom-up strategies (e.g., understanding each word, translation, fixation) reduced during the course.

Finally, Graham et al. (2008, 2011) examined the change in strategy use in a naturalistic setting. Graham et al. (2008) tracked two lower-intermediate learners of French at secondary schools in England for six months, collecting verbal reports during a multiple-choice listening task at two time points (October and April). It was found that their use of listening strategies did not change. Graham et al. (2011) improved the study design by increasing the sample size (from 2 to 15) and by using multiple data collection methods (a recall protocol, verbal reports, and teacher interviews). Despite the subtle change in listening proficiency and listening strategy use, it was observed that the learners who improved listening proficiency increased the use of selective attention and monitoring. These studies show that if conducted properly, qualitative methods offer an effective way to gain rich insights into the longitudinal development of listening strategies. However, care must be taken to

increase the trustworthiness of the findings. Overall, little research exists in this area; evidence for listening strategy development has been mixed and inconclusive (Yeldham, 2019).

2.3 Review questions

To highlight knowledge gaps in existing research on listening strategies in L2 lecture comprehension, the present systematic review attempts to map out the current state of an evidence base with the overarching question:

What empirical research has been carried out on listening strategies for second language lecture comprehension between 1975 and 2020?

In addition, an in-depth analysis of the available evidence is conducted to further our understanding of the field. Three in-depth review questions (IRQs) guided the in-depth review:

1. What evidence is available on the deployment of listening strategies in
second language lecture comprehension?
2. What evidence is available on the effectiveness of listening strategies in
second language lecture comprehension?
3. What evidence is available on the development of listening strategies in
second language lecture comprehension?

Chapter 3 Methodology

This chapter details the methodology in conducting the current systematic review. It first explains the principles adhered to by this review, followed by a description of the protocol. The review began with setting up eligibility criteria to identify studies relevant to the review questions. Next, information sources were searched for potentially relevant studies. Then, the identified studies were screened for eligibility. Once the studies included in this review were determined, relevant data (i.e., any information about a study, such as details of authors, publications, methods, and results/findings; Li et al., 2019) were extracted from those studies using a data extraction table. Two critical appraisal tools were employed to assess risk of bias in and across included studies. Finally, data were synthesized using a narrative approach.

3.1 Principles

A systematic review refers to “a review of existing research using explicit, accountable rigorous research methods” (Gough et al., 2012, p. 1), which is a

secondary level of analysis of the evidence from primary research (i.e., studies that produce original data; Gough et al., 2012). It helps us plan and interpret new primary research by clarifying “what is known, how it is known, how this varies across studies, and...what is not known from previous research” (Gough et al., 2012, p. 3). Macaro et al. (2018) argue that a systematic review entails the following five defining characteristics:

1. being transparent about the procedures from conception to final conclusions
by using a protocol which specifies how the review is to be conducted;
2. including studies based on comprehensive and reliable searching;
3. minimizing reviewer bias;
4. synthesizing evidence and its reliability;
5. ensuring the relevance and accessibility of the research

As manifold as primary research may be in terms of its aims, methodologies, and types of knowledge they produce, systematic reviews also diverge in the method of synthesizing research depending on the nature of the purpose, review questions, and evidence included in their reviews. Nevertheless, elucidating the purpose and method of a review lays the foundation for a systematic assessment of the quality

and relevance of existing evidence (Gough, 2007) by making explicit bias (i.e., a systematic error or deviation from the truth; Boutron et al., 2019) of individual studies as well as the review itself. Therefore, attempts were made to detail judgments and the reasons behind them in the course of conducting this systematic review as explicitly as possible in the following sections to come.

The current systematic review is primarily based on the recommendations by the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2019b) because of the inclusiveness of the protocols that are useful in guiding systematic reviews in applied linguistics. The method undertaken in this review is reported according to the PRISMA statement (Moher et al., 2009) to ensure the reporting quality except for the section concerning summary measures (e.g., risk ratio, the difference in means) and risk of bias assessment. Summary measures were not considered as a matter of critical importance because the review intends to configure (i.e., interpret) a heterogeneous set of data to explore the issues surrounding the review questions rather than test a hypothesis or theory, which can be achieved by a meta-analysis of a homogeneous set of studies employing similar designs and methods (Gough & Thomas, 2012; Snijlsteit et al., 2012).

3.2 Protocol

A protocol document was created following the PRISMA-P (Moher et al., 2015), which detailed administrative information (e.g., title, registration, authors, amendments, and support), introduction (i.e., rationale and objectives), and planned methods (i.e., eligibility criteria, information sources, search strategy, study records, data items, quality assessment of individual studies, data, and synthesis). The elements in the protocol were developed and revised in parallel as the review progressed. This allowed the monitored execution of the review protocol and modification of the method when necessary (see Appendix A for the protocol). I began working on the rationale and objectives on 6th November 2019. The protocol was finalized and verified by a content expert on 17th June 2020, before a collaborator was recruited (see Appendix B for the timeframe).

3.3 Eligibility criteria

Eligibility criteria were established, against which each study was assessed for its relevance to the review questions (see Table 1). Eligibility criteria were formulated

and refined in relation to the revision to the review questions and search method

from 6th November 2019 to 5th March 2020.

Table 1 Eligibility criteria

No.	Category	Inclusion criterion	Exclusion criterion	Rationale
1	Publication	The studies were published in peer-reviewed English language journals.	The studies were not published in peer-reviewed English language journals.	Publication bias is one of the reporting (or dissemination) biases (Boutron et al., 2019) and takes effect when there are systematic differences between the results of published and unpublished studies (Song et al., 2013). While publication bias has been problematized in social sciences (Franco et al., 2014), limiting the search to peer-reviewed journal articles has its own strengths. Firstly, journal articles are accessible and informative about methods and results/findings, allowing efficient data extraction (Li et al., 2019). Furthermore, published articles are systematically collected by publicly available electronic databases, enabling the replication of the search. On the other hand, access to dissertations, book chapters, and conference papers largely varies depending on individuals. Peer-reviewed journals also ensure the quality of published studies through comprehensive evaluation by experts in the same scholarly area, exemplifying the best research practices in the field. With these issues and the exploratory nature of the review questions in mind, the systematic assessment of the evidence from peer-reviewed journal articles was considered to serve the review's purpose. Similarly, English language bias occurs when a systematic review does

not take into account studies written in languages other than English (Egger et al., 1997). However, the assumed systematic difference between English language and non-English journals has not been consistently observed in previous systematic reviews (Boutron et al., 2019). As with other fields of research, English is the most widely used language in published peer-reviewed journals in linguistics, applied linguistics, and education research. While the asserted systematic differences may exist in these fields, English is the language that was accessible by the reviewers as well as many others who make use of this review. Rather than increasing the risk of misinterpretation of the studies written in non-English languages, English language journals were exclusively searched to maintain the quality, objectivity, and replicability of this review.				
2	Year	The studies were published between 1975 and 2020.	The studies were published before 1975.	The studies published after the pioneering work by Rubin (1975) and Stern (1975) were searched to delimit the scope of this review to language learning strategy research. This is because theories of strategies for language learning and use have been built upon Rubin and Stern's exploration of how good language learners actually learn (Macaro, 2006).
3	Topic	The studies provided empirical	The studies did not provide empirical	The studies that provided empirical evidence on listening strategies employed by L2 learners were screened for. This excluded review or

		evidence on listening strategies employed by L2 learners.	evidence (e.g., review or conceptual studies) or only addressed native speakers.	conceptual studies as well as the studies of listening (or any other) strategies employed by native speakers.
4	Task type	The spoken text was (hypothesized to be) unidirectional and academic, regardless of the location and timing of the actual listening. The actual listening can take place in and outside the classroom as well as during and outside the data-collection period.	The spoken text was not (hypothesized to be) unidirectional and academic.	The studies that included unidirectional and academic spoken texts were screened for to focus on listening strategies for lecture comprehension. Although the evidence on listening strategies deployed in actual lectures would be ideal, a pilot search located only a few such studies. Moreover, it was felt that the inclusion of other listening tasks, such as listening to audio recordings or lecture videos, often employed in English for Academic Purposes classes, may provide greater insights into the current state of evidence available on this topic. To expand the scope of this review, I focused on two key features of L2 lecture comprehension: unidirectional (i.e., one way) and academic spoken text (i.e., spoken English used in lectures and seminars; Flowerdew, 1994). Accordingly, the studies that included (or assumed) unidirectional and academic spoken text were screened for.

Participant characteristics, such as gender, educational level, proficiency, L1, and L2, were not restricted in the eligibility criteria because the review was interested in the current state of evidence in relation to those factors. While it was anticipated that most included studies would provide insights into listening strategies in L2 lecture comprehension through the medium of English, L2 was not restricted to English to avoid the initial exclusion of empirical studies on strategies in other languages (e.g., French listening comprehension; Graham et al., 2010). Moreover, although the review topic does not concern learners with disabilities, the trial search with the above eligibility criteria indicated no need for this kind of specification. Similarly, the review did not restrict context because no systematic differences were presupposed across different countries or educational contexts.

3.4 Information sources

Information sources include all the databases, journals, citations, contact with study authors that were used as sources of evidence (Moher et al., 2009). The relevant studies in linguistics and education research were searched in the electronic databases, such as Web of Science (Social Sciences Citation Index (SCCI) (1956-

present), British Education Index (BEI), Education Resources Information Center (ERIC), Scopus, Language and Linguistics Behavioural Abstracts (LLBA), and PsycINFO on 5th March 2020.

To complement the electronic search, I performed hand searching in Google Scholar, major journals (i.e., System, TESOL Quarterly, The Modern Language Journal, Language Teaching, Applied Linguistics, ELT Journal, Studies in Second Language Acquisition), forward citations (i.e., the studies that cited retrieved articles accessed via Scopus, Web of Science, and Google scholar) and backward citations (i.e., the reference lists of retrieved articles; see Hollands et al., 2016 for the example use of forward and backward citation searches) from 6th March 2020 to 10th August 2020.

It should be noted that the systematicity of this review derives from the replicability of the results obtained from electronic databases with the use of a predetermined string of keywords. Therefore, hand searching was regarded as an auxiliary. Most of the data for this review come from a sample of articles obtained through electronic databases.

3.5 Search

Searching can be accomplished by many methods (i.e., search strategies) to identify evidence of relevance to review questions (Dundar & Fleeman, 2019).

Following the recommendation by Lefebvre et al. (2019), search strategies were discussed with experienced librarians and experts at the Department of Education at the University of Oxford, trialed in the databases, and refined so that they enabled the identification of as many relevant studies as possible (i.e., specificity) without including irrelevant ones (i.e., sensitivity; Boland et al., 2017). As with review questions and eligibility criteria, tentative search strategies were piloted and revised 14 times between 20th January 2020 and 5th March 2020. The finalized search strategies were later (i.e., 10th March 2020) verified by a content expert (see Appendix C for the search strategies).

The final version of search terms was organized according to four key themes concerning “listening” “strategies” for “content learning” in “second language”, which reflected the underlying nature of evidence of interest (see Table 2 for the search terms). This helped increase relevance and decrease redundancy in search terms.

The terms related to “listening”, “strategies”, and “content learning” were searched in

titles and abstracts to identify topically relevant articles, whereas the terms pertaining to “second language” were searched anywhere in the document to specify a research area.

Table 2 Search terms

Skill title and abstract	Strategy title and abstract	Content learning title and abstract	Second language anywhere in the document
"listen"	AND "strateg**"	AND "academi**" "bilingual education" "CBI" "classroom**" "CLIL" "comprehen**" "content and foreign language integrated" "content and language integrated" "content based**" "content learning" "content-and-language integrated" "content-based**" "EAP" "EMI" "English as a lingua franca" "English for specific purposes" "English medium" "English taught program**" "ESP" "immersion" "lecture**" "medium education" "medium for instruction" "medium instruction" "medium of English" "medium of instruction" "parallel language education" "parallel-language education" "understand**"	AND "additional language" "EAL" "EFL" "ESL" "foreign language" "L2" "second language"

The database search was conducted on 5th March 2020 and produced the initial sample of 1216 articles. The initial sample of 1216 articles was exported to Mendeley, in which duplicates and 8 non-articles (e.g., irrelevant synopses or other nonsense entries) were removed. After the data were exported to Rayyan (Ouzzani

et al., 2016), two duplicates were found and removed. In this way, the electronic search identified 680 articles. Furthermore, the hand search of other information sources (i.e., Google Scholar, major journals, forward citations, and backward citations) was carried out between 6th March 2020 and 10th August 2020. The auxiliary search found two additional eligible articles, indicating the comprehensiveness and reliability of the database search. In total, 682 articles were subjected to study selection.

3.6 Study selection

Study selection is undertaken to select the studies that contain evidence relevant to review questions (Fleeman & Dundar, 2017). It follows two stages: title and abstract screening and full-text screening (Higgins, Thomas, et al., 2019). Title and abstract screening is a time-saving way to exclude apparently irrelevant articles. Full-text screening further excludes irrelevant articles based on more detailed screening of individual studies.

A second reviewer, who was studying for a master's degree program in applied linguistics at a university in the UK at the time when the review was conducted, was recruited to increase the reliability of study selection. While she had sufficient knowledge in applied linguistics and familiarity with the relevant research methodologies to be able to make valid decisions during study selection, she had not focused on studying the current review's topic extensively before. This was important because the inclusion of a collaborator with pre-formed opinions may lead to bias in their judgments about the quality of studies (Higgins, Thomas, et al., 2019). Before participating in the review, she was informed of the purpose of the review, read a review protocol fully, practiced the procedures with examples, and was given opportunities to ask questions about the present study.

3.6.1 Title and abstract screening

The title and abstract of each article were screened in Rayyan for its eligibility in the order of criteria 1 to 4. This stage carries a risk of excluding relevant articles, as titles and abstracts may not bear enough information. Therefore, the screening was over-inclusive (Higgins, Thomas, et al., 2019) in that if the title and abstract did not

include enough information to determine the eligibility, the article was included. I (the first reviewer) conducted the first screening of 682 articles (680 electronically searched articles screened from 6th March 2020 to 15th March 2020; two hand searched articles screened each time they were found), resulting in 180 potential includes.

Before the second screening, any disagreements on the screening of a few example studies were discussed and resolved to establish a common understanding of eligibility criteria between the first and the second reviewer. Then, the second reviewer screened 10 % of the 680 articles (i.e., 68 articles) from 1st July 2020 to 7th July 2020. The articles to be screened were preselected based on the numbers randomly generated from a uniform distribution (where every number has an equal probability of occurring) using R (see Appendix D for the R script).

The inter-rater agreement was 97 percent, yielding a Cohen's kappa of .92 (almost perfect; McHugh, 2012). The reasons for exclusion and the corresponding number of articles are presented in Table 3. The result of the first screening (i.e., 180 articles) was retained for full-text screening.

Table 3 Reasons for exclusion in title and abstract screening

Exclusion criteria	Number of articles
Exclusion criterion 1: the studies were not published in peer-reviewed English language journals.	2
Exclusion criterion 2: the studies were published before 1975.	1
Exclusion criterion 3: the studies did not provide empirical evidence (e.g., review or conceptual studies) or only addressed native speakers.	459
Exclusion criterion 4: the spoken text was not (hypothesized to be) unidirectional and academic.	40
Total	502

3.6.2 Full-text screening

Out of the 180 articles from the first screening, the full texts of 176 articles were retrieved through online databases, whereas two articles were made available by the authors upon request. Despite the attempt, two articles were unprocurable. As Li et al. (2019) state, multiple studies can be included in one report (i.e., article), whereas there are cases where one study is described in multiple reports. Therefore, report ID and study ID were differentiated; 184 studies were reported in 178 articles, whose full texts were obtained. The studies did not contain errata (i.e., critical weaknesses and retractions) or multiple reports of the same study (Li et al., 2019). Therefore, the full texts of the obtained 184 studies were assessed for eligibility.

An excel spreadsheet was used to organize a list of articles and the eligibility criteria, against which each article was assessed (see Appendix E for an example).

At this stage, the descriptions of criteria 3 and 4 were too broad to discriminate studies. Thus, these criteria were further clarified; each study was given a score of 1 or 2 in reference to the revised criteria 3 and 4 (see Table 4). Initially, the studies that received the summed scores of 3 and above were included in data collection.

However, it was later (while collecting data) found that criterion 4 was more essential. Whereas studies that did not collect empirical data on listening strategies may still be informative (e.g., intervention studies), the clear indication of the inclusion of academic spoken text was crucial in ensuring the relevance of included studies. Therefore, it was decided that the data be collected from only the studies that received a score of 1 or 2 for criterion 3 and a score of 2 for criterion 4.

Table 4 Clarification of criterion 3 and 4

No.	Clarification
3	Score 2 is given if the study collected data on listening strategies. Score 1 is given if the study indicated (the use or the development of) listening strategies without collecting data on listening strategies (e.g., intervention studies).
4	Score 2 is given if the study explicitly and mainly investigates listening strategies in response to the unidirectional and academic spoken text (i.e., listening to a lecturer). Instruments and listening materials usually include

relevant information. It has been reported that TED talks (Wingrove, 2017) and scripted lecture extracts in standardized tests (Wagner, 2016) do not resemble lecture discourse. For this review, the studies that explore TED talks are excluded because TED talks are not lectures. In contrast, the studies that include listening to audio/video recordings of lectures in standardized tests, such as the TOEFL or the IELTS, were regarded as relevant because such listening tasks directly related to lectures. Score 1 is given if the nature of the spoken text is unclear, although it can be inferred from the study design that (at least some) listening strategies are used to understand the unidirectional and academic spoken text. The study aim, research questions, study design, method, analysis, results, and discussion indicate that the study is (at least partly) interested in listening strategies for learning academic content. The study explores listening strategies in relation to academic-related variables (e.g., academic motivation).

I performed the first full-text screening from 9th May 2020 to 30th June 2020. This procedure excluded 163 out of 184 studies, resulting in the final sample of 21 studies. To prepare for the second screening, the second reviewer practiced full-text screening a few times, while she discussed and resolved any disagreements with the first reviewer. As with title and abstract screening, the studies subjected to second screening were chosen based on randomly generated numbers (see Appendix D for the R script). Then, screening of 10 % of 184 studies (i.e., 18 articles) was conducted by the second reviewer from 7th July 2020 to 14th July 2020.

Cohen's kappa was $k = .100$, showing 100% agreement between the first and second reviewers. Thus, the resulting 21 studies were included in this review and retained for data collection. The reasons for exclusion and the corresponding number of articles are presented in Table 5.

Table 5 Reasons for exclusion in full-text screening

Exclusion criteria	Number of articles
Exclusion criterion 1: the studies were not published in peer-reviewed English language journals.	2
Exclusion criterion 2: the studies were published before 1975.	0
Exclusion criterion 3: the studies did not provide empirical evidence (e.g., review or conceptual studies) or only addressed native speakers.	25
Exclusion criterion 4: the spoken text was not (hypothesized to be) unidirectional and academic.	136
Total	163

3.7 Data collection process

Data can be defined as “any information about (or derived from) a study, including details of methods, participants, setting, context, interventions, outcomes, results, publications, and investigators” (Li et al., 2019, para. 21). Data collection, on the other hand, involves the extraction of relevant data from data sources (Fleeman & Dundar, 2017). In the case of this review, 21 studies, which met the eligibility criteria through study selection, formed the data sources.

Data can be collected and stored by a data extraction tool, the use of which is encouraged to increase the consistency and transparency in the data extraction process among reviewers and minimize the necessity to refer to data sources during data synthesis and risk of bias assessment (Li et al., 2019). There are two commonly used data extraction tools. A data extraction form enables the extraction and recording of data from a single study, whereas a data extraction table is used to extract, record, and organize data from multiple studies at once.

The data extraction table by Macaro et al. (2018) and Rose et al. (2018) provided a blueprint for the data extraction tool used for this review. An electronic tabular format was chosen to allow consistent data collection and management in one place. The use of an electronic form was advantageous because it can be stored, edited, analyzed, and shared electronically, which helped reduce errors and increased accessibility during the data collection (Li et al., 2019). Some data items and question wordings were amended by referring to a well-established electronic data extraction form (i.e., EEF main data extraction v 1.0 October 2019[Standard]; EPPI-Centre, 2019), the Cochrane Handbook (Li et al., 2019), the PRISMA statement (Moher et al., 2009), textbooks (e.g., Boland et al., 2017), and previous systematic reviews (e.g., Macaro et al., 2005).

The table was piloted on the eligible studies and was further revised to gather evidence that is particularly important for the current review questions and accommodate variation in different research designs and reports. Moreover, some evaluative questions and the space for additional comments were created to facilitate data synthesis and risk of bias assessment at the latter stages. The piloting of the data extraction table took place from 12th May 2020 to 17th June 2020, on which date the final version of the table was verified for its coverage and usability by a content expert. The iterative process of revisions led to the development of an easy-to-use data extraction table.

The author extracted data from the 21 studies from 12th May 2020 until 20th July 2020. Although it was desired that at least two reviewers independently extract data from all the included studies to minimize errors and reviewer biases (Li et al., 2019), this was not accomplished due to the exhaustive nature of the task. Alternatively, it is suggested that information that requires subjective evaluation in data collection be extracted from individual studies by at least two reviewers (Li et al., 2019). This was not achieved either, which is one of the limitations of this review.

3.8 Data items

Before data collection, it is essential to decide clearly what data are relevant for the purpose of this review and how to collect them using a data extraction tool (Fleeman & Dundar, 2017; Li et al., 2019). In doing so, it is important to consider the exploratory nature of the current review questions. The review questions set out to evaluate the quantity (e.g., how many studies...?) and quality (i.e., strength and weakness) of the relevant evidence without dictating the characteristics of the included studies so that the review can produce a clearer picture of how the issues have been addressed in the field of applied linguistics so far. Therefore, data items should reflect the data categories that were shared among studies that employ different research designs.

In the data extraction table for this review, data items were arranged in the order they were commonly presented in study reports (Li et al., 2019) from descriptive (i.e., study information, abstract, introduction/rationale, literature review, and methodology) to analytical (i.e., results/findings, discussion, and conclusion) data categories (Fleeman & Dundar, 2017). This was intended to increase the convenience of the data collection tool. Each category included data items to be

elicited by questions (see Appendix F for the data items). To minimize subjective interpretation and judgment, it was attempted to ask closed-ended questions about what is reported in each study (e.g., “Did the study report...?”) to elicit data for as many data items as possible. Care was also taken so that the responses were not left blank by including “Yes”, “Partly”, “No”, and “N/A” options (Li et al., 2019). In addition, some questions were evaluative (e.g., “What was...?” or “Was the study...?”) and created to facilitate a risk of bias assessment.

As the data extraction table was designed to accommodate heterogeneous sets of studies, it did not collect some data included in other commonly used data extraction tools. For example, many data items were allocated to address information on intervention and experimental designs in EEF main data extraction v 1.0 October 2019[Standard] (EPPI-Centre, 2019) and Data collection form for intervention reviews for RCTs and non-RCTs Version 3 April 2014 (Cochrane, 2014). However, the table used in this review recorded the details of such information in the space for comments in the data collection section. Likewise, rather than creating an independent section for outcome variables, it was addressed in the section on data collection together with other types of variables. The information on study eligibility (Cochrane, 2014; EPPI-Centre, 2019) was not collected, as the included studies at

this stage were already considered to be eligible for this review. The information concerning ethics (e.g., whether ethical approval and/or informed consent were obtained) was not collected unless there was a violation of ethical considerations worth noting in the allocated comment section. Finally, as the present review is not purported to aggregate quantitative data (e.g., a meta-analysis), only the relevant numerical data (e.g., sample size, data for individual participants, mean, standard deviations, test statistics, P values, confidence intervals, and effect estimates; Li et al., 2019), where available, were collected in the results/findings section in the table to the extent that they can facilitate the configuration of data to answer the review questions.

3.9 Risk of bias in individual studies

A systematic review necessitates a risk of bias assessment for included studies because the review results and conclusions are affected by biased results from those studies (Boutron et al., 2019). For this purpose, a thorough quality assessment of each included study was conducted in the current review (Higgins, Savović, et al., 2019; Moher et al., 2009). Higgins, Savović, et al. (2019)

recommend that systematic reviews identify domains of quality and assess both the proposed overall and domain-level quality, which was separately addressed in the current review using two critical quality appraisal tools.

3.9.1 Overall quality assessment

The overall quality of evidence from each study was appraised by a Weight of Evidence (WoE) framework (Gough, 2007). Gough (2007) argues that the quality and relevance of knowledge can be evaluated according to generic and purpose-specific criteria. Generic criteria concern judgments about how well a study is executed in a general sense as it stands, that is, disregarding the purpose of a review. In contrast, review-specific criteria appraise a study in relation to review questions from three dimensions. Following Higgins, Savović, et al. (2019), signaling questions (i.e., a series of questions to elicit information about features of the study that are relevant to its quality) were formulated to evaluate the quality and relevance of a study based on four criteria, which are elaborated in Table 6.

Table 6 WoE criteria and signaling questions

Criterion	Component	Question	Rationale	Evaluation
WoE A refers to a generic (non-review specific) judgment about the trustworthiness (i.e., how convincing the finding is; Gorard, 2014; see Appendix G for how trustworthiness was assessed)	Research design; data collection; data analysis; results, discussion, conclusion	Was the study executed in a trustworthy way?	The trustworthiness of the quantitative findings was evaluated in terms of validity (i.e., whether meaningful inferences can be drawn from the data about the population of interest) and reliability (i.e., the extent to which measurements or instruments are repeatable over time and over a variety of conditions; Creswell & Creswell, 2017). In contrast, the trustworthiness of qualitative studies was assessed for qualitative validity (i.e., whether the findings are accurate from the standpoint of the researcher, the participant, or the people external to the study, such as reviewers or readers; Creswell & Creswell, 2017; Creswell & Miller, 2000) and qualitative reliability (i.e., the consistency of the process and findings of a study; Golafshani, 2003; Noble & Smith, 2015).	Trustworthiness was scored from 2 to 0. Score 2 (high) was assigned if the study's findings had high validity and high reliability. Score 1 (medium) was assigned if either validity or reliability was high. Score 0 was assigned if the study had low validity and low reliability. When a mixed-methods design was employed, a lower score assigned to either a quantitative or a qualitative method was

				used to rate the validity and reliability of the study.
WoE B involves a review-specific judgment about the appropriateness of research design. Each study was appraised for its appropriateness in addressing each of the three review questions (Gough, 2007).	Research design	Was the study design appropriate for answering the IRQ?	The first IRQ asks for the evidence available on the deployment of listening strategies in L2 lecture comprehension. Although potentially all the research design types identify listening strategy use, the study which incorporated a descriptive research design, such as a quantitative descriptive study design ¹ or a qualitative study design, was considered more appropriate than other research design types. In addition, the study needs to collect empirical data on listening strategies.	Appropriateness was scored from 2 to 0. Score 2 (high) was assigned if the study employed either a quantitative descriptive study design or a qualitative study design. Score 1 (medium) was assigned if the study used other research design types. Score of 0 (low) was assigned if a study did not collect empirical data on listening strategies.

¹ Quantitative descriptive studies describe the relationships between variables without assuming any causal relationships (Hong et al., 2018).

The second IRQ concerns the evidence on the effectiveness of listening strategies in L2 lecture comprehension. A causal relationship between variables needs to be empirically demonstrated to ensure that the effectiveness is not observed by chance (Davidson, 2011). For this purpose, true experiments, in which participants are randomly assigned to groups (i.e., randomized controlled trials; Creswell & Creswell, 2017), are regarded as “the gold standard” (Cartwright, 2010, p. 59). Quasi-experiments (i.e., non-randomized controlled trials; Creswell & Creswell, 2017), in contrast, bear similar design features except that unknown differences may exist between groups due to the lack of random assignments of participants. Despite the limited external validity, it was expected that quasi-experimental designs are more commonly employed than experimental designs (Creswell & Creswell, 2017). While cross-	Appropriateness was scored from 2 to 0. Score 2 (high) was given if the study employed a randomized or non-randomized controlled trial to explore the relationship between listening strategies and listening performance, abilities, skills, or proficiency with linguistic/prior knowledge controlled for. Score 1 (medium) was given if the study used other study designs to explore the relationship without controlling for linguistic/prior
--	---

	sectional analytic study designs² (one of the non-randomized controlled trials) or qualitative study designs might be used to investigate hypothesized relationships between the variables, such study designs are limited in their ability to draw valid conclusions about causality (Kesmodel, 2018). As the review is interested in the indication of the effectiveness rather than the extent of effect (which should be addressed in a meta-analysis; Gough, 2007), the research design types mentioned here were regarded as informative. Another important concern relates to the control for linguistic/prior knowledge that might affect the relationship between listening strategies and listening performance as a confounding variable (Macaro et al., 2007). To clarify the theoretical causal relationship between listening strategies and listening performance,	knowledge. A study was scored 0 (low) if the study did not explore the relationship between listening strategies and listening performance, abilities, skills, or proficiency.
--	---	---

² Cross-sectional analytic studies explore the theory-imposed relationship between variables at a particular point in time (Hong et al., 2018).

	a question was asked about whether the study design enabled the control for linguistic/prior knowledge.	
	The third IRQ involves the evidence on the development of listening strategies in L2 lecture comprehension. In the field of developmental psychology, either longitudinal or cross-sectional research designs have been used to observe human development within individuals (i.e., longitudinal) or between groups (i.e., cross-sectional; Putnick, 2018). Nevertheless, cross-sectional designs are subject to external validity threats due to the variability between cohorts, which may compromise the claims about development over time (Robinson et al., 2005). As with other studies (Grammer et al., 2013), longitudinal designs, which enable the observation of the developmental change within individuals, are regarded as more appropriate for the third IRQ.	Appropriateness was scored from 2 to 0. Score 2 (high) was given if the study employed a longitudinal study design to explore the development of listening strategies. Score 1 (medium) was assigned if the study used a longitudinal study design without collecting data on listening strategies at multiple time points or if the study used a cross-sectional design to explore the development of listening strategies

				across cohorts. Score 0 (low) was given when other research designs were incorporated.
WoE C concerns a review-specific judgment about the relevance of the study topic and context and relates to how directly the study addresses the focus of the review questions (Gough, 2007). The relevance of the study was assessed in terms of task and text characteristics of the listening activity employed in each study.	Data collection	Was the listening activity relevant to an authentic lecture situation?	The relevance of the topic and context of a study to the review questions concerns whether the study results/findings can be generalizable to listening strategies in L2 lecture comprehension. In this review, relevance depends on task characteristics and text characteristics of the listening activity employed in each study. As for task characteristics, listening to a lecturer or a teacher in an authentic lecture situation was regarded as realistic (Bronfenbrenner, 1976, 1979) or having high ecological validity (i.e., the extent to which the research conditions and the phenomenon reflect what happens in real-world settings; Gehrke, 2018). Text characteristics were judged to be authentic if there were explicit descriptions (i.e., evidence)	Relevance was scored from 2 to 0. Score 2 (high) was assigned if the listening activity of the study was both realistic (task characteristics) and authentic (text characteristics). Score 1 was assigned if the listening activity of the study was either realistic or authentic. Score 0 was assigned if the listening activity of the study was neither realistic nor authentic.

that suggested so. On the other hand, if the spoken texts were written, short, non-academic, or unknown, text characteristics were considered non-authentic.				
WoE D is a review-specific judgment about the overall contribution of the study to a specific review question.	WoE A, B, and C	Does the study contribute to the systematic review question?	Gough (2007) reports that research teams vary widely with respect to how the generic and review-specific components of quality judgment should be combined. While one may weight either generic or review-specific components relatively more highly than the other, it was decided, in the current review, that the overall judgment takes into account each of WoE A (trustworthiness), B (appropriateness), and C (relevance) equally.	The scores from the assessment of WoE A, B, and C were summated to produce a composite score in the range of 6 and 5 (high), 4 and 3 (medium), and 2 and 1 (low). If WoE B was scored 0, N/A (i.e. not applicable in relation to each review question) was assigned to the study.

The actual evaluation followed two steps. First, studies were evaluated against WoE A, B, and C. Then, the scores were summated to produce a composite score for WoE D. Since WoE B was evaluated for each of the three IRQs, three WoE D scores were produced. Therefore, the included studies were appraised separately in relation to the three IRQs. The advantage of the scoring system is its transparency. Each WoE criterion was explicitly described in the table above. The use of simple summation of scores makes the weighting procedure clear and traceable, thereby increasing the reliability of the evaluation. Although scores were assigned to levels arbitrarily without regard to the qualitative difference between high, medium, and low, the classification and its algorithm were considered useful for the purpose of interpretation and comparison of the overall qualities of individual studies in relative terms. In Chapter 4, the levels (low, medium, and high) for WoE A, B, C, and D are reported for each study.

3.9.2 Methodological quality assessment

The domain-level quality of individual studies was appraised with the Mixed Methods Appraisal Tool version 2018 (MMAT; Hong et al., 2018). The MMAT is the

third version of a critical appraisal tool and assesses the methodological quality of individual studies within five broad design types including qualitative, quantitative randomized controlled trials, quantitative non-randomized, quantitative descriptive, and mixed methods. It asks two screening questions to assess the eligibility of a study for further appraisal and five appraisal questions to evaluate the quality of each study. The five questions differ significantly depending on the type of research design employed and appraise how rigorously the study in question is conducted. Another distinctive characteristic of this appraisal tool involves the discouragement of presenting an overall numerical score for each study. Instead, it shows a rating for each criterion to enhance our understanding of what aspects of a study are problematic. For these characteristics, the MMAT was used to provide a detailed picture of the quality of included studies and add another layer of analysis for this review.

The WoE criteria and scoring procedure were verified by a content expert on 17th June 2020. Using the two critical appraisal tools, I assessed each study from 17th June 2020 to 20th July 2020. The judgments about the quality of included studies were further supported by written justifications in Chapter 4 (Higgins, Savović, et al., 2019).

3.10 Risk of bias across studies

Bias across studies concerns the bias that may exist in the cumulative evidence (Liberati et al., 2009). In systematic reviews that follow the PRISMA statement, risk of bias across studies is often assessed in terms of publication bias (i.e., incomplete picture of the evidence) and selective reporting of results within the included studies (Liabo et al., 2012; Liberati et al., 2009; Moher et al., 2009). However, it is questionable how directly “cumulative evidence” is relevant to the current review questions because the review does not intend to aggregate homogeneous data. The other similar criteria in the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) handbook, which was developed to assess intervention studies in health care research, relate to limitations in study design or execution (i.e., whether there is a risk of bias due to such limitations), the inconsistency of results across studies (i.e., whether unexplained heterogeneity exists across studies), indirectness of evidence (i.e., whether the evidence is generalizable to the population), and imprecision (i.e., the point estimates with large confidence intervals; Liabo et al., 2012; Schünemann et al., 2013). Although these

criteria are used in many systematic studies taking an aggregative approach, they may not be applicable to the purpose of this review.

While the above criteria address quantitative evidence, the GRADE-CERQual (“Confidence in the Evidence from Reviews of Qualitative research”; Lewin et al., 2018a) evaluates qualitative evidence. CERQual offers four criteria, such as methodological limitations (i.e., the extent to which there are concerns about the design or conduct of the included studies that contributed evidence to a review finding), coherence (i.e., how clear and well supported is the fit between evidence and a review finding), adequacy of data (i.e., the degree of richness and quantity of data supporting a review finding), and relevance (i.e., the extent to which the body of evidence from included studies supporting a review finding is applicable to the context [perspective, population, phenomenon of interest, setting] specified in the review question). It is important to note that this framework is originally created to assess qualitative evidence. Lewin et al. (2018a) specify that CERQual is not intended to appraise quantitative or mixed methods data (cf. a modified GRADE approach; Okada et al., 2020). However, CERQual’s four criteria seemed to be relatively congruent with the findings from the configurative synthesis and qualitative judgments made using two critical appraisal tools.

Since no established assessment tools for risk of bias across a heterogeneous set of studies were located, this review attempted to narratively synthesize the quality of evidence across studies with respect to CERQual's four criteria. This is accomplished in Section 5.1 in Chapter 5.

3.11 Synthesis of results

The PRISMA Statement (Moher et al., 2009) encourages the reporting of a four-phase flow diagram (i.e., the PRISMA flow diagram), whose format was obtained online in the form of a word document file (PRISMA, 2015). In this review, the PRISMA flow diagram was modified to present the number of studies screened, assessed for eligibility, and included in the review as well as the reasons for exclusions at each phase of the review process.

In answering the overarching review question, the results of overall study characteristics (i.e., publication details, research contexts, participants, and methodological characteristics) were shown in the form of an evidence map with

qualitative descriptions and visuals, such as tables and graphs for numerical data (Gough et al., 2012). Although there seems to be no agreement in how an evidence map should be reported (Miake-Lye et al., 2016), it was thought of as a useful way to identify gaps in knowledge in the existing literature on this topic.

As for the three IRQs, a narrative (i.e., qualitative) synthesis of included studies was drawn up to provide in-depth accounts of the findings and risk of bias in and across included studies. A narrative synthesis involves a commentary reporting of each study as well as differences and similarities across studies (Lucas et al., 2007). This approach was chosen to provide direct insights into the exploratory questions by synthesizing both quantitative and qualitative evidence (Boland et al., 2017; Snilstveit et al., 2012). Despite the flexibility in addressing heterogeneity in data, a narrative approach has been criticized for the lack of transparency and clarity in its procedure (Dixon-Woods et al., 2005; Mays et al., 2005). Nevertheless, the lack of robustness in evidence synthesis was complemented by incorporating established frameworks (the WoE framework and MMAT) to evaluate the quality of individual studies, as suggested by Popay et al. (2006). The overall and methodological quality of included studies was discussed, using tables and detailed descriptions. The narrative was organized according to quality ratings of included studies, which

helped approach the review questions in a structured way (J. Thomas & Harden, 2008).

Chapter 4 Results

This chapter reports the results of the systematic review. It explains the process of selecting the final studies included in the review. Next, descriptions of the included studies are presented. Then, the overall quality and the methodological rigor of the studies are reviewed in depth.

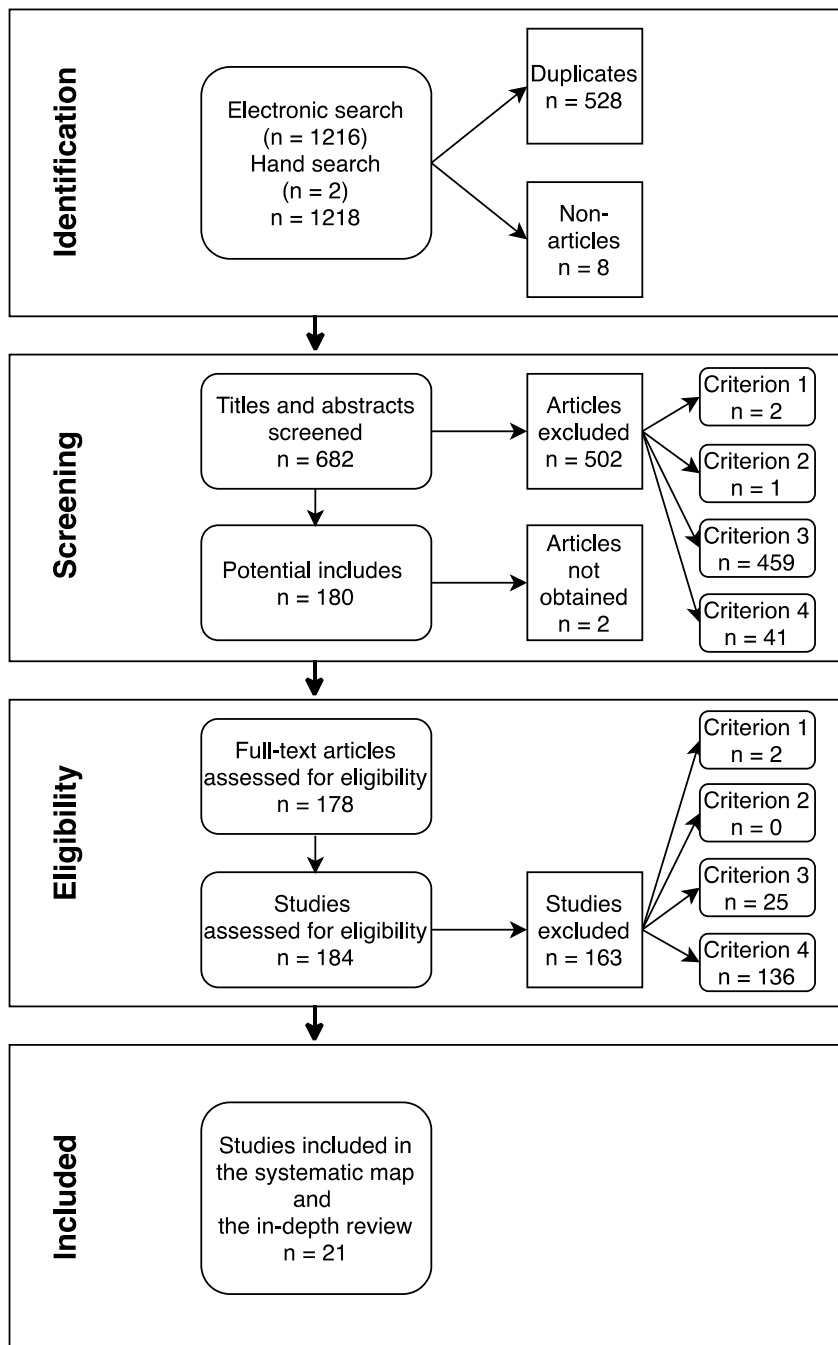
4.1 Included studies

Electronic and hand searching (backward citations) resulted in 1218 records, of which 536 records (duplicates and non-articles, such as synopses) were removed.

Title and abstract screening of 682 articles led to 178 articles, whose full-texts were retrieved either online or from the author(s). The full-text screening of these articles identified 184 studies, of which 163 studies did not meet the eligibility criteria.

Ultimately, a total of 21 studies were included in the review (see Appendix M for the annotated bibliography). An overview of searching and screening processes is summarized in the PRISMA flow diagram (Moher et al., 2009; see Figure 1).

Figure 1 Prisma flow diagram

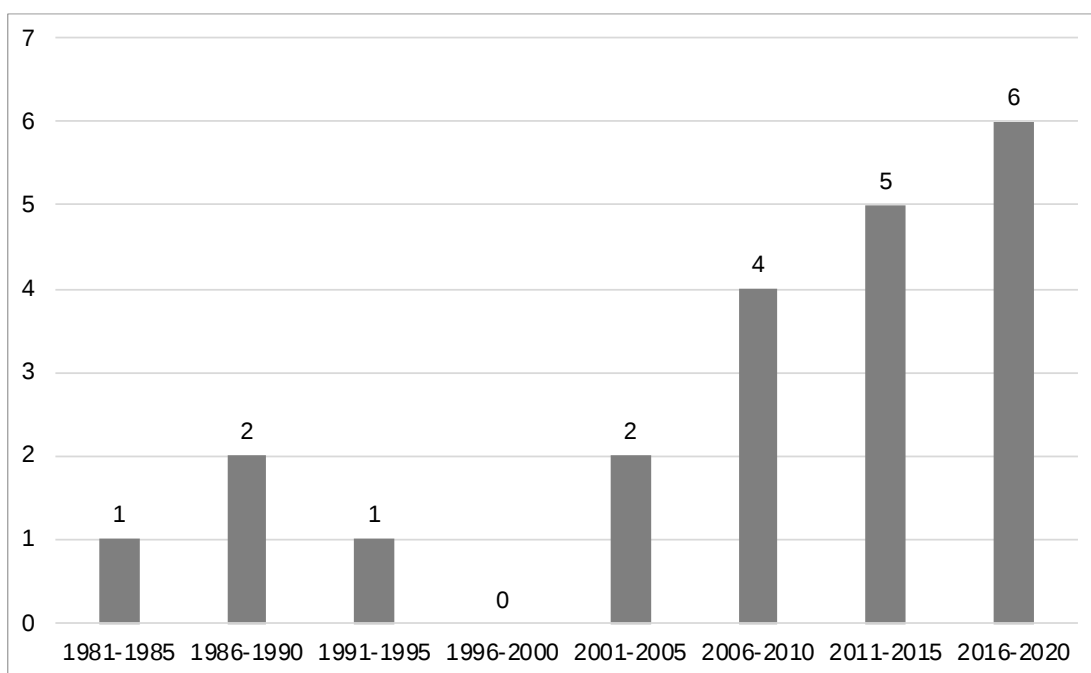


4.2 Characteristics of included studies

The included studies were published in journals from 1985 to 2020 (see Figure 2).

The number of publications fluctuated until the early 2000s. However, the number of studies has steadily increased since the late 2000s.

Figure 2 Studies by year (n = 21)



Most studies have been conducted at higher education institutions in the USA, China, Hong Kong, and Iran; there is little evidence available for secondary school learners in the USA and Hong Kong (see Table 7). This is reflected in Table 8,

showing the age of 18 and above is the most studied age range in the included studies. Moreover, Table 9 illustrates that about one-third of the included studies explored Chinese learners of English, whereas six studies collected data from learners of mixed L1s. One study explored learners of French in the USA without reporting the participants' L1s (see Table 10).

Table 7 Cross-tabulation: country by educational setting (n = 21)

	Higher education institution	Secondary school
USA	5	3
China	5	0
Hong Kong	1	1
Iran	2	0
UK	1	0
Japan	1	0
Taiwan	1	0
the Pacific Islands	1	0

Table 8 Studies by age (n = 21)

Age	Number
Around 20 to around 40	3
18 to 22	2
18	1
14 to 17	1
14	1
N/A	13
Total	21

Table 9 Studies by first language (n = 21)

Language	Number
Chinese	8
Persian	2
Japanese	1
Spanish	1
Thai	1
Mixed	6
N/A	2
Total	21

Table 10 Studies by second language (n = 21)

Language	Number
English as a foreign language	9
English as a second language	11
French	1
Total	21

About one-third of the included studies ($n = 7$) reported L2 proficiency (see Table 11). Slightly more studies ($n = 11$) reported L2 listening proficiency (see Table 12); these studies either divided listening proficiency into two groups (e.g., high vs. low) or included learners with intermediate listening proficiency levels according to standardized test results. Nevertheless, none of the studies used the same L2 proficiency or L2 listening proficiency measure.

Table 11 Studies by second language proficiency (n = 21)

L2 proficiency	Number
Intermediate by school	2
Intermediate by the CEFR	1
Upper intermediate by the OPT	1
Upper intermediate by the TOEFL PBT	1
A variety of proficiency levels by the TOEFL PBT and teacher judgement	1
A variety of proficiency levels by the IELTS	1
N/A	14
Total	21

Note. CEFR: the Common European Framework of Reference for Languages; OPT:

the Oxford Placement Test; TOEFL PBT: the Test of English as a Foreign Language

paper-based test; IELTS: the International English Language Testing System.

Table 12 Studies by second language listening proficiency (n = 21)

L2 listening proficiency	Number
High (effective) and low (ineffective) by teacher judgement and test scores	5
A variety of intermediate levels by test scores	4
Intermediate learners (by test scores) divided into good and weak listeners (by self-rating)	1
N/A	11
Total	21

Note. Flowerdew and Miller (1992) divided intermediate learners into good and

weak listeners using self-rating.

About half of the included studies ($n = 10$) reported participants' gender (see Table

13). Seven studies explored both female and male participants, although a majority

of the participants in the two studies were either female (Jin & Xu, 2017) or male

(Smidt & Hegelheimer, 2004). Furthermore, about one-third of the included studies

($n = 9$) did not report learners' discipline (see Table 14). The other studies included learners from either one discipline ($n = 6$) or a variety of disciplines ($n = 6$). Only one study (Olsen & Huckin, 1990) attempted to explore subject-specific (engineering) listening strategies.

Table 13 Studies by gender ($n = 21$)

Gender	Number
Both	5
Female (majority)	1
Male (majority)	1
Female	2
Male	1
N/A	11
Total	21

Note. Among the included studies, Smidt and Hegelheimer (2004) was the only study that investigated the relationship between listening strategies and gender.

Table 14 Studies by discipline ($n = 21$)

Disciplines	Number
English	3
Teaching English as a Foreign language	1
Teaching English as a Second language	1
Engineering	1
A variety of disciplines	6
N/A	9
Total	21

A questionnaire was the most commonly used strategy elicitation method in the included studies, followed by semi-structured interviews, retrospective interviews, and think-aloud protocols (see Table 15). “Other” methods included observation, diaries, or other varieties of interview techniques.

Table 15 Studies by strategy elicitation method (n = 26, not mutually exclusive)

Method	Number
Questionnaire	8
Semi-structured interview	4
Retrospective interview	3
Think-aloud protocol	2
Stimulated recall	1
Unstructured interview	1
Other	6
N/A	3
Total	28

Only four studies investigated listening strategies in authentic lectures (see Table 16). The other studies prepared audio or video recordings, which were often accompanied by follow-up tasks.

Table 16 Studies by task characteristics of listening activity (n = 21)

Task characteristics	WoE C	Number
Listening to a lecturer/a teacher	Realistic	4
Listening to audio recordings	Unrealistic	3
Listening to audio recordings and answering questions		5
Listening to audio recordings and summarizing the content		2
Listening to video recordings		1
Listening to video recordings and answering questions		3
Listening to audio and video recordings		2
A variety of listening activities (including listening to a lecturer)		1
	Total	21

Note. Task characteristics were judged to be either realistic or unrealistic according to the WoE C criterion (see Appendix H for further details).

Moreover, Table 17 shows that the text characteristics of listening activities included in nine studies were considered authentic. More than half of the included studies included spoken texts that were short and/or scripted ($n = 4$), non-academic ($n = 3$), or unknown ($n = 5$) to a non-negligible extent.

Table 17 Studies by text characteristics of listening activity (n = 21)

Text characteristics	WoE C	Number
Real lecture	Authentic	4
Long excerpt from real lecture		5
Short and scripted text	Non-authentic	3
Non-academic text		3
Scripted text		1
Unknown text		5
	Total	21

Note. Text characteristics were judged to be either authentic or non-authentic according to the WoE C criterion (see Appendix I for further details).

In summary, the current state of empirical research related to listening strategies in L2 lecture comprehension is characterized by the following features:

- Research in this area has received increased attention since the late 2000s.
- The clearest evidence is available for Chinese ESL or EFL learners in higher education in the USA and China. A few studies researched secondary school learners of English in the USA and Hong Kong.
- Some studies reported L2 proficiency or L2 listening proficiency; no studies used the same proficiency measure.
- About half of the included studies report information on gender or discipline; little evidence exists concerning the relationship between listening strategies and gender/academic discipline.
- Questionnaire and verbal reports are the most commonly used strategy elicitation methods; observation and diaries are the least common methods.
- A few studies explored listening strategies in response to listening to a lecturer/teacher; most studies used audio/video recordings.

- About half of the studies included authentic listening texts; the other half used non-authentic listening materials.

4.3 Overall quality of included studies

This section describes the overall quality of the included studies ($n = 21$) according to the WoE framework. The overall quality of the studies is reviewed in depth in relation to the first ($n = 18$), second ($n = 10$), and third ($n = 5$) IRQs.

4.3.1 Deployment of listening strategies in second language lecture comprehension

The first IRQ relates to the deployment of listening strategies in L2 lecture comprehension (see Appendix J for the summary). The appropriateness of three studies was judged to be “low” in relation to the first IRQ because they did not collect data on listening strategies. Thus, 18 studies, which informed the first IRQ, are reviewed in depth.

Three studies were judged to be of “high” overall quality (see Table 18): Flowerdew and Miller (1992), Fung and Macaro (2019), and O’Byryan and Hegelheimer (2009).

Flowerdew and Miller (1992) was an ethnography of 30 university students who studied Teaching English as a Second Language in Hong Kong. Using several data collection methods (observation, interviews, and diaries), the researchers identified the participants’ use of five strategies to facilitate lecture comprehension. The triangulation of data collection methods, the rich and thick description of the findings, and the presentation of disconfirming evidence enhanced the trustworthiness of the findings. However, strategies were neither defined nor situated in theories in the study. In Fung and Macaro (2019), listening strategies were elicited by a questionnaire that had been developed and validated by one of the authors based on Briggs’ (2015) distinction between strategies (as mental actions) and opportunities for strategic behavior (as observable actions; Fung, 2016). The questionnaire consisted of 65 items, which were summarized into 10 factors of strategies and three factors of opportunities for strategic behavior. Most strategies were related to the activation of prior knowledge (e.g., recall of prior knowledge), properly reflecting the nature of lecture comprehension. While the reliabilities of the instruments were not checked, the study successfully gained rich insights into the strategies that secondary school learners in Hong Kong used to

listen to a teacher. In the third study, O'Bryan and Hegelheimer (2009), used an open coding approach to identify listening strategies in ESL students' verbal data, which were later matched to the existing strategy taxonomies by O'Malley and Chamot (1990) and Vandergrift (1997). Triangulation of data collection methods, careful training on the think-aloud procedure, and the establishment of intercoder agreement resulted in high trustworthiness of the findings. Nevertheless, the artificial listening activity (listening to podcasts) decreased the relevance of this study.

Table 18 Weight of evidence assessment for high overall quality studies in relation to the first in-depth review question

Study	WoE A Trustworthiness of the study's findings	WoE B Appropriateness of the research design in relation with the first review question	WoE C Relevance of the study in relation with the first review question	WoE D Overall contribution of the study for the first review question
Flowerdew & Miller (1992)	High	High	High	High
Fung & Macaro (2019)	Medium	High	High	High
O'Bryan & Hegelheimer (2009)	High	High	Medium	High

Eight studies were evaluated as "medium" overall (see Table 19), of which three studies received "low" scores for trustworthiness. Schroeder (2016) developed and

validated the Academic Spoken English Strategies Survey (ASESS) to assess the listening and speaking strategy use of NNES graduate students in the US. As with Fung and Macaro (2019), the questionnaire underwent rounds of piloting. An exploratory factor analysis was run on 384 participants' responses to the questionnaire, resulting in five factors. However, the identified factors lacked a theoretical explanation. For instance, it was not clear why the strategy to "notice the speaker's facial expressions, gestures, and voice changes" loaded onto "strategies for keeping on track while listening" rather than "strategies for enhancing understanding while listening". In addition, the fifth factor consisted of only one strategy. These limitations reduced the (construct) validity and reliability of the questionnaire. In evaluating podcasts developed for academic listening in French, Weinberg et al. (2011) mainly relied on a questionnaire to collect data on the perceived change in participants' use of listening strategies before and after viewing podcasts. While the focus of the podcasts (on listening to authentic lectures in French) increased real-world relevance, the study did not make explicit reference to the definitions or theories of listening strategies.

In Smidt and Hegelheimer (2004), ESL learners were asked to watch an authentic lecture video and answer comprehension questions. Verbal data were elicited by

retrospective interviews from nine participants, which were coded into metacognitive and cognitive strategies following Vandergrift's (1997) procedure. Interestingly, elaboration (i.e., use of world knowledge) was not reported by the learners, possibly due to the lack of relevant prior knowledge (an issue with relevance) or the validity issues inherent in the retrospective interviewing technique (i.e. forgetting and fabrication; Russo et al., 1989). In Schroeder (2016) and Smidt and Hegelheimer (2004), no notable measures were taken to ensure the trustworthiness of data collection and analysis procedures.

Table 19 Weight of evidence assessment for medium overall quality studies in relation to the first in-depth review question

Study	WoE A Trustworthiness of the study's findings	WoE B Appropriateness of the research design in relation with the first review question	WoE C Relevance of the study in relation with the first review question	WoE D Overall contribution of the study for the first review question
Schroeder (2016)	Low	High	High	Medium
Weinberg et al. (2011)	Low	High	High	Medium
Smidt & Hegelheimer (2004)	Low	High	Medium	Medium
O'Malley et al. (1989)	High	High	Low	Medium
Rukthong & Brunfaut (2020)	High	High	Low	Medium
Lee & Cai (2010)	High	Medium	Low	Medium
Cai & Lee (2010)	High	Medium	Low	Medium
Cai & Lee (2012)	High	Medium	Low	Medium

The other “medium” overall quality studies utilized verbal report procedures, such as think-aloud protocols (O’Malley et al., 1989), stimulated recalls (Rukthong & Brunfaut, 2020), or interviews with or without retrieval cues (Cai & Lee, 2012, 2010; Lee & Cai, 2010). O’Malley et al. (1989) provided eight high school students with activities to practice verbalizing their thoughts in think-aloud protocols. The description of metacognitive, cognitive, and social/affective strategies in three stages of comprehension (J. R. Anderson, 1983) was rich in detail and included disconfirming evidence, increasing the trustworthiness of their findings. Rukthong and Brunfaut (2020) increased the validity of verbal reports by use of stimulated recalls, a piece of qualitative analysis software (NVivo to increase transparency), and triangulation. In the study, video-recordings were created while 12 Thai learners of English engaged in integrated listening-to-summarize tasks. These videos were later used to facilitate the participants’ recall of their strategy use. The verbal reports were coded into various cognitive and metacognitive strategies according to the strategy taxonomies by Goh (2002b), O’Malley et al. (1989), Vandergrift (2003a), Vandergrift and Goh (2012). The findings were corroborated by the participants’ notes as well as oral and written summaries of the listening contents.

Cai and Lee (2012, 2010) and Lee and Cai (2010) focused on the use of strategies and knowledge sources to process unfamiliar words in listening comprehension. In these studies, Chinese learners of English were asked to listen to nine texts (from a science textbook) and answered questions about how they made sense of the meaning of novel words with or without retrieval cues. The significance of these studies is the identification of different types of lexical inferencing based on knowledge sources, such as textual (e.g., information about target word or co-text) and extra-textual knowledge. Their coding procedures were clear, which helped achieve high intercoder reliabilities (around 95%). Despite the high trustworthiness, the five studies (Cai & Lee, 2012, 2010; Lee & Cai, 2010; O'Malley et al., 1989; Rukthong & Brunfaut, 2020) received low marks for relevance due to the low ecological validity of the listening activities.

The rest of the studies had a "low" overall quality (see Table 20). Two studies attempted to identify listening strategies that were unique to academic listening. In Du (2018), a new strategy taxonomy was developed by referring to interview data as well as existing taxonomies (Goh, 2002a; Graham et al., 2008; Santos et al., 2008), although the details of the procedure were not explained. The reporting of the qualitative findings contained inconsistencies and lacked clarity, which lowered the

trustworthiness of the study. Interested in learners' strategy use in academic listening practices, Jin and Xu (2017) developed a new questionnaire by incorporating O'Malley and Chamot's (1990) distinction of metacognitive, cognitive, and social-affective strategies and Richards' (1983) list of academic skills. However, the questionnaire was neither piloted nor validated in the study.

Table 20 Weight of evidence assessment for low overall quality studies in relation to the first in-depth review question

Study	WoE A Trustworth iness of the study's findings	WoE B Appropriaten ess of the research design in relation with the first review question	WoE C Relevance of the study in relation with the first review question	WoE D Overall contributio n of the study for the first review question
Du (2018)	Low	High	Low	Low
Jin & Xu (2017)	Low	High	Low	Low
Chand (2007)	Low	High	Low	Low
Rahimirad & Moini (2015)	Low	High	Low	Low
Bidabadi & Yamat (2013)	Low	High	Low	Low
Chou (2015)	Low	High	Low	Low
Hosogoshi (2016)	Low	Medium	Low	Low

The other low quality studies used existing strategy taxonomies with few or no modifications. Chand (2007) used O'Malley, Chamot, Stewner-Manzanares, et al.'s (1985) taxonomy to find listening strategies in the interview data from teachers and

learners who participated in a distance study skills course. However, identified themes lacked opposing views and were often based on only one participant's account. In Rahimirad and Moini (2015), participants were introduced to the concepts of metacognitive strategies proposed by Vandergrift (2003a) and were asked to answer questions concerning the effectiveness of strategy instruction during semi-structured interviews. Bidabadi and Yamat (2013) used semi-structured interviews to add more details to the findings from the MALQ. In the three studies, however, no measures were reported to increase the validity and reliability of the qualitative findings.

Chou (2015) explored the use of listening strategies in Taiwanese university students who participated in English for general/specific academic purposes (EGAP/ESAP). Listening strategies were identified via a questionnaire and retrospective interviews with reference to the frameworks for listening strategies (Vandergrift, 1997) and test-taking strategies (Cohen, 2006). Despite the high reliability of the questionnaire ($\alpha = .91$), the qualitative data collection and analysis procedures were not described in detail. Hosogoshi (2016) developed a new questionnaire using Vandergrift's (2003a) taxonomy of listening strategies to elicit Japanese EFL learners' strategies in response to lecture video with or without

captions/subtitles. This questionnaire was piloted on 86 students with similar study backgrounds, which was insufficient to establish external and statistical validity of the questionnaire.

4.3.2 Effectiveness of listening strategies in second language lecture

comprehension

The second IRQ concerns the effectiveness of listening strategies in L2 lecture comprehension (see Appendix K for the summary). Eleven studies had “low” appropriateness for the second IRQ because they 1) did not explore the effectiveness of listening strategies and 2) did not collect empirical data on listening performance, abilities, skills, or proficiency. Thus, the quality of evidence from 10 studies is discussed in detail.

Six studies resulted in “medium” overall ratings (see Table 21). O’Malley, Chamot, Stewner-Manzanares et al. (1985) randomly allocated 75 high school ESL students to two treatment and one control groups. For a period of eight days, the metacognitive group ($n = 27$) was taught one metacognitive strategy (selective

attention), one cognitive strategy (note-taking), and one socio-affective strategy (cooperation), whereas the cognitive group ($n = 26$) received instruction on the cognitive and socio-affective strategies. Overall, the treatment groups outperformed the control group. According to the prediction, the metacognitive group received significantly higher scores than the cognitive group in the first interim listening test ($p < .10$). However, the cognitive group performed better than the metacognitive group in the second ($p < .05$) and the third interim tests ($p < .01$). The difference between the three groups failed to reach significance in the last interim test ($p = .626$) as well as the posttest ($p = .162$). The study was able to control for the participants' linguistic and prior knowledge using a randomized controlled study design, although the details of the random allocation procedure were not reported. Treatment was consistent with the planned intervention. Nevertheless, the increased task difficulty in the latter listening tests may have caused mixed results. Moreover, the trustworthiness of the study decreased because there was no reporting of the reliabilities for the listening tests.

Table 21 Weight of evidence assessment for medium overall quality studies in relation to the second in-depth review question

Study	WoE A Trustworthiness of the study's findings	WoE B Appropriateness of the research design in relation with the second review question	WoE C Relevance of the study in relation with the second review question	WoE D Overall contribution of the study for the second review question
O'Malley et al. (1985)	Medium	High	Medium	Medium
O'Malley et al. (1989)	High	Medium	Low	Medium
O'Bryan & Hegelheimer (2009)	High	Medium	Low	Medium
Rukthong & Brunfaut (2020)	High	Medium	Low	Medium
Lee & Cai (2010)	High	Medium	Low	Medium
Cai & Lee (2012)	High	Medium	Low	Medium

The other five studies which received “medium” overall ratings can be highly trusted in a generic sense. In O'Malley et al. (1989), effective and ineffective listeners were differentiated according to subjective (e.g., attentiveness in class) and objective criteria (e.g., reading test scores). The analysis of their verbal reports showed that effective listeners maintained and directed their attention to the listening task (perception), listened for larger chunks, such as phrases or sentences, attended to intonation and pauses to use top-down processing (parsing), and used inferencing and elaboration with reference to their world knowledge, personal knowledge, and self-questioning. On the other hand, ineffective listeners relied more on bottom-up

processing and used multiple strategies, such as inferencing, translation, and self-monitoring, to compensate for the failure in parsing the text. Surprisingly, less use of metacognitive strategies by students with higher listening proficiency was reported in O'Bryan and Hegelheimer (2009), although it may just mean the students were able to understand the listening text effortlessly. The study also found that higher proficiency students used inferencing and elaboration at a global level (across sentences and propositions), whereas low proficiency students employed the same strategies at a local (word and sentence) level. In Rukthong and Brunfaut (2020), higher scorers in listening-to-summarize tasks used more types of metacognitive strategies and used inferencing more accurately. On the contrary, lower scorers were less successful in employing inferencing and elaboration and relied more on directed attention, fixation, and reconstruction strategies to recover the lack of understanding of the content.

Two studies focused on strategies and knowledge sources to process unfamiliar words in listening comprehension. Lee and Cai (2010) reported that learners with higher listening proficiency combined morphological clues, (local and global) co-textual clues, and extra-textual knowledge effectively to make lexical inferencing and noticed the dissonance between the clues, whereas learners with lower

listening proficiency failed to do so. The subsequent study by Cai and Lee (2012) confirmed that the use of morphological clues, global co-textual clues, and extra-textual knowledge actually led to better lexical inferencing and text comprehension. The above findings suggest that at least some of the differences in strategy use derive from varied linguistic knowledge. However, linguistic knowledge was never measured in these studies.

The five “low” overall quality studies are characterized by “low” trustworthiness and/or “low” relevance (see Table 22). In Rahimirad and Moini (2015), Iranian EFL learners who received metacognitive strategy instruction for five weeks achieved significantly higher scores than those who did not receive instruction in the posttest (equivalent to the IELTS listening exam). During the interviews, the learners in the treatment group reported increased use of metacognitive strategies, such as planning, directed attention, evaluation, and monitoring (except selective attention). The interview, however, relied on an understanding of the concept of each strategy, which may have been too demanding for the participants. Carrier (2003) provided seven high school ESL students with explicit strategy instruction on a variety of listening skills including discrete (bottom-up) and video (top-down) listening skills for six weeks. Posttest scores were significantly larger than pretest scores for both

discrete ($p < .025$, one-tailed) and video listening skills ($p < .01$, one-tailed), indicating the effectiveness of bottom-up and top-down listening strategies. Despite the high reliability of the listening tests, the small sample size and no inclusion of a control group decreased the trustworthiness of the findings.

Table 22 Weight of evidence assessment for low overall quality studies in relation to the second in-depth review question

Study	WoE A Trustworthiness of the study's findings	WoE B Appropriateness of the research design in relation with the second review question	WoE C Relevance of the study in relation with the second review question	WoE D Overall contribution of the study for the second review question
Rahimirad & Moini (2015)	Low	High	Low	Low
Carrier (2003)	Medium	Medium	Low	Low
Smidt & Hegelheimer (2004)	Low	Medium	Medium	Low
Chou (2015)	Low	Medium	Low	Low
Du (2018)	Low	Medium	Low	Low

Smidt and Hegelheimer (2004) investigated listening strategies in response to a 15-minute online video with overhead transparency notes and picture slides of an academic lecture. The analysis of the retrospective interviews of nine university ESL students indicated that high proficiency students used both metacognitive and

cognitive strategies and favored monitoring and inferencing, while low proficiency students used only cognitive strategies. Chou's (2015) study was unique in that it compared the listening strategies perceived to be most and least useful by Taiwanese lower-intermediate learners of English in EGAP and ESAP courses. The learners generally found "listening to main ideas" to be most useful and saw translation to be the least useful strategy. Strategies that involved the use of prior knowledge were perceived to be useful in EGAP listening. In contrast, note-taking and test wiseness (e.g., choosing an option that seemed to deviate from others) strategies were considered most helpful in ESAP listening. However, individual differences were also observed in terms of the perceived usefulness of note-taking and lexical inferencing strategies. There was an issue with the study design because the listeners participated in the ESAP course after taking part in the EGAP course. Thus, participation in the EGAP course may have affected the use of listening strategies in the ESAP course. Moreover, the reasons for the striking differences in listening strategy use between EGAP and ESAP courses were not explored in depth and remained unclear. Nevertheless, the two studies discussed here reported no measures to increase the trustworthiness of the qualitative findings.

Du (2018) was the only study that directly addressed the relationship between listening strategies, listening proficiency, and linguistic knowledge. Du divided nine Chinese university EFL students into high and low listening proficiency and linguistic knowledge groups and observed their use of strategies in response to different listening question types. The study identified strategies that were different as well as shared among the groups. However, the description of the findings often included inconsistencies and lacked evidence for their claims about the strategies commonly used by different groups of learners. It was not clear what strategies led to successful comprehension except selective attention used by learners with high listening proficiency.

4.3.3 Development of listening strategies in second language lecture comprehension

The third IRQ pertains to the development of listening strategies in L2 lecture comprehension (see Appendix L for the summary). Sixteen studies were excluded from the in-depth review for the third IRQ because the studies did not offer evidence

on the development of listening strategies. Therefore, the findings of the remaining five studies are synthesized below.

O'Bryan and Hegelheimer (2009) was the only study that received a "high" overall rating (see Table 23). In the study, four university ESL students received explicit strategy instruction on academic listening strategies over a 15-week semester. Instructor-designed podcasts were assigned as homework so that the students could review and practice those strategies outside the classroom. The development of metacognitive awareness was explored primarily by assessing the change between pre- and post-MALQ scores, while this was triangulated with semi-structured interviews. The largest increase was found in the use of problem-solving strategies (inferencing, elaboration, evaluation, monitoring, and summary), followed by personal knowledge, directed attention, and mental translation. There was no change in planning and evaluation strategies. The increase in mental translation was surprising because it contradicted the researchers' initial expectations that successful listeners would rely less on translation. The findings were trustworthy for reasons, such as the triangulation of data collection methods and the rich and thick description of the qualitative findings. Nevertheless, the small sample size and the

low ecological validity of the listening activities (which did not occur in authentic lectures) posed threats to the generalizability of the findings.

Table 23 Weight of evidence assessment for a high overall quality study in relation to the third in-depth review question

Study	WoE A Trustworth iness of the study's findings	WoE B Appropriat eness of the research design in relation with the third review question	WoE C Relevance of the study in relation with the third review question	WoE D Overall contributio n of the study for the third review question
O'Bryan & Hegelheimer (2009)	High	High	Medium	High

Two studies were rated to be of “medium” overall quality (see Table 24). The experiment by O'Malley, Chamot, Stewner-Manzanares, et al. (1985) suggested that the learners who received listening strategy instruction outperformed those who did not in a listening comprehension test. However, the limited period of intervention (eight days) may have undermined the efficacy of faded instructional cues for using metacognitive strategies, leading to the inconclusive findings of the relative benefits of the two treatment groups (with or without metacognitive strategy instruction).

Weinberg et al. (2011) administered a questionnaire to students after they had watched a podcast episode (for academic listening in French) on a DVD in a class for seven weeks. Student responses to a five-point Likert scale revealed notable changes in the use of listening strategies learned in the first (cognitive and socio-affective strategies), the second (planning), and the third (monitoring) episodes. The findings, however, suffered from “low” trustworthiness because the perceived change in strategy use was measured at one point in time; no data were collected before watching each episode.

Table 24 Weight of evidence assessment for medium overall quality studies in relation to the third in-depth review question

Study	WoE A Trustworthiness of the study's findings	WoE B Appropriateness of the research design in relation with the third review question	WoE C Relevance of the study in relation with the third review question	WoE D Overall contribution of the study for the third review question
O'Malley et al. (1985)	Medium	Medium	Medium	Medium
Weinberg et al. (2011)	Low	Medium	High	Medium

Finally, two studies received “low” overall ratings (see Table 25). In Carrier (2003), explicit strategy instruction on both bottom-up and top-down listening strategies turned out to be effective for secondary ESL students. However, since the effectiveness of bottom-up listening strategy instruction was assessed using a discrete listening skills test (e.g., discriminating sounds), the study offers limited empirical insights into whether or how bottom-up strategies benefit listening comprehension. Moreover, a top-down listening skills test featured only the note-taking strategies despite a variety of listening strategies covered during the intervention. Rahimirad and Moini (2015) presented the evidence for the effectiveness of the metacognitive strategy instruction, which was provided systematically based on Vandergrift (2003a) and for a longer period (five weeks) than O’Malley, Chamot, Stewner-Manzanares, et al. (1985). Academic listening sections of the IELTS were administered in pre- and posttests, showing a significant difference between treatment and control groups only in the posttest ($t = -2.431$, $p < .05$). In the interview held after the posttest, the participants who received instruction stated that they increased the use of planning and directed attention most, followed by evaluation and monitoring. The researchers reported, however, that most learners had no idea about selective attention. These findings further corroborated the effectiveness of instruction. Nevertheless, the possibility exists that

too much was expected on the part of the participants for understanding the concepts of strategies and verbalizing their thoughts in their L2 (English).

Table 25 Weight of evidence assessment for low overall quality studies in relation to the third in-depth review question

Study	WoE A Trustworthi ness of the study's findings	WoE B Appropriateness of the research design in relation with the third review question	WoE C Relevance of the study in relation with the third review question	WoE D Overall contribution of the study for the third review question
Carrier (2003)	Medium	Medium	Low	Low
Rahimirad & Moini (2015)	Low	High	Low	Low

4.4 Methodological quality of included studies

This section presents the methodological quality of the included studies, which was evaluated using the MMAT. Five study design types included qualitative studies ($n = 5$), quantitative randomized controlled trials ($n = 1$), quantitative non-randomized studies ($n = 2$), quantitative descriptive studies ($n = 3$), and mixed methods studies ($n = 10$). The results of the evaluation of individual studies are discussed according to each design type below.

Among the five qualitative studies (see Table 26), three qualitative studies were of high methodological quality. Using an ethnographic approach, Flowerdew and Miller (1992) combined different data collection methods effectively and justified the qualitative findings with a rich description of learner perspectives. Two studies provided a qualitative description of their findings without specifying its methodology. O'Malley et al. (1989) provided think-aloud training sessions with participants; a high intercoder agreement was achieved in the analysis of verbal data. In Olsen and Huckin (1990), the two researchers engaged in the analysis of the content and structure of a lecture, participants' lecture summaries, and interview data. Both studies provided detailed examples to discuss the interpretation of the qualitative findings.

Table 26 Methodological quality of qualitative studies

Study	Research design	1.1. Is the qualitative approach appropriate to answer the research question?	1.2. Are the qualitative data collection methods adequate to address the research question?	1.3. Are the findings adequately derived from the data?	1.4. Is the interpretation of results sufficiently substantiated by data?	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?	Overall WoE IRQ1	Overall WoE IRQ1	Overall WoE IRQ2
Flowerdew & Miller (1992)	Qualitative study (ethnography)	Yes	Yes	Yes	Yes	Yes	High	N/A	N/A
Chand (2007)	Qualitative study (grounded theory)	Yes	Yes	No	No	No	Low	N/A	N/A
O'Malley et al. (1989)	Qualitative study (qualitative description)	Yes	Yes	Yes	Yes	Yes	Medium	Medium	N/A
Olsen & Huckin (1990)	Qualitative study (qualitative description)	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Du (2018)	Qualitative study (qualitative description)	Yes	Yes	Yes	No	No	Low	N/A	N/A

Two studies had issues with the analysis, interpretation, and synthesis of qualitative findings. Despite the claim to have used a grounded theory approach, Chand (2007) identified the themes related to listening strategy use in interview data following O'Malley, Chamot, Stewner-Manzanares, et al.'s (1985) distinction of metacognitive, cognitive, and socio-affective strategies. Moreover, the themes were not convincing because the instances (that supported each theme) lacked opposing views and were often derived from a single person. In Du (2018), two researchers coded retrospective and unstructured interview data in reference to a predetermined strategy taxonomy. Despite the seemingly rigorous procedure, the reporting of the qualitative findings included inconsistencies and obscurities, which significantly reduced the trustworthiness of this study.

O'Malley, Chamot, Stewner-Manzanares, et al. (1985) conducted a quantitative randomized controlled trial (see Table 27). However, the method used to randomly allocate participants to groups was not explained in the study. Moreover, the teachers, who provided strategy instruction (intervention providers), also administered short listening comprehension tests (outcome measures) to the participants. This posed a threat to the validity of their findings.

Table 27 Methodological quality of quantitative randomized controlled trials

Study	Research design	2.1. Is randomization appropriately performed?	2.2. Are the groups comparable at baseline?	2.3. Are there complete outcome data?	2.4. Are outcome assessors blinded to the intervention provided?	2.5. Did the participants adhere to the assigned intervention?	Overall WoE IRQ1	Overall WoE IRQ1	Overall WoE IRQ2
O'Malley et al. (1985)	Quantitative randomized controlled trial	Can't tell	Yes	Yes	No	Yes	N/A	Medium	Medium

Two quantitative non-randomized studies did not explicitly state their target populations, justify their measurements; nor did they report on the fidelity of the intervention (see Table 28). Furthermore, Carrier (2003) did not include a control group or information about possible confounding variables.

Table 28 Methodological quality of quantitative non-randomized studies

Study	Research design	3.1. Are the participants representative of the target population?	3.2. Are measurements appropriate regarding both the outcome and exposure/intervention?	3.3. Are there complete outcome data?	3.4. Are the confounders accounted for in the design and analysis?	3.5. During the study period, is the intervention/exposure administered as intended?	Overall WoE IRQ1	Overall WoE IRQ1	Overall WoE IRQ2
Carrier (2003)	Quantitative non-randomized study (cross-sectional analytic study)	Can't tell	No	Yes	No	Can't tell	N/A	Low	Low
Hosogoshi (2016)	Quantitative non-randomized study (non-randomized controlled trial)	Can't tell	No	Yes	Yes	Can't tell	Low	N/A	N/A

Three quantitative descriptive studies conducted a questionnaire survey (see Table 29). Fung and Macaro (2019) presented a model case where the sample was matched with their target population (clearly specified); quantitative data collected from a large sample of participants via valid instruments underwent rigorous

statistical procedures according to their research questions. Nevertheless, the questionnaires developed by the two studies were not appropriate and lacked appropriate statistical analyses. Despite a number of pilot studies, the construct validity of Schroeder's (2016) strategy questionnaire was considered to be low because it lacked the theoretical accounts of what strategies refer to. The exploratory factor analysis may contain reliability issues because one of the identified factors loaded onto only one item. In Jin and Xu (2017), neither the population nor a sampling strategy was clearly explained. While low overall reliability ($\alpha = .726$) was achieved, no measures were taken to ensure the construct validity of the questionnaire. Moreover, the researchers did not clarify the types of correlations reported in the study.

Table 29 Methodological quality of quantitative descriptive studies

Study	Research design	4.1. Is the sampling strategy relevant to address the research question?	4.2. Is the sample representative of the target population?	4.3. Are the measurements appropriate?	4.4. Is the risk of nonresponse bias low?	4.5. Is the statistical analysis appropriate to answer the research question?	Overall WoE IRQ1	Overall WoE IRQ1	Overall WoE IRQ2
Fung & Macaro (2019)	Quantitative descriptive study (survey)	Yes	Yes	Yes	Yes	Yes	High	N/A	N/A
Schroeder (2016)	Quantitative descriptive study (survey)	Yes	Yes	No	Can't tell	No	Medium	N/A	N/A
Jin & Xu (2017)	Quantitative descriptive study (survey)	Can't tell	Can't tell	No	Yes	No	Low	N/A	N/A

Many of the included studies ($n = 10$) chose a mixed-methods design (see Table 30). On the positive side, most studies relied on both quantitative and qualitative data to answer research questions ($n = 9$) and brought together quantitative and qualitative findings in the results section ($n = 9$). The divergence was found in five studies in a way that qualitative findings added details to the quantitative findings; the rest of the studies did not report divergence or inconsistencies. Nevertheless, about half of the studies did not justify why they did so ($n = 6$) or did not ensure the quality of different components of the studies ($n = 7$). In particular, Bidabadi and Yamat, (2013) failed to supplement their quantitative results with rich and thick qualitative accounts of the participants' use of listening strategies. Nevertheless, other studies effectively compared and integrated quantitative and qualitative findings to shed more light on the topic.

Table 30 Methodological quality of mixed methods studies

Study	Research design	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?	5.2. Are the different components of the study effectively integrated to answer the research question?	5.3. Are the results adequately brought together into overall interpretations?	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	Overall WoE IRQ1	Overall WoE IRQ1	Overall WoE IRQ2
Rahimirad & Moini (2015)	Mixed methods study (convergent design) with quantitative randomized controlled trial and qualitative description	Yes	Yes	Yes	Yes	No	Low	Low	Low
Cai & Lee (2010)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	No	Yes	Yes	Yes	Yes	Medium	N/A	N/A
Lee & Cai (2010)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	No	Yes	Yes	Yes	No	Medium	Medium	N/A
Cai & Lee (2012)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	No	Yes	Yes	Yes	No	Medium	Medium	N/A
Chou (2015)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	Yes	Yes	Yes	Yes	No	Low	Low	N/A
Bidabadi & Yamat (2013)	Mixed methods study (convergent design) with survey and grounded theory	No	No	No	Yes	No	Low	N/A	N/A
Rukthong & Brunfaut (2020)	Mixed methods study (convergent design) with survey and grounded theory	Yes	Yes	Yes	Yes	Yes	Medium	Medium	N/A
O'Bryan & Hegelheimer (2009)	Mixed methods study (sequential exploratory design) with grounded theory and cohort study	Yes	Yes	Yes	Yes	Yes	High	Medium	High
Weinberg et al. (2011)	Mixed methods study (convergent design) with survey, qualitative description, and grounded theory	No	Yes	Yes	Yes	No	Medium	N/A	Medium
Smidt & Hegelheimer (2004)	Mixed methods study (convergent design) with the before-and-after trial without control, prevalence study, and qualitative description	No	Yes	Yes	Yes	No	Medium	Low	N/A

Chapter 5 Discussion and conclusion

This chapter discusses the results of this review. It attempts to answer each of the review questions with reference to the limitations of included studies. In the next section, the limitations of this review are explored. It concludes with the general interpretation of the results and implications for future research.

5.1 Summary of evidence

This section summarizes the main findings of the systematic review and discusses the quality of evidence across the included studies. It provides an overview of the current state of research available on this topic based on the characteristics of the included studies. Then, the quality of the reviewed studies is discussed in reference to the first, second, and third IRQs.

5.1.1 Overarching review question: What empirical research has been carried out on listening strategies for second language lecture comprehension between 1975 and 2020?

The descriptions of included studies suggest that research on this topic has been on the rise up until recently, reflecting the growth of the scholarly interest in language learning strategies (e.g., Mizumoto & Takeuchi, 2018) as well as English medium instruction (EMI; e.g., Macaro et al., 2018). Most studies explored learners of English in higher education with Chinese as their L1 or mixed language backgrounds in countries such as the USA, China, Hong Kong, and Iran. However, the pervasiveness of EMI (Macaro et al., 2018) necessitates more systematic efforts to investigate learners from different nations, educational phases, and different language and cultural backgrounds. Only one study investigated French as a medium of instruction; with a broader scope of encouraging multilingualism in higher education (Kirkpatrick, 2014), it is also meaningful to gain insights into other L2s as a medium of instruction.

Another major issue concerns the low ecological validity of the review findings. The included studies rarely controlled for participants' academic discipline. Only half of

the reviewed studies used authentic listening texts; even fewer studies explored listening strategies in real academic lectures. However, researchers have argued that L2 listeners go through different mental processes depending on learner characteristics as well as task and text characteristics of a listening activity (Field, 2011; Flowerdew, 1994; Richards, 2003; Rubin, 1994). Moreover, it is well known that discourse structure varies across subject areas (Dudley-Evans, 1994; Olsen & Huckin, 1990); it is likely that the mental processes required by learners of engineering significantly differ from those required by learners of English literature. Little attention paid to these factors makes the ecological validity of the listening activities included in many of the reviewed studies questionable. Therefore, more research is needed on how task characteristics (e.g., listening to a lecturer, audio recordings, or videos) and text characteristics (e.g., discourse structure in different disciplines) impact on listening strategies in L2 lecture comprehension.

5.1.2 First in-depth review question: What evidence is available for the *deployment* of listening strategies in second language lecture comprehension?

The in-depth review demonstrates that most studies (13 out of 18) relied on the existing taxonomies including those by Goh (2002b) or Vandergrift (1996, 1997, 2003a). These classifications expedit the use of prior knowledge, which allows rich insights into cognitive strategies, such as inferencing and elaboration that are crucial in understanding lectures. The affinity of these taxonomies for L2 lecture comprehension was found useful especially by those scholars who sought to identify listening strategies in verbal reports. Only two studies, Chand (2007) and Jin and Xu (2017), referred to simpler models (i.e., metacognitive, cognitive, and socio-affective strategies) proposed by O'Malley and his colleagues. And yet, Jin and Xu (2017) felt the need to incorporate Richards' (1983) taxonomy of academic listening skills into their questionnaire development. It should also be noted that some strategies (e.g., "socio-affective strategies" in O'Bryan & Hegelheimer, 2009; "elaboration" in Smidt & Hegelheimer, 2004) may not be elicited properly due to the artificial listening task (e.g., listening to recordings) or the heavy cognitive demands placed upon learners during verbal reports (e.g., "forgetting" often observed in retrospective interviews; Russo et al., 1989).

Unsatisfied with the existing strategy taxonomies, other studies, such as Du (2018), Flowerdew and Miller (1992), Fung and Macaro (2019), and Schroeder (2016),

strove to inductively identify new ways to categorize listening strategies. Such an inductive approach is important to capture a complex, dynamic state of strategic behaviors in a specific task and context (Gao, 2007; Rose et al., 2018). On the one hand, the actual attempts often had problems with a lack of theoretical underpinnings of the newly found strategies and taxonomies (e.g., Du, 2018; Flowerdew & Miller, 1992; Schroeder, 2016). On the other hand, Fung and Macaro (2019) presented a case in which listening strategies were clearly defined and grouped into mental actions with similar goals and intentions. Using a questionnaire which had been validated with learners with similar study backgrounds, the researchers were able to identify listening strategies most relevant to secondary school learners in Hong Kong. This was also one of only a few studies that explicitly stated the target population. It should be noted, however, that Fung and Macaro's (2019) conceptualization of listening strategies still lacks an explanation as to how each strategy relates to models of comprehension (cf. Hosogoshi, 2016; O'Malley et al., 1989; Rukthong & Brunfaut, 2020) or knowledge sources (cf. Cai & Lee, 2012, 2010; Lee & Cai, 2010), the issues common to many of the other included studies.

5.1.3 Second in-depth review question: What evidence is available for the effectiveness of listening strategies in second language lecture comprehension?

In general, the reviewed studies suggest that effective L2 listeners use metacognitive strategies and cognitive strategies that activate their prior knowledge more effectively than less effective L2 listeners. Another observed tendency was the use of strategies, such as translation, fixation, and reconstruction, by less effective L2 listeners. However, the research to date has not focused on the effective orchestration or combination of strategies to understand lectures. In general, previous studies have not been able to demonstrate high-quality evidence due to the failure to achieve high trustworthiness of the findings and/or low relevance of the listening activities employed in the studies. Furthermore, no studies used the same listening proficiency/performance measures or included authentic listening activities. These limitations make the findings less generalizable to listening strategies in L2 lecture comprehension.

Another issue with the existing studies concerns the neglect of the relationship between listening strategies and levels of comprehension. For example, no efforts

have been made to incorporate the Construction-Integration model (Kintsch, 1988, 1998) into listening performance assessment. One study by Cai and Lee (2012) rated the participants' recall of the listening text by counting the number of propositions according to Kintsch (1974). This approach, however, did not consider the depth of comprehension required to process each propositional statement. Likewise, Chou's (2015) claims concerning the most and least useful listening strategies in EGAP and ESAP courses could have been more convincing if the findings were situated in a model of comprehension.

Finally, it has not been clarified to what extent the effectiveness of listening strategies is dependent on linguistic knowledge or prior knowledge. Only one study by Du (2018) investigated the effectiveness of listening strategies in relation to linguistic knowledge. In Fung and Macaro (2019), participants were categorized into four groups according to their linguistic knowledge and how "strategic" they were based on the results of a strategy questionnaire. As with Du, the researchers reported that selective attention was used only by more strategic learners with higher linguistic knowledge, whereas translation was used by learners with lower linguistic knowledge. This study, however, included no data on listening performance and could not make claims about the effectiveness of strategies. In

both Du (2018) and Fung and Macaro (2019), linguistic knowledge was assessed by paper-based vocabulary and grammar tests. Although actual linguistic processing in listening involves auditory decoding of linguistic input, this has never been incorporated in operationalizing linguistic knowledge. Another important variable is prior knowledge, which was observed to have a significant impact on the strategy use and listening performance in Du's study. Nevertheless, the difficulty lies in controlling for prior knowledge. Although O'Bryan and Hegelheimer (2009) attempted to do this by choosing listening texts of general interest, participants' differential exposure to such topics may create more variability in prior knowledge.

Thus, the research to date has given some indications of the types of effective and ineffective listening strategies. Considering the small number of quality and relevant studies available in this area, the evidence is inconclusive in terms of the extent of effectiveness of listening strategies especially in association with levels of comprehension and linguistic/prior knowledge.

5.1.4 Third in-depth review question: What evidence is available for the development of listening strategies in second language lecture comprehension?

The in-depth review provides some evidence for the development of metacognitive listening strategies especially when explicit metacognitive strategy instruction is provided over a reasonable period of time (e.g., five weeks, 16 hours of instruction in total; Rahimirad & Moini, 2015). There are also indications of the development of other strategies, such as cognitive and socio-affective strategies (O'Malley, Chamot, Stewner-Manzanares, et al., 1985; Weinberg et al., 2011) as well as bottom-up and top-down strategies (Carrier, 2003) through the corresponding strategy instruction. These findings are in agreement with those obtained in listening strategy research in general (Chen, 2009; Graham et al., 2008, 2011; Yeldham & Gruba, 2014).

To date, however, there has been no reliable evidence available for the development of listening strategies in L2 lecture comprehension. The effectiveness of strategy instruction remains difficult to compare because researchers have used a variety of outcome measures for listening strategies or listening performance; only one study by O'Bryan and Hegelheimer (2009) offered direct evidence by collecting

data on listening strategies at multiple time points. O'Bryan and Hegelheimer was also the only study that extensively made use of both quantitative and qualitative data to arrive at their conclusions. Moreover, the review sheds light on study design issues, such as no explicit statements of the target population, small sample sizes, the unexplained random assignment procedures, inappropriate strategy instruction, no mentioning of the fidelity of the intervention, a lack of delayed posttests, and low ecological validity of listening activities. No studies have investigated this topic in a naturalistic setting. Therefore, it can be argued that the least attention has been paid to the development of listening strategies.

5.2 Limitations

One limitation of this review is the involvement of a collaborator only in the study selection phase. Other phases, such as data collection, quality assessment, and evidence synthesis, were conducted only by one researcher (the author of this review) due to the time constraints. However, efforts were made to increase the validity and reliability of procedures by clarifying the decision-making processes and having them checked by a content expert.

A second limitation concerns the systematicity of electronic and hand searching. In addition to electronic database searching using a predetermined search string, hand searching was conducted to ensure the comprehensiveness of this review. Two studies were included through backward citation searching; further attempts may have resulted in the identification of other relevant studies. This could also mean the limitation of the current search string, despite the attempt to be as exhaustive in searching as possible.

A third limitation relates to the eligibility criteria. The present review limited itself to the studies published in peer-reviewed English language journals with the intention to ensure the systematicity in searching and the quality of evidence. Therefore, the review does not represent all the research that has ever been conducted about listening strategies in L2 lecture comprehension. A more comprehensive understanding of this topic should consider the evidence from other sources including dissertations, book chapters, and conference papers as well as the studies written in non-English languages.

5.3 Conclusion

This review set out to explore the current state of research on listening strategies in L2 lecture comprehension. The 21 studies undertaken so far provide little generalizable evidence concerning this topic. Nevertheless, the review showcased important directions that may be taken up in future studies.

Firstly, there is a need for a joint effort to diversify research participants and contexts to reach a more comprehensive understanding of this topic. More research is clearly required to explore learners of different language, study, and cultural backgrounds in a variety of L2 medium instruction contexts, in addition to the current focus on learners of English from a narrow range of nationalities in the USA, China, Hong Kong, and Iran. Scholars should also bear in mind the impact of task and text characteristics on listeners' mental processes and control for these factors when investigating listening strategies in L2 lecture comprehension. In more general terms, future research will benefit from the efforts to increase the methodological quality and trustworthiness of their findings.

Secondly, the findings of this review illustrate the importance of clarifying the definitions and taxonomies of strategies in identifying listening strategies (Macaro, 2006; Macaro et al., 2007). Researchers may find ready-made strategy taxonomies, such as Goh (2002b) and Vandergrift (1996, 1997, 2003a), useful for this purpose. Another promising avenue is to operationalize and theorize listening strategies so that identified strategies are relevant to lecture comprehension of the target population, as showcased by Fung and Macaro (2019). In doing so, it is necessary to clarify the theoretical link between strategies and models of comprehension.

Thirdly, previous studies demonstrated broadly two ways to explore the effectiveness of listening strategies. One way is to employ a (quasi) experimental design to compare listening performances of groups who use and do not use listening strategies. Although a randomized controlled trial is “the gold standard” (Cartwright, 2010, p. 59), researchers should specify a method of allocation of participants. When a non-randomized controlled trial is chosen, efforts should be made to control for linguistic/prior knowledge. Another approach is to explore the relationship between listening strategies and linguistic/prior knowledge instead of controlling for (thereby diminishing the effect of) linguistic/prior knowledge. As instantiated by Du (2018) and Fung and Macaro (2019), one could conduct a cross-

sectional study or a quasi-experimental study where participants are divided into groups according to linguistic/prior knowledge, self-regulation, and listening proficiency.

Finally, more studies should either quantitatively or qualitatively collect data on listening strategies at multiple time points to provide direct evidence for or against the development of listening strategies. In addition to well-designed strategy instruction studies, a longitudinal investigation into strategy development in a naturalistic setting should also be informative in prioritizing the types of strategies and learners in strategy instruction.

References

An, S. (2013). Schema theory in reading. *Theory and Practice in Language Studies*, 3(1), 130–134. <https://doi.org/10.4304/tpls.3.1.130-134>

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge University Press.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). W H Freeman.

Anderson, K., & Lynch, T. (1988). *Listening*. Oxford University Press.

Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies*, 5(2), 272–281.
<http://hdl.handle.net/123456789/256>

Bacon, S. M. (1992a). Phases of listening to authentic input in Spanish: A descriptive study. *Foreign Language Annals*, 25(4), 317–333.
<https://doi.org/10.1111/j.1944-9720.1992.tb00552.x>

Bacon, S. M. (1992b). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *The Modern Language Journal*, 76(2), 160–178.
<https://doi.org/10.1111/j.1540-4781.1992.tb01096.x>

- Benson, M. J. (1994). Lecture listening in an ethnographic perspective. In J. Flowerdew (Ed.), *Academic listening: Research perspective* (pp. 181–198). Cambridge University Press.
- Bialystok, E. (1983). Some factors in the selection and implementation of communication strategies. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 100–118). Longman.
- Bidabadi, F. S., & Yamat, H. (2013). EFL learners' perceptions towards meta-cognitive strategy use in English language listening. *GEMA Online Journal of Language Studies*, 13(3), 31–43.
<http://ejournal.ukm.my/gema/article/view/4219/2336>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Bilbow, G. T. (1989). Towards an understanding of overseas students' difficulties in lectures: A phenomenographic approach. *Journal of Further and Higher Education*, 13(3), 85–99.
<https://www.tandfonline.com/doi/abs/10.1080/0309877890130308>
- Boland, A., Cherry, G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd ed.). Sage Publications. <https://uk.sagepub.com/en-gb/eur/doing-a-systematic-review/book251308>

- Boutron, I., Page, M. J., Higgins, J. P., Altman, D. G., Lundh, A., & Hróbjartsson, A. (2019). *Considering bias and conflicts of interest among the included studies* (J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (eds.)). Cochrane Handbook for Systematic Reviews of Interventions Version 6.0; Cochrane. <https://training.cochrane.org/handbook/current/chapter-07>
- Briggs, J. G. (2015). A context-specific research tool to probe the out-of-class vocabulary-related strategies of study-abroad learners. *International Journal of Applied Linguistics*, 25(3), 291–314. <https://doi.org/10.1111/ijal.12070>
- Bronfenbrenner, U. (1976). The experimental ecology of education. *Educational Researcher*, 5, 5–15. <https://doi.org/10.3102/0013189X005009005>
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Brown, G. (1995). Dimensions of difficulty in listening comprehension. In D. J. Mendelsohn & J. Rubin (Eds.), *A Guide for the Teaching of Second Language Listening* (pp. 59–73). Dominie Press.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8(1), 67–91. <https://doi.org/10.1177/026553229100800105>
- Byrnes, H. (1984). The role of listening comprehension: A theoretical base. *Foreign*

Language Annals, 17(4), 317–329. <https://doi.org/10.1111/j.1944-9720.1984.tb03235.x>

Cai, W., & Lee, B. P. H. (2012). Processing unfamiliar words: Strategies, knowledge sources, and the relationship to text and word comprehension. *Canadian Journal of Applied Linguistics / Revue Canadienne de Linguistique Appliquee*, 15(1), 122–145.

<https://search.proquest.com/docview/1033568066?accountid=13042>

Cai, W., & Lee, B. P. H. H. (2010). Investigating the effect of contextual clues on the processing of unfamiliar words in second language listening comprehension. *Australian Review of Applied Linguistics*, 33(2), 18.1-18.28.

<https://doi.org/10.2104/aral1018>

Call, M. E. (1985). Auditory short-term memory, listening comprehension, and the Input Hypothesis. *TESOL Quarterly*, 19(4), 765.

<https://doi.org/10.2307/3586675>

Carrell, P. L. (1983). Some issues in studying the role of schemata, or background knowledge, in second language comprehension. *Reading in a Foreign Language*, 1(2), 81–92. <http://nflrc.ill.hawaii.edu/rfl/PastIssues/rfl12carrell.pdf>

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, 18(3), 441–469. <https://doi.org/10.2307/3586714>

- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21(3), 461–478. <https://doi.org/10.2307/3586498>
- Carrier, K. A. (2003). Improving high school English language learners' second language listening through strategy instruction. *Bilingual Research Journal*, 27(3), 383–408. <https://doi.org/10.1080/15235882.2003.10162600>
- Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical Studies*, 147(1), 59–70. <https://doi.org/10.1007/s11098-009-9450-2>
- Chand, R. K. (2007). Same size doesn't fit all: Insights from research on listening skills at the University of the South Pacific (USP). *International Review of Research in Open and Distance Learning*, 8(3), 1–22. <https://doi.org/10.19173/irrodl.v8i3.383>
- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7, 113–127. [https://scholarspace.manoa.hawaii.edu/bitstream/10125/38555/1/Chaudron %20 Richards \(1985\)_WP4\(2\).pdf](https://scholarspace.manoa.hawaii.edu/bitstream/10125/38555/1/Chaudron%20&Richards(1985)_WP4(2).pdf)
- Chen, A. (2009). Listening strategy instruction: Exploring Taiwanese college students' strategy development. *Asian EFL Journal*, 11(2), 54–85.
- Chou, M. H. (2015). The influence of topics on listening strategy use for English for academic purposes. *English Language Teaching*, 8(2), 44–54.

<https://doi.org/10.5539/elt.v8n2p44>

Cochrane. (2014). *Data collection form for intervention reviews for RCTs and non-RCTs version 3*. Cochrane Developmental, Psychosocial and Learning Problems. <https://dplp.cochrane.org/data-extraction-forms>

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 34(34), 307–331.

<https://doi.org/10.1080/15434300701333129>

Cohen, A. D. (2007). Coming to terms with language learner strategies: Surveying the experts. In A. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 29–45). Oxford University Press.

<https://global.oup.com/academic/product/language-learner-strategies-30-years-of-research-and-practice-9780194422543?lang=en&cc=gb>

Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315833200>

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications Inc. <https://us.sagepub.com/en-us/nam/research-design/book255675>

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130.

https://doi.org/10.1207/s15430421tip3903_2

Crystal, D. (2012). *English as a global language* (2nd ed.). Cambridge University

Press. <https://doi.org/10.1017/CBO9781139196970>

Cutler, A., & Clifton, C. J. (1999). Comprehending spoken language: a blueprint of

the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of*

language (pp. 123–166). Oxford University Press.

Davidson, J. (2011). Effectiveness. In *Encyclopedia of evaluation* (p. 122). Sage

Publications, Inc. <https://dx.doi.org/10.4135/9781412950558>

DiPiro, J. T. (2009). Why do we still lecture? *American Journal of Pharmaceutical*

Education, 73(8), 137. <https://doi.org/10.5688/aj7308137>

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in

second language acquisition. In *The psychology of the language learner:*

Individual differences in second language acquisition. Routledge.

<https://doi.org/10.4324/9781410613349>

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language

learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second*

language acquisition (pp. 589–630). Blackwell Publishing Ltd. [https://ezproxy-](https://ezproxy-prd.bodleian.ox.ac.uk:2106/doi/book/10.1002/9780470756492)

[prd.bodleian.ox.ac.uk:2106/doi/book/10.1002/9780470756492](https://ezproxy-prd.bodleian.ox.ac.uk:2106/doi/book/10.1002/9780470756492)

Drost, E. A. (2011). Validity and reliability in social science research. *Education*

Research and Perspectives, 38(1), 105–124.

Du, F. (2018). Question type, language proficiency and listening strategy use: An exploratory study. *Asian EFL Journal*, 20(9), 193.

https://search.proquest.com/docview/2153600344?accountid=13042%0Ahttp://oxfordsfx.hosted.exlibrisgroup.com/oxford?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Allba&atitle=Question+Type%2C+Language+Proficiency+

Dudley-Evans, T. (1994). Variations in the discourse patterns favoured by different disciplines and their pedagogical implications. In J. Flowerdew (Ed.), *Academic listening research perspective* (pp. 146–158). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524612>

Duffy, G. G. (2002). The case for direct explanation of strategies. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 28–41). Guilford.

Dundar, Y., & Fleeman, N. (2019). Developing my search strategy. In A. Boland, G. Cherry, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (2nd ed., pp. 61–77). Sage Publications. <https://uk.sagepub.com/en-gb/eur/doing-a-systematic-review/book251308>

Egger, M., Zellweger-Zähner, T., Schneider, M., Junker, C., Lengeler, C., & Antes,

- G. (1997). Language bias in randomised controlled trials published in English and German. *Lancet*, 350(9074), 326–329. [https://doi.org/10.1016/S0140-6736\(97\)02419-7](https://doi.org/10.1016/S0140-6736(97)02419-7)
- Ehrman, M. E., Leaver, B. Lou, & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, 31(3), 313–330. [https://doi.org/10.1016/S0346-251X\(03\)00045-9](https://doi.org/10.1016/S0346-251X(03)00045-9)
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press. <https://global.oup.com/academic/product/the-study-of-second-language-acquisition-9780194422574?lang=en&cc=it>
- EPPI-Centre. (2019). *EEF main data extraction v 1.0 October 2019[Standard]*. EPPI-Centre.
- Færch, C., & Kasper, G. (1986). The role of comprehension in second-language learning. *Applied Linguistics*, 7(3), 257–274. <https://doi.org/10.1093/applin/7.3.257>
- Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal*, 52(2), 110–118. <https://doi.org/10.1093/elt/52.2.110>
- Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, 32(3), 363–377. <https://doi.org/10.1016/j.system.2004.05.002>

Field, J. (2008). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly*, 42(3), 411–432.

<https://doi.org/10.1002/j.1545-7249.2008.tb00139.x>

Field, J. (2009). *Listening in the language classroom*. Cambridge University Press.

<https://doi.org/10.1093/elt/ccq026>

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes*, 10(2), 102–112.

<https://doi.org/10.1016/j.jeap.2011.04.002>

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening* (pp. 77–151). Cambridge University Press.

<https://www.cambridgeenglish.org/research-and-validation/published-research/silt/>

Field, J. (2019). *Rethinking the second language listening test: From theory to practice*. Equinox Publishing Ltd.

<https://www.equinoxpub.com/home/rethinking-second-language/>

Fleeman, N., & Dunder, Y. (2017). Data extraction: Where do I begin? In A. Boland, G. Cherry, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (2nd ed., pp. 93–105). Sage Publications. <https://uk.sagepub.com/en-gb/eur/doing-a-systematic-review/book251308>

- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231–244. [https://doi.org/10.1016/0346-251X\(93\)90044-H](https://doi.org/10.1016/0346-251X(93)90044-H)
- Flowerdew, J. (1994). *Academic listening: Research perspective*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524612>
- Flowerdew, J., & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal*, 23(2), 60–80. <https://doi.org/10.1177/003368829202300205>
- Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16(1), 27–46. [https://doi.org/10.1016/S0889-4906\(96\)00030-0](https://doi.org/10.1016/S0889-4906(96)00030-0)
- Flowerdew, J., & Miller, L. (2010). Listening in a Second Language. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 158–177). Wiley-Blackwell. <https://doi.org/10.1002/9781444314908.ch1>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>
- French, S., & Kennedy, G. (2017). Reassessing the value of university lectures. *Teaching in Higher Education*, 22(6), 639–654. <https://doi.org/10.1080/13562517.2016.1273213>

Fung, D. (2016). *Comprehension strategies when listening to the teacher in the ESL classroom (Doctoral dissertation)*. University of Oxford.

Fung, D., & Macaro, E. (2019). Exploring the relationship between linguistic knowledge and strategy use in listening comprehension. *Language Teaching Research*, 00(0), 1–25. <https://doi.org/10.1177/1362168819868879>

Gao, X. (2007). Has language learning strategy research come to an end? A response to Tseng et al. (2006). *Applied Linguistics*, 28(4), 615–620. <https://doi.org/10.1093/applin/amm034>

Gehrke, P. J. (2018). Ecological validity. In *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 564–565).

Genesee, F. (2004). What do we know about bilingual education for majority language students. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of bilingualism and multiculturalism*. (pp. 547–576). Wiley-Blackwell. <https://www.wiley.com/en-gb/The+Handbook+of+Bilingualism+and+Multilingualism%2C+2nd+Edition-p-9781444334906>

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23(3), 265–304. <https://doi.org/10.1080/01638539709544994>

Goh, C. C. M. (1997). Metacognitive awareness and second language listeners.

- ELT Journal*, 51(4), 361–369. <https://doi.org/10.1093/elt/51.4.361>
- Goh, C. C. M. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, 2(2), 124–147. <https://doi.org/10.1177/136216889800200203>
- Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75. [https://doi.org/10.1016/S0346-251X\(99\)00060-3](https://doi.org/10.1016/S0346-251X(99)00060-3)
- Goh, C. C. M. (2002a). Exploring listening comprehension tactics and their interaction patterns. *System*, 30(2), 185–206. [https://doi.org/10.1016/S0346-251X\(02\)00004-0](https://doi.org/10.1016/S0346-251X(02)00004-0)
- Goh, C. C. M. (2002b). Learners' self-reports on comprehension and learning strategies for listening. *Asian Journal of English Language Teaching*, 12, 45–68. <https://www.airitilibrary.com/Publication/alDetailedMesh?docid=P20180321001-200212-201803260023-201803260023-45-68>
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4), 597–607. <https://nsuworks.nova.edu/tqr/vol8/iss4/6/>
- Gorard, S. (2014). A proposal for judging the trustworthiness of research findings.

Radical Statistics, 110, 47–59.

<http://www.radstats.org.uk/journal/journalissue110/>

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213–228.

<https://doi.org/10.1080/02671520701296189>

Gough, D., Oliver, S., & Thomas, J. (2012). Introducing systematic reviews. In D.

Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews*

(pp. 1–17). SAGE. [https://uk.sagepub.com/en-gb/eur/an-introduction-to-](https://uk.sagepub.com/en-gb/eur/an-introduction-to-systematic-reviews/book245742)

[systematic-reviews/book245742](https://uk.sagepub.com/en-gb/eur/an-introduction-to-systematic-reviews/book245742)

Graham, S. (1997). *Effective language learning: Positive strategies for advanced level language learning*. Multilingual Matters Ltd.

Graham, S. (2003). Learner strategies and advanced level listening comprehension.

Language Learning Journal, 28(1), 64–69.

<https://doi.org/10.1080/09571730385200221>

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*,

34(2), 165–182. <https://doi.org/10.1016/j.system.2005.11.001>

Graham, S. (2011). Self-efficacy and academic listening. *Journal of English for*

Academic Purposes, 10(2), 113–117.

<https://doi.org/10.1016/j.jeap.2011.04.001>

- Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching*, 50(1), 107–119.
<https://doi.org/10.1017/S0261444816000306>
- Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of french. *Language Learning*, 58(4), 747–783.
<https://doi.org/10.1111/j.1467-9922.2008.00478.x>
- Graham, S., & Santos, D. (2015). Strategies for second language listening: Current scenarios and improved pedagogy. In *Strategies for Second Language Listening: Current Scenarios and Improved Pedagogy*. Palgrave Macmillan.
<https://doi.org/10.1057/9781137410528>
- Graham, S., Santos, D., & Vanderplank, R. (2008). Listening comprehension and strategy use: A longitudinal exploration. *System*, 36(1), 52–68.
<https://doi.org/10.1016/j.system.2007.11.001>
- Graham, S., Santos, D., & Vanderplank, R. (2010). Strategy clusters and sources of knowledge in French L2 listening comprehension. *Innovation in Language Learning and Teaching*, 4(1), 1–20.
<https://doi.org/10.1080/17501220802385866>
- Graham, S., Santos, D., & Vanderplank, R. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching*

Research, 15(4), 435–456. <https://doi.org/10.1177/1362168811412026>

Grammer, J. K., Coffman, J. L., Ornstein, P. A., & Morrison, F. J. (2013). Change over time: Conducting longitudinal studies of children's cognitive development.

Journal of Cognition and Development, 14(4), 515–528.

<https://doi.org/10.1080/15248372.2013.833925>

Hansen, C., & Jensen, C. (1994). Evaluating lecture comprehension. In J.

Flowerdew (Ed.), *Academic listening: Research perspective* (pp. 241–268).

Cambridge University Press. <https://doi.org/10.1017/CBO9781139524612>

Hellekjær, G. O. (2017). Lecture comprehension in English-medium higher education. *HERMES - Journal of Language and Communication in Business*,

23(45), 11–34. <https://doi.org/10.7146/hjlc.v23i45.97343>

Higgins, J. P., Savović, J., Page, M. J., Elbers, R. G., & Sterne, J. A. (2019).

Assessing risk of bias in a randomized trial (J. P. Higgins, J. Thomas, J.

Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (eds.)). Cochrane

Handbook for Systematic Reviews of Interventions Version 6.0; Cochrane.

<https://training.cochrane.org/handbook/current/chapter-08>

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V.

(Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*

version 6.0. Cochrane. www.training.cochrane.org/handbook

- Hildyard, A., & Olson, D. R. (1982). On the comprehension and memory of oral versus written discourse. In D. Tannen (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 19–32). Ablex Publishing Corporation.
<https://blackwells.co.uk/bookshop/product/Spoken-and-Written-Language-Exploring-Orality-and-Literacy-by-Tannen-Deborah/9780893910945>
- Hollands, G. J., French, D. P., Griffin, S. J., Prevost, A. T., Sutton, S., King, S., & Marteau, T. M. (2016). The impact of communicating genetic risks of disease on riskreducing health behaviour: Systematic review with meta-analysis. *BMJ (Online)*, 352, 1–11. <https://doi.org/10.1136/bmj.i1102>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Horváthová, B. (2013). Tracing the development of listening strategies in the absence of explicit instruction. *XLinguae*, 6(3), 32–41.
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84879757704&partnerID=40&md5=0eb2dade26726967572321195087f1e8>
- Hosogoshi, K. (2016). Effects of captions and subtitles on the listening process:

- Insights from EFL learners' listening strategies. *JALT CALL Journal*, 12(3), 153–178. <https://eric.ed.gov/?id=EJ1125240>
- Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *International Journal of Phytoremediation*, 21(1), 413–425. <https://doi.org/10.1076/call.16.5.413.29488>
- Jeon, J. (2007). *A study of listening comprehension of academic lectures within the Construction-Integration Model (Doctoral dissertation)* [The Ohio State University]. https://etd.ohiolink.edu/!etd.send_file?accession=osu1196100462&disposition=inline
- Jin, B., & Xu, W. (2017). Strategy use awareness in academic listening practices relative to L2 motivation among Chinese tertiary students. *Journal of Language Teaching and Research*, 8(4), 722. <https://doi.org/10.17507/jltr.0804.11>
- Jones, S. E. (2007). Reflections on the lecture: Outmoded medium or instrument of inspiration? *Journal of Further and Higher Education*, 31(4), 397–406. <https://doi.org/10.1080/03098770701656816>
- Joyce, P. (2019). The relationship between L2 listening proficiency and L2 aural language processing. *Journal of Language Teaching and Learning in Thailand*, 57, 9–32. <https://files.eric.ed.gov/fulltext/EJ1224420.pdf>

- Jurkovič, V. (2010). Language learner strategies and linguistic competence as factors affecting achievement test scores in English for specific purposes. *TESOL Journal*, 1(4), 449–469. <https://doi.org/10.5054/tj.2010.234765>
- Kesmodel, U. S. (2018). Cross-sectional studies - what are they good for? *Acta Obstetricia et Gynecologica Scandinavica*, 97(4), 388–393. <https://doi.org/10.1111/aogs.13331>
- Kintsch, W. (1974). *The representation of meaning in memory*. Lawrence Erlbaum.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, 16(1–2), 193–202. <https://doi.org/10.1080/01638539309544837>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press. <https://psycnet.apa.org/record/1998-07128-000>
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kirkpatrick, A. (2014). The language(s) of HE: EMI and/or ELF and/or

multilingualism? *The Asian Journal of Applied Linguistics*, 1(1), 4–15.

<http://caes.hku.hk/ajal>

Kohler, D. B. (2002). *The effects of metacognitive language learning strategy training on lower-achieving second language learners (Doctoral dissertation)*

[Brigham Young University]. <https://psycnet.apa.org/record/2002-95021-063>

Lebauer, R. S. (1984). Using lecture transcripts in EAP lecture comprehension courses. *TESOL Quarterly*, 18(1), 41–54. <https://doi.org/10.2307/3586334>

Lee, B. P. H., & Cai, W. (2010). The effects of language proficiency on unfamiliar word processing in listening comprehension. *Hong Kong Journal of Applied Linguistics*, 12(2), 61–82.

Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M.-

I., Noel-Storr, A., Rader, T., Shokraneh, F., Thomas, J., & Wieland, L. S.

(2019). *Searching for and selecting studies* (J. P. Higgins, J. Thomas, J.

Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.)). Cochrane

Handbook for Systematic Reviews of Interventions Version 6.0; Cochrane.

<https://training.cochrane.org/handbook/current/chapter-04>

Li, T., Higgins, J. P., & Deeks, J. J. (2019). *Collecting data* (J. P. Higgins, J.

Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.)).

Cochrane Handbook for Systematic Reviews of Interventions Version 6.0.

<https://training.cochrane.org/handbook/current/chapter-05>

Liabo, K., Gough, D., & Harden, A. (2012). Developing justifiable evidence claims. In

D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews*

(pp. 251–278). SAGE. [https://uk.sagepub.com/en-gb/eur/an-introduction-to-](https://uk.sagepub.com/en-gb/eur/an-introduction-to-systematic-reviews/book245742)

[systematic-reviews/book245742](https://uk.sagepub.com/en-gb/eur/an-introduction-to-systematic-reviews/book245742)

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P.

A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA

statement for reporting systematic reviews and meta-analyses of studies that

evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical*

Research Ed.), 339. <https://doi.org/10.1136/bmj.b2700>

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Littlemore, J. (2001). The use of metaphor in university lectures and the problems

that it causes for overseas students. *Teaching in Higher Education*, 6(3), 333–

349. <https://doi.org/10.1080/13562510120061205>

Lucas, P. J., Baird, J., Arai, L., Law, C., & Roberts, H. M. (2007). Worked examples

of alternative methods for the synthesis of qualitative and quantitative research

in systematic reviews. *BMC Medical Research Methodology*, 7(4), 1–7.

<https://doi.org/10.1186/1471-2288-7-4>

Lund, R. J. (1991). A comparison of second language listening and reading

comprehension. *The Modern Language Journal*, 75(2), 196–204.

<https://doi.org/10.2307/328827>

Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10(2), 79–88.

<https://doi.org/10.1016/j.jeap.2011.03.001>

Lynch, T., & Mendelsohn, D. (2019). Listening. In N. Schmitt & M. P. H. Rodgers (Eds.), *An introduction to applied linguistics* (3rd ed., pp. 222–239). Taylor and Francis. <https://www.routledge.com/An-Introduction-to-Applied-Linguistics-3rd-Edition/Schmitt-Rodgers/p/book/9781138290136>

Ma, R., & Oxford, R. L. (2014). A diary study focusing on listening and speaking:

The evolving interaction of learning styles and learning strategies in a motivated, advanced ESL learner. *System*, 43(1), 101–113.

<https://doi.org/10.1016/j.system.2013.12.010>

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *Modern Language Journal*, 90(3), 320–337.

<https://doi.org/10.1111/j.1540-4781.2006.00425.x>

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1),

36–76. <https://doi.org/10.1017/S0261444817000350>

- Macaro, E., Graham, S., & Vanderplank, R. (2007). A review of listening strategies: focus on sources of knowledge and on success. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 165–186). Oxford University Press.
- <https://global.oup.com/academic/product/language-learner-strategies-30-years-of-research-and-practice-9780194422543?lang=en&cc=gb>
- Macaro, E., Vanderplank, R., & Graham, S. (2005). A systematic review of the role of prior knowledge in unidirectional listening comprehension. In *Research Evidence in Education Library* (Issue September).
- <http://eppi.ioe.ac.uk/%0Ahttp://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=0YJczAHkdBg=&tabid=299&mid=1149>
- Mareschal, P. M. (2007). How the west was won: an inside view of the seu's strategies and tactics for organizing home care workers in oregon. *International Journal of Organization Theory & Behavior*, 10(3), 386–412.
- <https://doi.org/10.1108/ijotb-10-03-2007-b006>
- McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models*. The MIT Press.
- <https://mitpress.mit.edu/books/parallel-distributed-processing-volume-1>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*,

22(3), 276–282. <https://doi.org/10.11613/bm.2012.031>

Miake-Lye, I. M., Hempel, S., Shanman, R., & Shekelle, P. G. (2016). What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Systematic Reviews*, 5(28), 1–21.

<https://doi.org/10.1186/s13643-016-0204-x>

Michael, Y. (2018). L2 listening instruction: More bottom-up or more top-down? *The Journal of AsiaTEFL*, 15(3), 805–810.

<https://doi.org/10.18823/asiatefl.2018.15.3.805>

Mizumoto, A., & Takeuchi, O. (2018). Modeling a Prototypical Use of Language Learning Strategies: Decision Tree-Based Methods in Multiple Context. In R. L. Oxford & C. M. Amerstorfer (Eds.), *Language Learning Strategies and Individual Learner Characteristics: Situating Strategy Use in Diverse Contexts* (pp. 99–122). Bloomsbury Academic.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

Morley, J. (1991). The pronunciation component in Teaching English to Speakers of Other Languages. *TESOL Quarterly*, 25(3), 481–520.

<https://doi.org/10.2307/3586981>

- Murphy, J. M. (1985). *An investigation into the listening strategies of ESL college students* (Doctoral dissertation). Colombia University.
- Murphy, V. A. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford University Press. <https://doi.org/10.1093/elt/ccu020>
- Nassaji, H. (2002). Schema Theory and Knowledge-Based Processes in Second Language Reading Comprehension: A Need for Alternative Perspectives. *Language Learning*, 52(2), 439–481. <https://doi.org/10.1111/0023-8333.00189>
- Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. In *Evidence-Based Nursing* (Vol. 18, Issue 2, pp. 34–35). BMJ Publishing Group. <https://doi.org/10.1136/eb-2015-102054>
- O'Bryan, A., & Hegelheimer, V. (2009). Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics*, 12(1), 9–38. <https://search.proquest.com/docview/216987857?accountid=13042>
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.
- O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418–437.

<https://doi.org/10.1093/applin/10.4.418>

O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Russo, R. P., Küpper, L., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly*, 19(3), 557–584.

<https://doi.org/10.2307/3586278>

O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L., & Russo, R. P. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning*, 35(1), 21–46. <https://doi.org/10.1111/j.1467-1770.1985.tb01013.x>

Okada, M., Mitchell, P., Finger, R. P., Eldem, B., Talks, S. J., Hirst, C., Paladini, L., Barratt, J., Wong, T. Y., & Loewenstein, A. (2020). Non-adherence or non-persistence to intravitreal injection therapy for neovascular age-related macular degeneration: A mixed-methods systematic review. *Ophthalmology*.

<https://doi.org/10.1016/j.ophtha.2020.07.060>

Olsen, L. A., & Huckin, T. H. (1990). Point-driven understanding in engineering lecture comprehension. *English for Specific Purposes*, 9(1), 33–47.

[https://doi.org/10.1016/0889-4906\(90\)90027-A](https://doi.org/10.1016/0889-4906(90)90027-A)

Osada, N. (2001). What strategy do less proficient learners employ in listening comprehension?: A reappraisal of bottom-up and top-down processing. *Journal*

of Pan-Pacific Association of Applied Linguistics, 5(1), 73–90.

<https://search.proquest.com/docview/85553545?accountid=13042%0Ahttp://oxf>

ordsfx.hosted.exlibrisgroup.com/oxford?url_ver=Z39.88-

2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%

<3Allba&atitle=What+Strategy+Do+Less+Proficient+Learner>

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Heinle.

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge.

[https://www.routledge.com/Teaching-and-Researching-Language-Learning-](https://www.routledge.com/Teaching-and-Researching-Language-Learning-Strategies-Self-Regulation-in/Oxford/p/book/9781138856806)

<Strategies-Self-Regulation-in/Oxford/p/book/9781138856806>

Oxford, R. L. (2018). Language learning strategies. In A. Burns & J. C. Richards (Eds.), *The cambridge guide to learning English as a second language* (pp. 81–90). Cambridge University Press.

[https://www.cambridge.org/gb/cambridgeenglish/catalog/teacher-training-](https://www.cambridge.org/gb/cambridgeenglish/catalog/teacher-training-development-and-research/cambridge-guide-learning-english-second-language)

<development-and-research/cambridge-guide-learning-english-second-language>

Oxford, R. L., Cho, Y., Leung, S., & Hae-Jin, K. (2004). Effect of the presence and difficulty of task on strategy use: An exploratory study. *International Review of Applied Linguistics in Language Teaching*, 42, 1–47.

<https://doi.org/10.1515/iral.2004.001>

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61(4), 993–1038.

<https://doi.org/10.1111/j.1467-9922.2011.00663.x>

PRISMA. (2015). *PRISMA flow diagram*. <http://prisma-statement.org/PRISMAStatement/FlowDiagram>

Putnick, D. L. (2018). Longitudinal versus cross-sectional design, analysis. In *The SAGE encyclopedia of lifespan human development* (pp. 1309–1311).

Rahimirad, M., & Moini, M. R. (2015). The challenges of listening to academic lectures for EAP learners and the impact of metacognition on academic lecture listening comprehension. *SAGE Open*, 5(2), 215824401559060.

<https://doi.org/10.1177/2158244015590609>

Rahimirad, M., & Zare-ee, A. (2015). Metacognitive strategy instruction as a means to improve listening self-efficacy among Iranian undergraduate learners of English. *International Journal of Instruction*, 8(1), 117–132.

<https://doi.org/10.12973/iji.2015.819a>

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17(2), 219–240. <https://doi.org/10.2307/3586651>

Richards, J. C. (2003). Current trends in teaching listening and speaking. *The*

Language Teacher, 27, 3–6.

http://www.finchpark.com/courses/tkt/Unit_07/teacher3.pdf

Robinson, K., Schmidt, T., & Teti, D. M. (2005). Issues in the use of longitudinal and cross-sectional designs. In *Handbook of research methods in developmental science* (pp. 3–20). Blackwell Publishing Ltd.

<https://doi.org/10.1002/9780470756676.ch1>

Rose, H., Briggs, J. G., Boggs, J. A., Sergio, L., & Ivanova-Slavianskaia, N. (2018).

A systematic review of language learner strategy research in the face of self-regulation. *System*, 72, 151–163. <https://doi.org/10.1016/j.system.2017.12.002>

Rost, M. (2002). *Teaching and researching listening*. Longman.

Rost, M., & Ross, S. (1991). Learner use of strategies in interaction: Typology and teachability. *Language Learning*, 41(2), 235–268.

<https://doi.org/10.1111/j.1467-1770.1991.tb00685.x>

Roussel, S. (2011). A computer assisted method to track listening strategies in second language learning. *ReCALL*, 23(2), 98–116.

<https://doi.org/10.1017/S0958344011000036>

Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers: A genre-based study. *English for Specific Purposes*, 21(1), 19–40.

[https://doi.org/10.1016/S0889-4906\(00\)00024-7](https://doi.org/10.1016/S0889-4906(00)00024-7)

- Rubin, J. (1975). What the “Good Language Learner” can teach us. *TESOL Quarterly*, 9(1), 41–51. <https://doi.org/10.2307/3586011>
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199–221. <https://doi.org/10.1111/j.1540-4781.1994.tb02034.x>
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53. <https://doi.org/10.1177/0265532219871470>
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, 17(6), 759–769. <https://doi.org/10.3758/BF03202637>
- Santos, D., Graham, S., & Vanderplank, R. (2008). Second language listening strategy research: Methodological challenges and perspectives. *Evaluation and Research in Education*, 21(2), 111–133. <https://doi.org/10.1080/09500790802152183>
- Schroeder, R. M. (2016). The development and validation of the “Academic Spoken English Strategies Survey (ASESS)” for non-native English speaking graduate students. *Journal of International Students*, 6(2), 394–414. <https://search.proquest.com/docview/1826533336?accountid=13042>

Schünemann, H., Brožek, J., Guyatt, G., & Oxman, A. (2013). *GRADE Handbook*.

<https://gdt.gradepro.org/app/handbook/handbook.html#h.kh98hjt91nl>

Seo, K. (2000). *Intervening in tertiary students' strategic listening in Japanese as a foreign language (Doctoral dissertation)*. Griffith University.

Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8(1), 23–40.

<https://doi.org/10.1177/026553229100800103>

Siegel, J. (2015a). *Exploring listening strategy instruction through action research*.

Palgrave Macmillan. <https://doi.org/10.1057/9781137521903>

Siegel, J. (2015b). Listening and how it is taught. In J. Siegel (Ed.), *Exploring listening strategy instruction through action research* (pp. 23–56). Palgrave

Macmillan. https://doi.org/10.1057/9781137521903_2

Skehan, P. (1989). *Individual differences in second-language learning*. Arnold.

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use.

Computer Assisted Language Learning, 17(5), 517–556.

<https://doi.org/10.1080/0958822042000319692>

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and

practice. *Journal of Development Effectiveness*, 4(3), 409–429.

<https://doi.org/10.1080/19439342.2012.710641>

Song, F., Hooper, L., & Loke, Y. K. (2013). Publication bias: what is it? how do we measure it? how do we avoid it? *Open Access Journal of Clinical Trials*, 5, 71–81.

https://ueaeprints.uea.ac.uk/id/eprint/43035/1/f_OAJCT_34419_publication_bias_what_is_it_how_do_we_measure_it_how_do_w_070313_16644.pdf

Sparks, R. L., & Ganschow, L. (1993). The impact of native language learning problems on foreign language learning: Case study illustrations of the Linguistic Coding Deficit Hypothesis. *The Modern Language Journal*, 77(1), 58–74.

<https://doi.org/10.2307/329559>

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16(1), 32–71. <https://doi.org/10.2307/747348>

Stanovich, K. E. (1984). The Interactive-Compensatory Model of Reading. *Remedial and Special Education*, 5(3), 11–19.

<https://doi.org/10.1177/074193258400500306>

Stern, H. H. (1975). What can we learn from the Good Language Learner? *Canadian Modern Language Review*, 31(4), 304–319.

<https://doi.org/10.3138/cmlr.31.4.304>

Stevick, E. W. (1990). Research on What? Some terminology. *The Modern*

Language Journal, 74(2), 143–153. <https://doi.org/10.2307/328118>

Swan, M., & Walter, C. (2017a). A response to Jonathan Newton. *ELT Journal*,

71(2), 245–246. <https://doi.org/10.1093/ELT/CCW095>

Swan, M., & Walter, C. (2017b). Misunderstanding comprehension. *ELT Journal*,

71(2), 228–236. <https://doi.org/10.1093/elt/ccw094>

Taguchi, N. (2001). L2 learners' strategic mental processes during a listening test.

JALT Journal, 23(2), 176–201.

Tauroza, S., & Allison, D. (1994). Expectation-driven understanding in information

systems lecture comprehension. In J. Flowerdew (Ed.), *Academic listening:*

Research perspective (pp. 35–54). Cambridge University Press.

<https://doi.org/10.1017/CBO9781139524612>

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative

research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 1–

10. <https://doi.org/10.1186/1471-2288-8-45>

Thomas, N., & Rose, H. (2019). Do Language Learning Strategies Need to Be Self-

Directed? Disentangling Strategies From Self-Regulated Learning. *TESOL*

Quarterly, 53(1), 248–257. <https://doi.org/10.1002/tesq.473>

- Thomas, N., Rose, H., & Pojanapunya, P. (2019). Conceptual issues in strategy research: Examining the roles of teachers and students in formal education settings. *Applied Linguistics Review*, 1(ahead-of-print), 1–17.
<https://doi.org/10.1515/applirev-2019-0033>
- Thompson, I., & Rubin, J. (1996). Can strategy instruction improve listening comprehension? *Foreign Language Annals*, 29(3), 331–342.
<https://doi.org/10.1111/j.1944-9720.1996.tb01246.x>
- Thompson, S. E. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes*, 2(1), 5–20. [https://doi.org/10.1016/S1475-1585\(02\)00036-X](https://doi.org/10.1016/S1475-1585(02)00036-X)
- Tseng, W. T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78–102. <https://doi.org/10.1093/applin/ami046>
- Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics*, 19(4), 432–451.
<https://doi.org/10.1093/applin/19.4.432>
- Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language*, 26(6), 638–657.
[https://doi.org/10.1016/0749-596X\(87\)90107-0](https://doi.org/10.1016/0749-596X(87)90107-0)

- Vandergrift, L. (1996). The listening comprehension strategies of core French high school students. *Canadian Modern Language Review*, 52(2), 200–223.
<https://doi.org/10.3138/cmlr.52.2.200>
- Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, 30(3), 387–409.
<https://doi.org/10.1111/j.1944-9720.1997.tb02362.x>
- Vandergrift, L. (1998). Successful and less successful listeners in French : What are the strategy differences? *The French Review*, 71(3), 370–395.
<https://www.jstor.org/stable/398969>
- Vandergrift, L. (2003a). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning*, 53(3), 463–496.
<https://doi.org/10.1111/1467-9922.00232>
- Vandergrift, L. (2003b, December 7). From prediction through reflection: Guiding students through the process of L2 listening. In *Canadian Modern Language Review* (Vol. 59, Issue 3, pp. 425–440). University of Toronto Press.
<https://doi.org/10.3138/cmlr.59.3.425>
- Vandergrift, L. (2004). Listening To Learn or Learning To Listen? *Annual Review of Applied Linguistics*, 24, 3–25. <https://doi.org/10.1017/s0267190504000017>
- Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive

awareness and proficiency in L2 listening. *Applied Linguistics*, 26(1), 70–89.

<https://doi.org/10.1093/applin/amh039>

Vandergrift, L. (2007). Recent developments in second and foreign language

listening comprehension research. *Language Teaching*, 40(3), 191–210.

<https://doi.org/10.1017/S0261444807004338>

Vandergrift, L. (2011). Second language listening: Presage, process, product, and

pedagogy. In E. Hinkel (Ed.), *Handbook of research in second language*

teaching and learning volume II (pp. 455–471). Taylor & Francis e-Library.

[https://www.routledge.com/Handbook-of-Research-in-Second-Language-](https://www.routledge.com/Handbook-of-Research-in-Second-Language-Teaching-and-Learning-Volume-2/Hinkel/p/book/9780415998727)

[Teaching-and-Learning-Volume-2/Hinkel/p/book/9780415998727](https://www.routledge.com/Handbook-of-Research-in-Second-Language-Teaching-and-Learning-Volume-2/Hinkel/p/book/9780415998727)

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language*

listening: Metacognition in action. Routledge.

<https://doi.org/10.4324/9780203843376>

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The

Metacognitive Awareness Listening Questionnaire: Development and

validation. *Language Learning*, 56(3), 431–462. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-9922.2006.00373.x)

[9922.2006.00373.x](https://doi.org/10.1111/j.1467-9922.2006.00373.x)

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen

does make a difference: An empirical study. *Language Learning*, 60(2), 470–

497. <https://doi.org/10.1111/j.1467-9922.2009.00559.x>

Vanderplank, R. (2013). Listening and understanding. In P. Driscoll, E. Macaro, & A. Swarbrick (Eds.), *Debates in modern languages education* (pp. 53–65).

Routledge. <https://doi.org/10.4324/9781315856551>

Vann, R. J., & Abraham, R. G. (1990). Strategies of unsuccessful language learners. *TESOL Quarterly*, 24(2), 177. <https://doi.org/10.2307/3586898>

Wagner, E. (2016). Authentic texts in the assessment of L2 listening ability. In J. Banerjee & D. Tsagari (Eds.), *Contemporary second language assessment: Contemporary applied linguistics volume 4* (pp. 103–124). Bloomsbury Academic.

Weinberg, A., Knoerr, H. H., Vandergrift, L., Knoerr, ne, & Vandergrift, L. (2011). Creating podcasts for academic listening in French: Student perceptions of enjoyment and usefulness. *CALICO Journal*, 28(3), 588–605. <https://doi.org/10.11139/cj.28.3.588-605>

Wingrove, P. (2017). How suitable are TED talks for academic listening? *Journal of English for Academic Purposes*, 30, 79–95. <https://doi.org/10.1016/j.jeap.2017.10.010>

Wolf, M. C., Muijselaar, M. M. L., Boonstra, A. M., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: Shared and

modality-specific components. *Reading and Writing*, 32(7), 1747–1767.

<https://doi.org/10.1007/s11145-018-9924-8>

Yeldham, M. (2019). Developing a framework for investigating L2 listeners'

longitudinal development. *IRAL - International Review of Applied Linguistics in*

Language Teaching, 57(2), 235–263. <https://doi.org/10.1515/iral-2017-0004>

Yeldham, M., & Gruba, P. (2014). The development of individual learners in an L2

listening strategies course. *Language Teaching Research*, 20(1), 9–34.

<https://doi.org/10.1177/1362168814541723>

Zare, J., & Keivanloo-Shahrestanaki, Z. (2017). The language of English academic

lectures: The case of field of study in highlighting importance. *Lingua*, 193, 36–

50. <https://doi.org/10.1016/j.lingua.2017.04.005>

Appendices

Appendix A: Review protocol

Protocol

Title

A systematic review of research on listening strategies for second language lecture comprehension

Registration

The protocol has not been registered in the existing databases of systematic reviews.

Authors

Atsushi Kanayama
A postgraduate student in MSc Applied Linguistics and Second Language Acquisition at the University of Oxford
Address: 4SS(Building)/08(Room), Ship Street, Ship Street Houses, Oxford, OX1 3DA
Mobile: 44 7721 379311
Email: jesu3578@ox.ac.uk

Contributions

AK is the main contributor to the protocol development and analyses.
ZH is a second reviewer and contributes to study selection.

Amendments

Important protocol amendments will be reflected in the protocol upon the consultation with a supervisor.

Support

No sources of support or funding were provided for this review.

INTRODUCTION

Rationale

The growing number of lectures are delivered in second language (English in particular; Flowerdew, 1994; Lynch, 2011). Language barriers have been claimed to be one of the main sources of difficulties for content acquisition in second language lecture contexts (Doiz et al., 2012). Among language skills, listening has been identified as the most difficult skill to learn due to its complex and implicit nature (Goh, 2018; Graham, 2017; Hulstijn, 2003). And yet, listening comprehension is a vital skill that is critical for academic success.

Listening to lectures has traditionally been distinguished from listening for interaction (e.g., Richards, 1983). While the latter multidirectional communication (to maintain relationships) is more commonly seen in many social domains of our life, the former unidirectional communication (to receive information) tends to be the way learners listen to learn in university lectures (Mendelsohn, 1998; Vandergrift, 2004).

Listening comprehension can be seen as 'a highly complex problem-solving activity' (Byrnes, 1984, p.318) rather than passive reciepience of speech code. Compared to conversations where multiple contextual cues are available, one-way communication such as lecture comprehension can be even more cognitively demanding on the part of the listener (O'Malley et al., 1985; Wipf, 1984). Therefore, strategies, or conscious and purposeful deployment of mental action (Macaro, 2006), may become of vital importance to facilitate comprehension processes and avoid the problems that may occur while listening (O'Malley et al., 1989).

Although empirical studies on listening strategies have been gaining popularity, especially since the development of a new strategy elicitation tool (e.g., Graham, 2017; Vandergrift et al., 2006), the overall picture of listening strategy use in second language comprehension and its developmental patterns is not yet clear. Most reviews on academic listening (e.g., Lynch, 2011) or listening strategies (e.g., Graham, 2017) were not conducted in a systematic way. The only exception was Macaro et al. (2007), in which the evidence from 44 empirical studies on listening strategies was evaluated in depth, although the review did not focus on academic listening.

Objectives

To fill the gap in this field, the study adopts a systematic research methodology to synthesize previous literature on listening strategies in second language comprehension, to further understanding of what, how, why those strategies are used and developed, and provides insights for future research and practice. The present study seeks to answer the following questions:

1. What empirical research was carried out on listening strategies for second language lecture comprehension between 1975 and 2020?
2. What evidence is available for listening strategies deployed in second language lecture comprehension?
3. What evidence is available for the effectiveness of listening strategies in second language lecture comprehension?
4. What evidence is available for how listening strategies are developed over time?

METHODS

Eligibility criteria

The review employs the following eligibility criteria:

- ☐ Publication: The studies were published in peer-reviewed English language journals.
- ☐ Year: The studies were published between 1975 and 2020.
- ☐ Topic: The studies provided empirical evidence on listening strategies (i.e., strategies to understand a spoken text) employed by foreign or second language learners.
- ☐ Task type: The spoken text was (hypothesized to be) unidirectional and academic, regardless of the location and timing of the actual listening (the actual listening can take place in and outside the classroom as well as during and outside the data-collection period).

Information sources

The search will be performed from 20th January 2020 using the following data sources:

- ☐ Electronic databases: Web of Science (Social Sciences Citation Index (SCCI) (1956-present), British Education Index (BEI), Education Resources Information Center (ERIC), Scopus, Language and Linguistics Behavioural Abstracts (LLBA), PsycINFO, and Google Scholar
- ☐ Journals: System, TESOL Quarterly, The Modern Language Journal, Language Teaching, Applied Linguistics, ELT Journal, Studies in Second Language Acquisition
- ☐ Forward (via Scopus, Web of Science, and Google scholar) and backward citations (via reference lists)
- ☐ Contact with study authors

Search strategy

The search strategy will include search terms that are centered around four key themes concerning 'listening' 'strategies' for 'content learning' in 'second language'

Search terms			
Listening	Strategy	Content learning	Second language
title and abstract	title and abstract	title and abstract	anywhere in the document
"listen**	AND	"strateg**	AND
		"comprehen**	"second language"
		"understand**	"foreign language"
		"lecture**	"additional language"
		"classroom**	"L2"
		"medium instruction"	"ESL"
		"medium of instruction"	"EFL"
		"medium for instruction"	"EAL"
		"medium education"	
		"medium of English"	
		"English medium"	
		"EMI"	
		"English as a lingua franca"	
		"English taught program**"	
		"parallel language education"	
		"parallel-language education"	
		"content and language integrated"	
		"content-and-language integrated"	
		"CLIL"	
		"content and foreign language integrated"	
		"content learning"	
		"content based**"	
		"content-based**"	
		"CBI"	
		"bilingual education"	
		"immersion"	
		"academi**"	
		"EAP"	
		"English for specific purposes"	
		"ESP"	

Study records**Data management**

Study records will be managed through a reference management software called Mendeley.

Selection process

The author will assess identified studies against the eligibility criteria (i.e., first screening) for title and abstract as well as full text screening. A collaborator screens 10% of the articles the author screened (i.e., second screening) to check the reliability of the first screening. As for title and abstract screening, the author will perform the first screening via Rayyan (Ouzzani et al., 2016) from 6th March 2020 to 15th March 2020. Then, the second screening will take place from 1st July 2020 to 31st July 2020. Only the studies that clearly did not meet the eligibility criteria were excluded. The first full text screening will be conducted by the author from 9th May 2020 to 30th June 2020. Afterward, the second screening will be conducted from 1st July 2020 to 31st July 2020.

Data collection process

Data will be collected using a data extraction table developed specifically for this review. The author will extract data from eligible articles from 4th June 2020 to 31st June 2020.

Data items

The following data items are extracted from each study:

- Study information
- Topic
- Purpose
- Relevant definitions and theoretical frameworks
- Research questions
- Research methods and designs
- Sampling strategy
- Research context
- Actual sample
- Data collection method
- Data analysis method
- Results
- Conclusions
- Implications for teaching, learning, and future research
- Limitations

Quality assessment of individual studies

The author will evaluate each study according to two critical appraisal tools from 4th June 2020 to 31st June 2020. The overall quality of evidence will be evaluated by Weight of Evidence Framework (Gough, 2007). The Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018) will enable more detailed assessment of the methodological quality of individual studies despite the heterogeneity of the study designs.

Data

The results of overall study characteristics (i.e., study characteristics, study design characteristics, methodological characteristics, definitions, and theoretical frameworks) will

be presented quantitatively in the form of an evidence map, using descriptions and visuals such as tables and graphs (Gough et al., 2012). The results of the in-depth analysis will be reported using a narrative approach (e.g., Snilstveit et al., 2012).

Synthesis

A narrative (i.e., qualitative) synthesis, involving a commentary reporting of each study as well as differences and similarities across studies (Lucas et al., 2007), is undertaken to synthesize the heterogeneity of the studies included in this review (Boland et al., 2017).

References

- Boland, A., Cherry, G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd ed.). Sage Publications. <https://uk.sagepub.com/en-gb/eur/doing-a-systematic-review/book251308>
- Byrnes, H. (1984). The Role of Listening Comprehension: A Theoretical Base. *Foreign Language Annals*, 17(4), 317–329. <https://doi.org/10.1111/j.1944-9720.1984.tb03235.x>
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2012). Future challenges for english-medium instruction at the tertiary level. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-Medium Instruction At Universities: Global Challenges* (pp. 213–221). Multilingual Matters. <https://doi.org/10.21832/9781847698162-015>
- Flowerdew, J. (1994). *Academic listening: Research perspective*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524612>
- Goh, C. C. M. (2018). Metacognition in Second Language Listening. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118784235.eelt0572>
- Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22(2), 213–228. <https://doi.org/10.1080/02671520701296189>
- Gough, D., Thomas, J., & Oliver, S. (2012). Clarifying differences between review designs and methods. *Systematic Reviews*, 1(1), 1–9. <https://doi.org/10.1186/2046-4053-1-28>
- Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting. *Language Teaching*, 50(1), 107–119. <https://doi.org/10.1017/S0261444816000306>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Hulstijn, J. H. (2003). Connectionist Models of Language Processing and the Training of Listening Skills With the Aid of Multimedia Software. *International Journal of Phytoremediation*, 21(1), 413–425. <https://doi.org/10.1076/call.16.5.413.29488>
- Lucas, P. J., Baird, J., Arai, L., Law, C., & Roberts, H. M. (2007). Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Medical Research Methodology*, 7(4), 1–7. <https://doi.org/10.1186/1471-2288-7-4>
- Lynch, T. (2011). Academic listening in the 21st century: Reviewing a decade of research. *Journal of English for Academic Purposes*, 10(2), 79–88. <https://doi.org/10.1016/j.jeap.2011.03.001>
- Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *Modern Language Journal*, 90(3), 320–337. <https://doi.org/10.1111/j.1540-4781.2006.00425.x>
- Macaro, E., Graham, S., & Vanderplank, R. (2007). A review of listening strategies: focus on sources of knowledge and on success. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies* (pp. 165–186). Oxford University Press. <https://global.oup.com/academic/product/language-learner-strategies-30-years-of-research-and-practice-9780194422543?lang=en&cc=gb>
- Mendelsohn, D. J. (1998). Teaching Listening. *Annual Review of Applied Linguistics*, 18, 81–101. <https://doi.org/10.1017/S0267190500003494>
- O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418–437. <https://doi.org/10.1093/applin/10.4.418>
- O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Kupper, L., & Russo, R. P. (1985). Learning Strategies Used By Beginning and Intermediate Esl Students. *Language Learning*, 35(1), 21–46. <https://doi.org/10.1111/j.1467-1770.1985.tb01013.x>

- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 1–10. <https://doi.org/10.1186/s13643-016-0384-4>
- Richards, J. C. (1983). Listening Comprehension: Approach, Design, Procedure. *TESOL Quarterly*, 17(2), 219–240. <https://doi.org/10.2307/3586651>
- Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness*, 4(3), 409–429. <https://doi.org/10.1080/19439342.2012.710641>
- Vandergrift, L. (2004). 1. Listening To Learn or Learning To Listen? *Annual Review of Applied Linguistics*, 24, 3–25. <https://doi.org/10.1017/s0267190504000017>
- Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The Metacognitive Awareness Listening Questionnaire: Development and Validation. *Language Learning*, 56(3), 431–462. <https://doi.org/10.1111/j.1467-9922.2006.00373.x>
- Wipf, J. A. (1984). Strategies for Teaching Second Language Listening Comprehension. *Foreign Language Annals*, 17(4), 345–348. <https://doi.org/10.1111/j.1944-9720.1984.tb03240.x>

Appendix B: Timeframe of this review

Section	Review stage	First reviewer	Second reviewer
3.2	Protocol	06/11/2019 - 17/06/2020	
3.3	Eligibility criteria	06/11/2019 - 05/03/2020	
3.5	Search strategy	20/01/2020 - 05/03/2020	
3.5	Electronic search	05/03/2020	
3.5	Hand search	06/03/2020 - 10/08/2020	
3.6.1	Title and abstract screening	06/03/2020 - 15/03/2020	01/07/2020 - 07/07/2020
3.6.2	Full text screening	09/05/2020 - 30/06/2020	07/07/2020 - 14/07/2020
3.7	Data collection tool	12/05/2020 - 17/06/2020	
3.7	Data collection	12/05/2020 - 20/07/2020	
3.9, 3.10	Risk of bias assessment	17/06/2020 - 20/07/2020	
3.11	Synthesis of results	21/07/2020 - 23/08/2020	

Appendix C: Search strategies

Database	Specific setting	Finalized Boolean search string	Result
Social Science Citation Index (available through Web of Science - all disciplines)	• Social Sciences Citation Index (SSCI) --1956-present • Articles • 1975-2020 • English	((TS=("listen") AND TS=("strateg") AND TS=("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP") AND TS=("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL"))))	160
British Education Index (BEI)	• Scholarly (Peer Reviewed) Journals • Publication Date: January 1975- December 2020 • Publication Type: Academic Journals • Document Type: Journal Article • Language: English • Educational Level: Higher Education • Age Level: All	((("listen") OR ab("listen")) AND ((("strateg") OR ab("strateg")) AND ((("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP") OR ab("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP")) AND ("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL"))	0
Education Resources Information Center (ERIC)	• Date: From 1975 to 2020 • Source type: Scholarly Journals • Document type: Article • Language: English • Exclude duplicate documents	((("listen") OR ab("listen")) AND ((("strateg") OR ab("strateg")) AND ((("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP") OR ab("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP")) AND ("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL"))	302
Scopus	• Subject area: Social Sciences • Document type: Article • Source type: Journals • Language: English • 1975-2020	TITLE-ABS-KEY ("listen") AND TITLE-ABS-KEY ("strateg") AND TITLE-ABS-KEY ("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP") AND ALL ("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL")	373
Language and Linguistics Behavioural Abstracts (LLBA)	• Full text • Peer reviewed • Source type: Scholarly Journals • Publication date: 1975-2020 • Document type: Article • Language: English	((("listen") OR ab("listen")) AND ((("strateg") OR ab("strateg")) AND ((("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP") OR ab("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP")) AND ("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL"))	291
PsycINFO - psychology	• Ovid Full Text Available • All journals • English language • 1975-2020	("listen" AND "strateg" AND ("comprehen" OR "understand" OR "lecture" OR "classroom" OR "medium instruction" OR "medium of instruction" OR "medium for instruction" OR "medium education" OR "medium of English" OR "English medium" OR "EMI" OR "English as a lingua franca" OR "English taught program" OR "parallel language education" OR "parallel-language education" OR "content and language integrated" OR "content-and-language integrated" OR "CLIL" OR "content and foreign language integrated" OR "content learning" OR "content based" OR "content-based" OR "CBI" OR "bilingual education" OR "immersion" OR "academi" OR "EAP" OR "English for specific purposes" OR "ESP")).ab,ti, and ("second language" OR "foreign language" OR "additional language" OR "L2" OR "ESL" OR "EFL" OR "EAL").af. limit 5 to (all journals and english language and yr=1975 - 2020)	90
		Total	1216

Appendix D: R script for generating random numbers using uniform distribution

Study selection	R script
Title and abstract screening	<pre>sample_titleandabstractscreening <- runif(n=68,min=1,max=680) #Drawing from the uniform distribution sample_titleandabstractscreening_integer <- round(sample_titleandabstractscreening) #Each number was rounded to the nearest integer sample_titleandabstractscreening_integer</pre>
Full text screening	<pre>sample_fulltextscreening <- runif(n=18,min=1,max=179) #Drawing from the uniform distribution sample_fulltextscreening_integer <- round(sample_fulltextscreening) #Each number was rounded to the nearest integer sample_fulltextscreening_integer</pre>

Appendix E: Excel sheet for full-text screening

Article	Report ID	Study ID	Criterion 1: the studies were published in peer-reviewed English language journals.	Criterion 2: the studies were published between 1975 and 2020.	Criterion 3: the studies provided empirical evidence on listening strategies (i.e., strategies to understand a spoken text) employed by L2 learners.	Criterion 3: further clarification 2: the study collected data on listening strategies 1: the study indicated (the use or the development of) listening strategies without collecting data on listening strategies (e.g., intervention studies) N/A: not applicable	Criterion 4: the spoken text was (hypothesized to be) unidirectional and academic, regardless of the location and timing of the actual listening. The actual listening can take place in and outside the classroom as well as during and outside the data-collection period.	Criterion 4: further clarification 2: the study explicitly and mainly investigates listening strategies in response to unidirectional and academic spoken text (i.e., listening to a lecturer) 1: the nature of the spoken text is unclear, although it can be inferred from the study design that (at least some) listening strategies are used to understand unidirectional and academic spoken text N/A: not applicable	Relevance to the criteria 4: high 3: medium 2: low	Eligible 1: eligible 0: not eligible
Ahour, T., & Sakhaei, S. R. (2015). The	1	1	1	1	0	N/A	N/A	N/A	N/A	0
Al-Ahwan, A., Asassfeh, S., & Al-Shboul	2	2	1	1	1	2	0	N/A	N/A	0
Al-Jahwari, M., Al Mekhlafi, A. M., Al-B	3	3	1	1	1	2	0	N/A	N/A	0
Alavi, S., & Mireskandari, N. (2015). Br	4	4	1	1	1	2	0	N/A	N/A	0
Alavinia, P., & Mollahosseini, H. (2012).	5	5	1	1	1	2	0	N/A	N/A	0
Amaluddin, Salasiah, A., & Mardiah. (2	6	6	1	1	1	2	0	N/A	N/A	0
Ardasheva, Y., Howell, P. B., & Vidrio M	7	7	1	1	0	N/A	N/A	N/A	N/A	0
Armiun, N., Rahmatian, R., Safa, P., &	8	8	1	1	1	2	0	N/A	N/A	0
Anyadoust, V. (2015). Application of Ev	9	9	1	1	1	2	0	N/A	N/A	0
BACON, S. M., & FINNESEMAN, M. D. (1	10	10	1	1	1	2	0	N/A	N/A	0
Baleghizadeh, S., & Hossein Rahimi, A.	11	11	1	1	1	2	0	1	3	0
Bekleyen, N. (2009). Helping teachers t	12	12	1	1	0	N/A	N/A	N/A	N/A	0
Bidabadi, F. S., & Yamat, H. (2013). Efi	13	13	1	1	1	2	1	2	4	1
Bork, W. J. (2000). Second language le	14	14	1	1	1	1	0	N/A	N/A	0
Bozorgian, H. (2014). Less-skilled learn	15	15	1	1	1	1	0	N/A	N/A	0
Bozorgian, H., & Pilla, H. (2013). Enhanc	16	16	1	1	1	1	0	N/A	N/A	0
Cai, W., & Lee, B. (2012). Processing U	17	17	1	1	1	2	1	2	4	1
Cai, W., & Lee, B. (2012). Processing U	18	18	1	1	1	2	1	2	4	1
Carrier, K. A. (2003). Improving high scl	19	19	1	1	1	1	1	2	3	1
Chand, R. K. (2007). Same size doesn't	20	20	1	1	1	2	1	2	4	1
Chang, C., & Chang, C.-K. (2014). Devel	21	21	1	1	1	2	0	N/A	N/A	0
Chen, A. (2009). Listening strategy inst	22	22	1	1	1	2	0	1	3	0
Chen, A. H. (2010). Effects of listening	23	23	1	1	1	2	0	1	3	0
Chen, K. T., Lee, I. Y. N., & Lin, C. Y. (20	24	24	1	1	1	2	0	1	3	0
Chen, L., Zhang, R., & Liu, C. (2014). Lis	25	25	1	1	1	2	0	1	3	0
Chen, Y. (2005). A Qualitative Study on	26	26	1	1	1	2	0	N/A	N/A	0
Chen, Y. (2005). Barriers to Acquiring Li	27	27	1	1	1	2	0	N/A	N/A	0
Chien, C. N., & Li, W. (1998). The strate	28	28	1	1	1	2	0	N/A	N/A	0
Chien, C. N., & van Heyst, N. (2014). Ot	29	29	1	1	1	2	0	N/A	N/A	0
Chien, C.-N., & Wei, L. (1997). Developi	30	30	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0
Ching-Shyang Chang, A. (2007). The Im	31	31	1	1	1	2	0	N/A	N/A	0
Chou, M. H. (2015). The influence of to	32	32	1	1	1	2	1	2	4	1
Chou, M. H. (2016). Strategy Use for Lis	33	33	1	1	1	2	0	1	3	0
Chou, M. H. (2017). A Task-based Lang	34	34	1	1	1	2	0	N/A	N/A	0
Chou, M. H. (2017). English language e	35	35	1	1	1	2	0	N/A	N/A	0
Chou, M. H. (2019). Predicting self-effici	36	36	1	1	1	2	0	1	3	0
Chow, B. W. Y., Chiu, H. T., & Wong, S.	37	37	1	1	1	1	0	1	2	0
Cubillos, J. H., Chieffo, L., & Fan, C. (201	38	38	1	1	1	2	0	N/A	N/A	0
Cubillo, J., & Winke, P. (2013). Redefini	39	39	1	1	0	N/A	N/A	N/A	N/A	0
Danah, M. (2016). Enhancing Listening	40	40	1	1	1	2	0	N/A	N/A	0
de la Peña, A. M. M., & Soler, L. R. (1	41	41	1	1	0	N/A	N/A	N/A	N/A	0
Dong, J. (2016). A dynamic systems the	42	42	1	1	1	2	0	N/A	N/A	0
Dousti, H., & Abolfathi, H. (2013). Tr	43	43	1	1	1	2	0	N/A	N/A	0
Du, F. (2018). Question Type, Language	44	44	1	1	1	2	1	2	4	1
Dunkel, P., & Davy, S. (1989). The heuri	45	45	1	1	0	N/A	N/A	N/A	N/A	0
Fahim, M., & Fakri Alamdari, E. (2014	46	46	1	1	1	2	0	1	3	0
Fathi, J., & Homidzadeh, R. (2019). Th	47	47	1	1	1	2	0	1	3	0
Fernández-Toro, M. (2005). The role of	48	48	1	1	1	2	0	N/A	N/A	0
Field, J. (2011). Into the mind of the ac	49	49	1	1	1	1	0	N/A	N/A	0
Field, J. (2011). 2	N/A	50	1	1	0	N/A	N/A	N/A	N/A	0
Field, J. (2011). 3	N/A	51	1	1	0	N/A	N/A	N/A	N/A	0
Flowerdew, J., & Miller, L. (1992). Stud	50	52	1	1	1	2	1	2	4	1

Note. Full-text screening of the first 50 articles

Appendix F: Data items

Component	Question (or information category)
General information	Reviewer
Study information	Date
	Report ID
	Study ID
	Author
	Title
	Year
	Journal
	Database
Abstract	Does the abstract provide sufficient information to understand the study (Yes, Partly, No, N/A)?
Introduction	Did the study report the topic of the study clearly (Yes, Partly, No, N/A)? Please specify.
	Did the study explain why it was carried out (Yes, Partly, No, N/A)? Please specify.
Literature review	Did the study define listening strategies (Yes, Partly, No, N/A)? Please specify.
	Did the study report the theoretical frameworks to understand listening strategies (Yes, Partly, No, N/A)? Please specify.
	Did the study report any other core constructs (e.g., strategies, comprehension; Yes, Partly, No, N/A)? Please specify.
	Did the study report any other core models, theories, frameworks (e.g., strategies, comprehension; Yes, Partly, No, N/A)? Please specify.
	Did the study report RQs clearly (Yes, Partly, No, N/A)? Please specify.
Research design	What was the methodology used (Hong et al., 2018, p.8)?
	• Qualitative study • Quantitative randomized controlled trial • Quantitative non-randomized study • Quantitative descriptive study • Mixed methods study
	Was the study cross-sectional or longitudinal? Please specify (e.g., the length of the study).
	Did the study report variables or concepts and their relationships (e.g., dependent/independent/control) the study aimed to measure or examine (Yes, Partly, No, N/A)? Please specify (did the study control for linguistic knowledge and/or prior knowledge?).
	Did the study report how participants were assigned to groups (Yes, No, N/A)? Please specify (e.g., the number of treatment groups, whether there was a control group; see EPPI-Centre, 2019 for more details).
	• Random • Non-random but matched • Non-random not matched prior to treatment • Not assigned (Retrospective Quasi Experimental Design or Regression discontinuity) • N/A (e.g., there is only one treatment group)
	Were there other important features of the research design (Yes, Partly, No)? Please specify (e.g., details of the study design).
Sampling strategy	Did the study report the population (Yes, Partly, No, N/A)? Please specify.
	Did the study report the planned sample (Yes, Partly, No, N/A)? Please specify.
	Did the study report the planned sample size (Yes, Partly, No, N/A)? Please specify.
	Did the study report a sampling strategy (Yes, Partly, No, N/A)? Please specify the strategy and the reason for choosing the sample.
	• Random sampling • Purposeful sampling • Convenience sampling
	Did the study report the actual sample and sample size (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' ages (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' genders (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' L1 (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' L2 (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' L2 proficiency (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' linguistic knowledge (i.e., vocabulary and grammatical knowledge) in their L2 (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' discipline/major (Yes, Partly, No, N/A)? Please specify.
	Did the study report the participants' educational level (Yes, Partly, No, N/A)? Please specify.
	Did the study report the country/countries where the study was carried out (Yes, Partly, No, N/A)? Please specify.
Data collection	Were there other important features of the participants/the research context (Yes, Partly, No)? Please specify.
	Did the study report data collection instruments (Yes, Partly, No, N/A)? Please specify.
	Did the study report the data collection procedure (Yes, Partly, No, N/A)? Please specify.
	Did the study report the setting where data were collected? (Yes, Partly, No, N/A)? Please specify.
	Did the study report any ways they have addressed the validity and reliability of their data collection tools/methods (Yes, Partly, No, N/A)? Please specify.
	Were listening strategies operationalized (Yes, Partly, No, N/A)? Please specify.
Data analysis	Was the spoken text unidirectional and academic (Yes, Partly, No, N/A)? Please specify.
	Were there other important features of the data collection (Yes, Partly, No)? Please specify.
	Did the study report how data were analyzed (Yes, Partly, No, N/A)? Please specify.
	Did the study report any ways they have addressed the validity and reliability of data analysis (Yes, Partly, No, N/A)? Please specify.
	Did the study report any ways to control for confounding variables (Yes, Partly, No, N/A)? Please specify.
Results, findings, discussion, and conclusion	Were there other important features of the data analysis (Yes, Partly, No)? Please specify.
	What did the study report the actual results were?
	What did the study conclude about the findings of the study?
	Did their conclusions match your assessment of the findings/results (Yes, Partly, No, N/A)? Please specify.
	Did the study report implications for teaching and learning (Yes, Partly, No)? Please specify.
	Did the implications match the study findings (Yes, Partly, No, N/A)? Please specify.
	Did the study report suggestions for further research (Yes, Partly, No, N/A)? Please specify.
	Did the study report implications for theory-building in listening strategies (Yes, Partly, No, N/A)? Please specify.
	Did the study report limitations of the study (Yes, Partly, No, N/A)? Please specify.

Appendix G: Evaluation of the quality of quantitative and qualitative findings

Quantitative validity

Validity in a quantitative study concerns the meaningfulness of study components (Drost, 2011), which influences whether meaningful inferences can be drawn from the data about the population of interest (Creswell & Creswell, 2017). The current review assessed each study according to the four types of validity suggested by Drost (2011; Table 31).

Table 31 Four quantitative validity types and signaling questions

Type	Component	Question
<i>Construct validity</i> refers to how well indicators (e.g., questionnaire items) measure constructs or concepts (Creswell & Creswell, 2017).	Data collection	Did the study report or indicate how it ensured that indicators represented constructs or concepts (from a subjective point of view; face validity; Drost, 2011)? If not, were there any other reasons to believe that they were so?
<i>Internal validity</i> involves the extent to which the relationships between variables (e.g., causality) were	Research design	Did the study report or indicate whether the results were influenced by one of the following threats to internal validity (Creswell & Creswell, 2017)?: the unrelated events that occurred during the experiment (i.e., history)

clarified (Creswell & Creswell, 2017; Lincoln & Guba, 1985).	• the maturation or change in participants during the experiment (i.e., maturation) • the inclusion of the participants who were known, before the experiment, to exhibit extreme scores (i.e., regression to the mean) • the systematic differences between the groups caused by the certain characteristics of the participants (i.e., selection) • dropout from participants (i.e., attrition) • the interaction between the participants from different groups (i.e., diffusion of treatment) • unequal benefits received by the groups (i.e., compensatory demoralization) • the participants' feeling undervalued due to the participation in the control group (i.e., compensatory rivalry) • the participants' becoming familiar with the test used in the experiment (i.e., testing) • the difference in the instruments used for a pretest and posttest (i.e., instrumentation)
<i>External validity</i> relates to the extent to which the results can be	Did the study report or indicate whether the generalizability of the results was influenced by one of the following threats

generalizable (Creswell & Creswell, 2017; Drost, 2011; Lincoln & Guba, 1985).	to external validity (Creswell & Creswell, 2017) • the narrow characteristics of the participants, which prevents the generalization of the results to individuals who do not have the characteristics (i.e., the interaction of selection and treatment, requiring additional experiments with groups with different characteristics) • the characteristics of the setting of the participants in an experiment, which prevents the generalization of the results to individuals in other settings (i.e., interaction of setting and treatment, requiring additional experiments in new settings) • the time-bound nature of the experiment (i.e., interaction of history and treatment, requiring further replication of the study at later times)
<i>Statistical conclusion validity</i> concerns the extent to which conclusions about the relationships are reasonable based on the given data (Creswell &	Sampling strategy Data collection Data analysis Did the study report or indicate any threats that may have lowered the statistical conclusion validity (Creswell & Creswell, 2017; Drost, 2011)?: • low statistical power (e.g., a sufficient sample size) • low reliability (e.g., measurements or observations, treatments or interventions)

Creswell, 2017;
Drost, 2011).

Quantitative reliability

Reliability can be defined as the extent to which measurements or instruments are repeatable over time and over a variety of conditions (Creswell & Creswell, 2017; Drost, 2011). Reliability involves equivalence (i.e., whether identical responses can be observed via alternative forms of a test), stability over time (i.e., whether identical responses can be observed via identical tests administered at different times), and consistency of a test (i.e., whether the same behavior or characteristic is measured by different items of a test; Drost, 2011). Table 32 shows the methods typically used to test the three aspects of reliability. Each included study was checked if it employed any of the methods to ensure the reliability of the test(s) or questionnaire(s).

Table 32 Methods to test reliability

Aspect	Method	Explanation
Equivalence	Alternative forms	Alternative forms of a test are used to calculate the correlation between them. A low correlation between alternative forms indicates that the sampling of items on the tests is not consistent.

Stability over time	Test-retest reliability	Test-retest reliability concerns the temporal stability of a test over time. This is checked by administering the same test to the same group of respondents at different times. However, this method has some limitations. The correlation between the two occasions may be influenced by the practice effect (caused by a short interval between tests), maturation (caused by a long interval between tests).
Internal consistency	Split-half approach	In the split-half approach, a test is split into half. The correlation between the two new measures is calculated and corrected to produce a reliability coefficient for the test. This method is more convenient and desirable than alternative forms or test-retest reliability, as it can be administered at the same time. Nevertheless, the correlation between the two halves may vary depending on how the items are split.
	Inter-rater reliability	Inter-rater reliability concerns the combined internal consistency of judgments made by different raters or judges.
	Cronbach alpha	Cronbach's alpha (α) is the most popular method of testing for internal consistency. The alpha value ranges between 0 and 1. α values between .7 and .9 are considered optimal.

Qualitative validity and reliability

Validity in a qualitative study can be described as whether the findings are accurate from the standpoint of the researcher, the participant, or the people external to the study, such as reviewers or readers (Creswell & Creswell, 2017; Creswell & Miller, 2000). Qualitative reliability, on the other hand, refers to the consistency of the process and findings of a study (Golafshani, 2003; Noble & Smith, 2015). Table 33 exhibits the strategies to enhance qualitative validity and reliability proposed by Anney (2014) and Creswell and Creswell (2017). The signaling questions were asked to evaluate whether each study employed those strategies to increase its validity and reliability.

Table 33 Strategies to increase qualitative validity and reliability and signaling questions

Strategy	Component	Question	Validity	Reliability
Combined strategies (Creswell & Creswell, 2017)	N/A	Did the study use multiple validity strategies to ensure the accuracy of the findings?	○	
<i>Triangulation</i> (Creswell & Creswell, 2017)	Data collection Data analysis Findings	Did the study combine evidence from different data sources to justify the themes?	○	

<i>Member checking</i> (Creswell & Creswell, 2017)	Data collection Data analysis Findings	Did the study take back the qualitative findings to the participants to see whether they feel that those findings were accurate?	○
<i>Rich, thick description</i> (Creswell & Creswell, 2017)	Findings	Did the study provide detailed descriptions of the setting or perspectives about a theme so that the findings become more realistic and richer for the reader?	○
<i>Researcher reflexivity</i> (Creswell & Creswell, 2017)	Data collection Data analysis Findings	Did the study report elements of self-reflection about the researchers' background (e.g., gender, culture, history, or socioeconomic origin) for the narrative to become open and honest that would resonate well with the reader?	○
<i>Disconfirming evidence</i> (Creswell & Creswell, 2017)	Findings	Did the study discuss information that contradicts the general perspectives of the theme so that the account becomes more realistic?	○

<i>Prolonged engagement in the field</i> (Creswell & Creswell, 2017)	Data collection	Did the researcher(s) of the study spend prolonged time in the field, develop an in-depth understanding of the phenomenon of interest, and account for detail about the site and the people under study?	○	
	Data analysis			
	Findings			
<i>Peer debriefing</i> (Creswell & Creswell, 2017)	Data collection	Did the study report the involvement of a peer debriefer who reviewed and asked questions about the study to ensure the account could resonate with other people?	○	○
	Data analysis			
	Findings			
<i>The audit trail</i> (Creswell & Creswell, 2017; Creswell & Miller, 2000)	Research design	Did the study report the involvement of an external auditor who is not familiar with the researcher or the project and assessed the quality of the study (e.g., the accuracy of transcription, the link between the research questions and the data, data analysis) during or at the end of the project? Did the study report whether the	○	○
	Sampling strategy			
	Data collection			
	Data analysis			
	Findings			

		researcher(s) kept a research log to increase the transparency of the data collection and analysis procedures?	
<i>Stepwise replication</i> (Anney, 2014)	Data analysis	Did the study report the involvement of two or more researchers to analyze the same data independently, compare the results, and resolved any inconsistencies?	○
<i>Code-Recode Strategy</i> (Anney, 2014)	Data analysis	Did the study explore whether the coding done by researchers (i.e., inter-rater or inter-observer coding) or the same researcher to see if the coding results were in agreement?	○
<i>Peer Examination</i> (Anney, 2014)	Data analysis	Did the study report whether the researcher(s) discussed their research process and findings with neutral colleagues (e.g., doctoral students)?	○

Validity and reliability of verbal protocols

The validity of verbal protocols concerns how accurately verbalization can reflect what they attended to while engaging in the task. Russo et al. (1989) argue that verbal protocols may have two validity concerns: reactivity (i.e., changes of the primary process caused by verbalization) and nonveridicality (i.e., an inaccurate reflection of the underlying primary process in verbalization). On the other hand, the reliability of verbal protocols relates to how data can be collected consistently. Table 34 summarizes some relevant issues to consider when evaluating the validity and reliability of verbal protocol procedures.

Table 34 Issues concerning the validity and reliability of verbal protocols

Validity	Issue	Question	Verbal protocol
Reactivity	Additional cognitive demand	Were there any reasons to believe that slack cognitive resources were available for the participants to be able to verbalize their thoughts without interrupting the primary process?	Think-aloud protocols Retrospective protocols
	Enhanced learning	Were there any reasons to believe that the additional opportunities of reflection led to the discovery, acquisition,	Think-aloud protocols Retrospective protocols

		or development of new or old strategies?	
	Motivational shift	Were there any reasons to believe that the participants' behavior reflected the perceived preferences of the experimenter or themselves in the face of public exposure?	Think-aloud protocols Retrospective protocols
	Auditory feedback	Were there any reasons to believe that the memory reinforcement (caused by auditory feedback) affected the participants' task performance?	Think-aloud protocols
Nonveridicality	Forgetting	Were there any reasons (e.g., a long interval between the task and verbalization) to believe that forgetting (i.e., memory effects) may have occurred?	Retrospective protocols
	Fabrication	Were there any reasons (e.g., cueing a retrospective protocol) to believe that the accounts provided by the participants were fabricated (i.e., reconstructed) and differed from the truth?	Retrospective protocols

Suggestions have been made to increase the validity and reliability of verbal protocols as follows (Field, 2011; Macaro et al., 2007):

- Pilot testing verbal protocol procedures to see how they affect the listening process
- Careful think-aloud training
- Improving inter-coder reliability
- Triangulation of a verbal protocol (e.g., think-aloud) with other data collection methods (e.g., retrospective interviews, questionnaires)
- the accurate use of two tape recorders for the listening text and verbalization
(In case of think-aloud protocols)
- Ensuring the immediate verbal report by asking the participants questions soon after the listening task

Appendix H: Further details of studies by task characteristics of listening activity

Study	Task characteristics	WoE C
Flowerdew & Miller (1992)	Listening to a lecturer/a teacher	Realistic
Weinberg et al. (2011)	Listening to a lecturer/a teacher	
Schroeder (2016)	Listening to a lecturer/a teacher	
Fung & Macaro (2019)	Listening to a lecturer/a teacher	
O'Malley et al. (1989)	Listening to audio recordings of academic lecture	Unrealistic
O'Bryan & Hegelheimer (2009)	Listening to audio recordings of academic lecture	
Rahimirad & Moini (2015)	Listening to audio recordings of academic lecture	
Rukthong & Brunfaut (2020)	Listening to audio recordings of academic lecture and summarizing the content	
Cai & Lee (2010)	Listening to audio recording and answering questions	
Lee & Cai (2010)	Listening to audio recording and answering questions	
Cai & Lee (2012)	Listening to audio recording and answering questions	
Chou (2015)	Listening to audio recording and answering questions	
Du (2018)	Listening to audio recording and answering questions	
Olsen & Huckin (1990)	Listening to audio recordings and summarizing the content	
O'Malley et al. (1985)	Listening to video recordings	
Smidt & Hegelheimer (2004)	Listening to video recordings and answering questions	
Hosogoshi (2016)	Listening to video recordings and answering questions	
Jin & Xu (2017)	Listening to video recordings and answering questions	
Carrier (2003)	Listening to audio and video recordings	
Chand (2007)	Listening to audio and video recordings	
Bidabadi & Yamat (2013)	A variety of listening activities (including listening to a lecturer)	

Note. Although Weinberg et al. (2011) provided strategy instruction using video

podcasts, the participants were expected to apply the knowledge learned through

the podcasts in real academic lectures.

Appendix I: Further details of studies by text characteristics of listening activity

Study	Text characteristics	WoE C
Flowerdew & Miller (1992)	Real lecture	Authentic
Weinberg et al. (2011)	Real lecture	
Schroeder (2016)	Real lecture	
Fung & Macaro (2019)	Real lecture	
O'Malley et al. (1985)	Long excerpt from real lecture	
Olsen & Huckin (1990)	Long excerpt from real lecture	
Smidt & Hegelheimer (2004)	Long excerpt from real lecture	
O'Bryan & Hegelheimer (2009)	Long excerpt from real lecture	
Hosogoshi (2016)	Long excerpt from real lecture	
Cai & Lee (2010)	Short and scripted text	Non-authentic
Lee & Cai (2010)	Short and scripted	
Cai & Lee (2012)	Short and scripted text	
Carrier (2003)	Non-academic text	
Bidabadi & Yamat (2013)	Non-academic text	
Chou (2015)	Non-academic text	
Chand (2007)	Scripted text	
O'Malley et al. (1989)	Unknown text	
Rahimirad & Moini (2015)	Unknown text	
Jin & Xu (2017)	Unknown text	
Du (2018)	Unknown text	
Rukthong & Brunfaut (2020)	Unknown text	

Note. O'Bryan and Hegelheimer (2009) included TED talks although they regarded them as academic lectures. The researchers also included a radio broadcast, whose text characteristics were “more scripted than an academic lecture” (p.17). However, three of the four podcasts included long excerpts from academic lectures.

Furthermore, text characteristics were judged to be unknown in case insufficient explanation was given (O'Malley et al., 1989) or listening tasks were taken from standardized tests (Du, 2018; Jin & Xu, 2017; Rahimirad & Moini, 2015; Rukthong & Brunfaut, 2020).

Appendix J: Summary of the studies related to the first in-depth review question

Overall WoE	Study	Observed strategies	Strategy definition	Strategy taxonomy	Comprehension model	Strategy elicitation
High	Flowerdew & Miller (1992)	• Reading • Peer help • Lecturer/tutor help • Attempts to concentrate harder • Marking the book and note-taking	N/A	N/A	N/A	• Observation • Interviews • Diaries
High	Fung & Macaro (2019)	Strategies • contextualization for the present lesson • selective attention on difficult words or segments • recall of prior knowledge • relational: understanding through recalling teacher's approach • summarization • translation • selective attention on simple words or segments • auditory representation and imagery • evaluation • repetition Opportunities for strategic behaviors • Utilization of personal physical resources • Hide and seek • Direct help seeking from the teacher	Mental actions (adopted from Briggs, 2015)	N/A	N/A	• Questionnaire
High	O'Bryan & Hegelheimer (2009)	<i>Vandergrift (1997)</i> Metacognitive strategies • Advance organization • Selective attention • Comprehension monitoring • Double-check monitoring • Evaluation • Problem identification Cognitive strategies • Inferencing • Linguistic inferencing • Extralinguistic inferencing • Between-parts inferencing • Elaboration • Personal elaboration • World elaboration • Creative elaboration • Summary • Repetition <i>Vandergrift et al. (2006)</i> Metacognitive strategies • Problem-solving • Planning and evaluation • Mental translation • Person knowledge • Directed attention	Steps or actions selected by learners to either improve the learning of an L2, the use of it, or both (adopted from Cohen, 1996a, p. 5)	• O'Malley & Chamot (1990) • Vandergrift (1997) • Vandergrift et al. (2006)	N/A	• Think-aloud protocols • Semi-structured interviews • MALQ

Overall WoE	Study	Observed strategies	Strategy definition	Strategy taxonomy	Comprehension model	Strategy elicitation
Medium	Schroeder (2016)	- Preparing for the Listening Task - Strategies for Keeping on Track while Listening - Strategies for Enhancing Understanding While Listening - Strategies for Clarifying Understanding after Listening - Strategies for Using Technologies to Help Listening	N/A	N/A	N/A	Questionnaire
Medium	Weinberg et al. (2011)	Content of podcasts - Listening to lectures: Applying the strategies of good listeners, - Preparing for the lecture before class and when arriving in class, - Monitoring and problem-solving during and after the lecture, - Best practices in note-taking, - The Cornell note-taking system, - Reviewing and evaluating the lecture, and - New technology and note-taking (wikis, semantic mapping, and Microsoft Office OneNote).	N/A	N/A	N/A	- Questionnaire - Focus group interview
Medium	Smidt & Hegelheimer (2004)	<i>Vandergrift (1996, 1997)</i> Metacognitive strategies - Planning - Monitoring - Evaluating - Problem identification Cognitive strategies - Inferencing - Elaboration - Summarization - Translation - Transfer - Repetition Socio-affective strategies - Questioning for clarification - Cooperation - Lowering anxiety - Self-encouragement - Taking emotional temperature	Deliberate, cognitive steps used by students to enhance comprehension, learning and retention of the target language (adopted from Vandergrift, 1996, p. 202)	Vandergrift (1996, 1997)	N/A	Retrospective interview
Medium	O'Malley et al. (1989)	- Selective attention - Directed attention - Self-monitoring - Elaboration - Inferencing	Mental processes that are activated in order to understand new information that is ambiguous or to learn or retain new information	- O'Malley et al. (1985) - Wenden & Rubin (1987) - Chamot & O'Malley (1987)	Anderson (1983, 1985)	Think-aloud protocols

Overall WoE	Study	Observed strategies	Strategy definition	Strategy taxonomy	Comprehension model	Strategy elicitation
Medium	Rukthong and Brunfaut (2020)	Goh (2002); O'Malley et al. (1989); Vandergrift (2003a); Vandergrift & Goh (2012) Metacognitive strategies • Preparing for listening • Selective attention • Directed attention • Monitoring comprehension • Real-time assessment of input • Comprehension evaluation Cognitive strategies • Inferencing • Elaboration • Prediction • Translation • Fixation • Reconstruction	Strategies are used to facilitate the listening process and fill gaps in understanding, and they are employed with a certain degree of consciousness (adopted from Vandergrift & Goh, 2012)	• Goh (2002) • O'Malley et al. (1989) • Vandergrift (2003a) • Vandergrift & Goh (2012)	• Anderson (1985) • Kintsch & van Dijk (1978) • Rost (2016) • Field (2013) • Vandergrift & Goh (2012)	• Stimulated recalls • Notes made while listening • Oral and written summaries
Medium	Lee & Cai (2010)	• Interencing strategy • Ignoring strategy • No attention	Effortful processing (adopted from Tomlin and Villa, 1994, p. 187)	• Hastrup (1991)	N/A	• Interviews (with or without retrieval cues)
Medium	Cai & Lee (2010)	• Interencing strategy • Ignoring strategy • No attention	Effortful processing (adopted from Tomlin and Villa, 1994, p. 187)	• Hastrup (1991)	N/A	• Interviews (with or without retrieval cues)
Medium	Cai & Lee (2012)	• Interencing strategy • Ignoring strategy • No attention	Effortful processing (adopted from Tomlin & Villa, 1994, p. 187)	• Hastrup (1991)	N/A	• Interviews (with or without retrieval cues)
Low	Chand (2007)	O'Malley et al. (1985) Metacognitive Strategies • Advance organiser • Directed attention • Comprehension and evaluation • Selective attention Cognitive Strategies • Translation • Seeking clarification • Repetition • Note-taking • Visualisation Social Mediation • Cooperation	N/A	• O'Malley et al. (1985)	N/A	• Semi-structured interviews

Overall WoE	Study	Observed strategies	Strategy definition	Strategy taxonomy	Comprehension model	Strategy elicitation
Low	Bidabadi & Yamat (2013)	<i>Vandergrift et al. (2006)</i> Metacognitive strategies • Planning and evaluation • Problem-solving • Personal knowledge • Directed knowledge	N/A	• Vandergrift et al. (2006)	N/A	• MALQ • Semi-structured interviews
Low	Rahimirad & Moini (2015)	<i>Vandergrift (2003a)</i> Metacognitive strategies • Planning • Directed attention • Evaluation • Monitoring • Selective attention	Effortful and conscious behaviors that play significant roles in facilitating listening comprehension and development (adopted from Chamot, 1995)	• Vandergrift (2003a)	N/A	• Semi-structured interviews
Low	Jin & Xu (2017)	Metacognitive strategies • Planning • Selective attention • Self-monitoring • Self-evaluation Cognitive strategies • Elaboration • Note-taking • Summarization • Inference Social-affective strategies • Self-encouragement • After-class interaction	N/A	• O'Malley & Chamot (1990) • Richards (1983)	N/A	• Questionnaire
Low	Du (2018)	• Match lexis heard to lexis in options • Elaboration • Hypothesis formation • Selective attention • Monitoring • Integration • Deduction • Real-time assessment of input • Fixation • Transfer • Avoidance strategy • Incompletion • Filtering • Approximation • Skipping	N/A	Goh (2002) Graham et al. (2008) Santos et al. (2008)	N/A	• Retrospective interviews • Unstructured interviews

Overall WoE	Study	Observed strategies	Strategy definition	Strategy taxonomy	Comprehension model	Strategy elicitation
Low	Chou (2015)	<i>Vandergrift (1997)</i> Metacognitive strategies • paying attention to the main points • paying attention to details • paying attention to understanding • making decisions as to whether the words heard were right in the context • skipping the unknown parts to complete the task • using the process of elimination. • changing my responses as appropriate Cognitive strategies • guessing the meaning of unknown words by linking them to known words • guessing by means of the tone of voice • using prior personal experience to comprehend the task • using my world knowledge to comprehend the task • using mental imagery to create a picture of what was happening. • making a mental summary of what I heard • translating what is heard in L2 to L1 • using knowledge about Chinese to facilitate listening to English • taking notes of key words or information as I followed the spoken texts <i>Cohen (2006)</i> Test wiseness strategies • answering the questions in chronological order in the listening. • choosing an option that seemed to deviate from the others, was special, was different, or conspicuous. • selecting the option because it appeared to have a word or phrase from the listening text – possibly a key word. • selecting the option because it had a word or phrase that appeared in the question.	N/A	• Vandergrift (1997) • Cohen (2006)	N/A	• Questionnaire • Retrospective interviews
Low	Hosogoshi (2016)	<i>Vandergrift (2003a)</i> Metacognitive strategies • Planning • Monitoring • Evaluation • Problem identification Cognitive strategies • Inferencing • Elaboration • Imagery • Summarization • Translation • Transfer • Repetition	Mental processes that are activated in order to understand new information that is ambiguous or to learn or retain new information (adopted from O'Malley, Chamot, and Kupper, 1989, p.422)	• Vandergrift (2003a)	• Anderson (1985)	• Questionnaire

Appendix K: Summary of the studies related to the second in-depth review question

Overall WoE	Study	Research design	Sample	Proficiency	Strategy instruction	Treatment	Outcome measure	Findings
Medium	O'Malley et al. (1985)	Quantitative randomized controlled trial	75 high school ESL students in the US	English proficiency • intermediate • determined by school	• 50 minutes; 10 days (the first and last days were spent for pre and posttests) • Metacognitive group: selective attention (metacognitive), note taking (cognitive, cooperation (social/affective)) • Cognitive group: note taking (cognitive, cooperation (social/affective))	Metacognitive (n = 27) Cognitive (n = 26) Control (n = 22)	• a short listening comprehension test with multiple-choice recognition items designed to assess Bloom and Krathwohl's (1977) knowledge, comprehension, and analysis levels	<i>Analysis of covariance</i> Pretest • the results were controlled for. Interim assessments • test 1 metacognitive > cognitive > control (p < .10) • test 2 cognitive > metacognitive > control (p < .05) • test 3 cognitive > metacognitive > control (p < .01) • test 4 metacognitive > cognitive > control (p = .626) Posttest • metacognitive > cognitive > control (p = .162)
Medium	O'Malley et al. (1989)	Qualitative study (qualitative description)	11 high school ESL students in the US	English proficiency • intermediate • determined by school	N/A	N/A	N/A	<i>Qualitative analysis</i> Perceptual processing • effective listeners were able to maintain and direct their attention to the listening task • ineffective listeners often failed to redirect their attention when their concentration is disrupted Parsing • effective listeners listened for larger chunks such as phrases or sentences and attended to intonation and pauses to use top-down processing. Bottom-up processing was used only out of necessity. • ineffective listeners attended to smaller units such as words and sentences and relied heavily on bottom-up processing. When the listeners found it difficult to parse the text, they used multiple strategies such as inferencing, translation, and self-monitoring to support their comprehension. Utilization • effective listeners used elaboration with reference to their world knowledge, personal knowledge, and self-questioning. These listeners also inferred the meaning of unfamiliar words and phrases using their prior knowledge. • ineffective listeners elaborated on the meaning of text on fewer occasions and did not associate the information with their personal experiences.
Medium	O'Bryan & Hegelheimer (2009)	Mixed methods study (sequential exploratory design) with grounded theory and cohort study	4 university ESL students in the US	Listening proficiency • high-intermediate, medium-intermediate, low-intermediate • determined by the listening scores of the English Placement Test (EPT)	• 15-week semester • classroom instruction: listening for lecture cues, taking notes, listening to numbers and statistics, and listening for additional indicators of academic lecture organization • 14 instructor-designed podcasts: either demonstrating or encouraging students to review and practice the listening strategies discussed in class • textbook: different types of classroom-based listening strategies	All participants	N/A	<i>Qualitative analysis</i> Higher proficiency students • used the least percentage of metacognitive strategies, indicating successful understanding of the listening text • applied between-parts inferencing and personal elaboration to fill in the gap in knowledge to understand the text Lower proficiency students • used inferencing and elaboration strategies to process information at the local (word and sentence) level
Medium	Rukthong & Brunfaut (2020)	Mixed methods study (convergent design) with grounded theory and survey	72 university ESL students in the UK (survey) 12 of these students (stimulate d recalls)	English proficiency • a wide range of proficiency levels • determined by IELTS scores	N/A	N/A	• listening-to-summarize tasks (i.e., their written or oral summaries) • assessed by two markers	<i>Qualitative analysis</i> Higher scorers (based on the marks of their written or oral summaries) • used more types of metacognitive strategies. • used inferencing more accurately • did not use of reconstruction strategies Lower scorers (based on the marks of their written or oral summaries) • relied more on directed attention, location, and reconstruction strategies to compensate for the lack of understanding of the content • employed inferencing and elaboration less successfully
Medium	Lee & Cai (2010)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	20 university EFL students	Listening proficiency • high and low • determined by listening tests and by the instructor	N/A	N/A	N/A	<i>Chi-square analysis</i> • a statistically significant difference in the use of strategies across proficiency groups ($\chi^2 (2, N=180) = 20.550, p < .000$) • HP (16.6% of the use of strategies) and LP learners (13.3%) used the ignoring strategy similarly • differences were found in the way the inferencing strategy (73.3% in HP learners vs. 47.8% in LP learners) and the 'no attention' strategy (10.0% in HP learners vs. 38.9% in LP learners) were used • a statistically significant difference in the use of knowledge sources across proficiency groups ($\chi^2 (3, N=82) = 38.375, p < .000$) • while HP and LP learners both used semantics in the local co-text or Semantics 1 combined with extra-textual knowledge frequently (42.3% and 36.7% respectively), the stark differences were observed in the use of semantics of the overall co-text or Semantics 3 (40.4% and 3.3% respectively) and morphology combined with extra-textual knowledge or Semantics 2 combined with extra-textual knowledge (1.9% and 56.7% respectively). <i>Qualitative description</i> • LP learners tended to fail in using correct clue words (both in local and global co-text) and morphological clues, which led to the overreliance of their extra-textual knowledge. Although HP learners also misused clues, they were able to identify the dissonance between the inappropriate use of knowledge sources and the overall understanding of the text (i.e., Semantics 3). • although HP learners also misused clues, they were able to identify the dissonance between the inappropriate use of knowledge sources and the overall understanding of the text (i.e., Semantics 3).
Medium	Cai & Lee (2012)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	20 university EFL students in China	N/A	N/A	N/A	• listening to nine texts and answering questions about the meaning of novel words • lexical inferencing performance • listening comprehension	<i>ANOVA and the Tukey's post hoc test</i> Text comprehension • Inferencing > Ignoring (t = 3.5994, p < .05) • Inferencing > No attention (t = 5.5916, p < .05) <i>ANOVA and the Tukey's post hoc test</i> Lexical inferencing • the use of morphology combined with extra-textual knowledge < other knowledge sources • the use of semantics of specific words in the global co-text (Semantics 2) combined with extra-textual knowledge < other knowledge sources

Overall WoE	Study	Research design	Sample	Proficiency	Strategy instruction	Treatment	Outcome measure	Findings
Low	Rahimrad & Moini (2015)	Mixed methods study (convergent design) with quantitative randomized controlled trial and qualitative description	15 adult EFL students in Iran	Listening proficiency • ranged from lower intermediate to intermediate level • determined by a listening pretest	• 89 minutes: two sessions a week; 5 weeks (16 hours in total) • four stages of metacognitive strategy instruction (Vandergrift, 2003)	Treatment group ($n = 8$) Control group ($n = 7$)	• a listening section of IELTS	<i>Independent t-test</i> Pretest • no statistically significant differences between the groups ($t = -.932, p \geq .05$) Posttest • treatment group > control group ($t = -2.431, p \leq .05$) <i>Qualitative analysis (Interviews)</i> • treatment group increased the use of planning and directed attention most, followed by evaluation and monitoring • participants had no idea about how to selective attention appropriately nor its usefulness even after the strategy instruction
Low	Carrier (2003)	Quantitative non-randomized study (cross-sectional analytic study)	7 high school ESL students in the US	English proficiency • intermediate • determined by school	• 20 to 30 minutes; 15 class sessions; 6 weeks • explicit strategy instruction (Chamot & O'Malley, 1994) • bottom-up (i.e., discrete) and top-down (i.e., video) listening processing (Interactive-Constructivist model of listening; Vandergrift, 1999) • effective note taking • important academic skills	All participants	Discrete (bottom-up) listening skills • a test to measure their ability to discriminate sounds • participants listened to an audiotape and checked off or wrote their responses to the questions on the answer sheet provided Video (top-down) listening skills • listening to an academic lecture video • participants took notes on a 2-minute video; the number of correct facts used as a measure of a top-down listening ability	<i>Nonparametric Wilcoxon Signed-Rank test</i> Discrete listening tests • posttest scores > pretest scores ($W = 1, p < .025$, one-tailed) Video listening tests • posttest scores > pretest scores ($W = 0, p < .01$, one-tailed)
Low	Smidt & Hegelheimer (2004)	Mixed methods study (convergent design) with the before-and-after trial without control, prevalence study, and qualitative description	9 university ESL students in the US	English proficiency • high, intermediate, low • determined by self-reported TOEFL scores	N/A	N/A	• listening to lecture video and answering multiple-choice questions	<i>Qualitative analysis</i> High proficiency group • favor monitoring and inferencing strategies • use both metacognitive and cognitive strategies Low proficiency group • use only cognitive strategies • use the strategy "remembering words" • answer the questions by referring to the wrong media
Low	Chou (2015)	Mixed methods study (convergent design) with cross-sectional analytic study and qualitative description	20 university EFL students in Taiwan	English proficiency • B1 threshold (i.e., intermediate) • determined by the Common European Framework (CEFR)	EGAP (the first semester) • once a week • the participants listened to audio recording and answered text items for 15 minutes. • retrospective interviews carried out with 20 participants (Week 17) ESAP (the second semester) • similar to that of the first semester	All participants	• listening comprehension tests with multiple-choice, one-word short answer, gap filling, and information transfer items	<i>Qualitative analysis</i> Most useful strategies • listening to main ideas (both EGAP and ESAP) • strategies that involved their prior knowledge (e.g., using imagination, making inferences, using knowledge about Chinese, guessing from the tone, making a mental summary; EGAP) • taking notes of key words and test wiseness strategies (ESAP) Least useful strategies • metacognitive strategies such as monitoring comprehension or making decisions as to whether the words heard matched the context (EGAP) • guessing by the tone of voice, using imagination, and translation (ESAP) (although the survey results suggested that translation was found to be used more for the ESAP task than the EGAP task) Strategies of varied usefulness • the note-taking strategy and the lexical inferencing strategy
Low	Du (2018)	Qualitative study (qualitative description)	9 university EFL students in China	Listening proficiency • high and low • results of the listening comprehension test in a larger program Linguistic knowledge • high and low • determined by a grammar test (Murphy, 2001) and a vocabulary recognition test (Schmitt et al., 2001)	N/A	N/A	• listening to audio recording and answering multiple-choice questions and open-ended questions	<i>Qualitative analysis</i> Strategies used by learners with high linguistic knowledge and/or high listening proficiency • selective attention (by HLP) • general deduction strategy ? • hypothesis confirmation strategy Strategies used by learners with low linguistic knowledge and/or low listening proficiency • elimination deduction • matching lexis heard to lexis • monitoring for sense Strategies used by all learners • elaboration • comprehension monitoring • real-time assessment of input

Appendix L: Summary of the studies related to the third in-depth review question

Overall WoE	Study	Sample	Research design	Strategy instruction	Treatment	Allocation	Outcome measure	Findings
High	O'Bryan & Hegelheimer (2009)	4 university ESL students in the US	Mixed methods study (sequential exploratory design) with grounded theory and cohort study	• 15-week semester • classroom instruction: listening for lecture cues, taking notes, listening to numbers and statistics, and listening for additional indicators of academic lecture organization • 14 instructor-designed podcasts: either demonstrating or encouraging students to review and practice the listening strategies discussed in class • textbook: different types of classroom-based listening strategies	All participants	N/A	• MALQ	<i>Descriptive statistics</i> • largest increase in problem-solving strategies (inferencing, elaboration, evaluation, monitoring and summary) • clearer development of personal knowledge strategies in a high proficiency student (although a lower proficiency student also reported the change in an interview) • slight gains in both directed attention strategies and mental translation strategies • no change in planning and evaluation strategies
Medium	O'Malley et al. (1985)	75 high school ESL students in the US	Quantitative randomized controlled trial	• 50 minutes; 10 days (the first and last days were spent for pre and posttests) • Metacognitive group: selective attention (metacognitive), note taking (cognitive, cooperation (socioaffective)) • Cognitive group: note taking (cognitive, cooperation (socioaffective))	Metacognitive (<i>n</i> = 27) Cognitive (<i>n</i> = 26) Control (<i>n</i> = 22)	Random	• a short listening comprehension test with multiple-choice recognition items designed to assess Bloom and Krathwohl's (1977) knowledge, comprehension, and analysis levels	<i>Analysis of covariance</i> Pretest • the results were controlled for. Interim assessments • test 1 metacognitive > cognitive > control (<i>p</i> < .10) • test 2 cognitive > metacognitive > control (<i>p</i> < .05) • test 3 cognitive > metacognitive > control (<i>p</i> < .01) • test 4 metacognitive > cognitive > control (<i>p</i> = .626) Posttest • metacognitive > cognitive > control (<i>p</i> = .162)
Medium	Weinberg et al. (2011)	75 university French as a second language (FSL) students in the US	Mixed methods study (convergent design) with survey, qualitative descriptions, and grounded theory	• watching one podcast on a DVD per week over a period of seven weeks (to enhance student listening strategies and note-taking skills) • podcast contents 1. listening to lectures: Applying the strategies of good listeners 2. preparing for the lecture before class and when arriving in class 3. monitoring and problem-solving during and after the lecture 4. best practices in note-taking 5. Cornell note-taking system 6. reviewing and evaluating the lecture 7. new technology and note-taking (wikis, semantic mapping, and Microsoft Office OneNote).	N/A	N/A	Questionnaire (main source) • administered for about 20 minutes after watching a DVD in a 90-minute class each week • the second section of the individual questionnaire (a five-point scale) asked students to indicate the degree to which they used the listening strategies before and after viewing the podcasts • 50 and 70 responses per week in total from all four disciplines Focus group discussion	<i>Descriptive statistics</i> • increased use of listening strategies (Podcasts #1 through #6) • most substantive change on the socio-affective strategies (Podcast #1) • no change or slight decrease in the use of new technologies (Podcast # 7) <i>Qualitative analysis</i> (verbal data from focus group discussion) • increased use of the strategies presented in the podcasts in general, especially for metacognitive, cognitive, and socio-affective strategies
Low	Carrier (2003)	7 high school ESL students in the US	Quantitative non- randomized study (non- randomized controlled trial)	• 20 to 30 minutes; 15 class sessions; 6 weeks • explicit strategy instruction (Chamot & O'Malley, 1994) • bottom-up (i.e., discrete) and top-down (i.e., video) listening processing (Interactive-Constructivist model of listening; Vandergrift, 1999) • effective note taking • important academic skills	All participants	N/A	Discrete (bottom-up) listening skills • a test to measure their ability to discriminate sounds • participants listened to an audiotape and checked off or wrote their responses to the questions on the answer sheet provided Video (top-down) listening skills • listening to an academic lecture video • participants took notes on a 2-minute video; the number of correct facts used as a measure of a top-down listening ability	<i>Nonparametric Wilcoxon Signed-Rank test</i> Discrete listening tests • posttest scores > pretest scores (<i>W</i> = 1, <i>p</i> < .025, one-tailed) Video listening tests • posttest scores > pretest scores (<i>W</i> = 0, <i>p</i> < .01, one-tailed)
Low	Rahimirad & Moini (2015)	15 adult EFL learners in Iran	Mixed methods study (convergent design) with quantitative randomized controlled trial and qualitative description	• 89 minutes: two sessions a week; 5 weeks (16 hours in total) • four stages of metacognitive strategy instruction (Vandergrift, 2003)	Treatment group (<i>n</i> = 8) Control group (<i>n</i> = 7)	Random	• a listening section of IELTS	<i>Independent t-test</i> Pretest • no statistically significant differences between the groups (<i>t</i> = -.932, <i>p</i> ≥ .05) Posttest • treatment group > control group (<i>t</i> = -2.431, <i>p</i> ≤ .05) <i>Qualitative analysis (Interviews)</i> • treatment group increased the use of planning and directed attention most, followed by evaluation and monitoring • participants had no idea about how to selective attention appropriately nor its usefulness even after the strategy instruction

Appendix M: Annotated bibliography of included studies

1. Bidabadi and Yamat (2013)

Bidabadi, F. S., & Yamat, H. (2013). EFL learners' perceptions towards meta-cognitive strategy use in English language listening. *GEMA Online Journal of Language Studies*, 13(3), 31–43. <http://ejournal.ukm.my/gema/article/view/4219/2336>

The authors conducted a single-case study of Iranian female EFL university students to explore their perceptions towards the use of metacognitive processes and strategies when listening to tapes or a lecturer in English. The participants studied Teaching English as a Foreign Language (TEFL) and were all at an intermediate level. Quantitative data were collected from 84 participants through a questionnaire. Four factors (i.e., planning and evaluation, problem-solving, personal knowledge, and directed knowledge) were adapted from MALQ (Vandergrift et al., 2006) and measured through a five-point Likert-scale format from one (Strongly Disagree) to five (Strongly Agree). The descriptive statistics showed that directed knowledge ($M = 3.68$, $SD = 0.53$) received the highest mean scores, followed by planning and evaluation ($M = 3.55$, $SD = 0.49$), problem-solving ($M = 3.44$, $SD = 0.46$), and personal knowledge

($M = 3.33$, $SD = 0.59$). From the results of the descriptive statistics, the authors claim that directed knowledge is a particularly important strategy for these learners. Semi-structured interviews were also conducted for six participants who participated in the questionnaire survey. The interview data were transcribed and coded, which resulted in four thematic categories of planning and evaluation, problem-solving, personal knowledge, and directed knowledge. As for planning and evaluation, the participants acknowledged the influence of their background knowledge and mood, the clarity of voice, and environmental noise. They also perceived the importance of understanding and preparing for the listening tasks before entering classrooms. For problem-solving, the participants deduced the meaning of unknown words from context (e.g., familiar words) or prior knowledge, evaluated the plausibility of their interpretation based on their topical knowledge, took notes on unclear points and asked questions to the lecturer, sat in front of the classroom to hear the voice more easily, and created the quiet learning environment. Regarding personal knowledge, the participants noted the listening difficulties (e.g., speech rate or accents), the need for further practice (instead of using metacognitive strategies), and anxieties about being laughed at by peers and teachers. Lastly, in terms of directed knowledge, the participants

employed various strategies (e.g., taking notes, raising questions) to maintain their focus. Thus, the author provided further evidence for the participants' use of each strategy type identified by the MALQ. Although the study took a mixed-methods design that converged quantitative and qualitative data analyses, the qualitative data analysis was only used to confirm the findings from the MALQ. The validity of the qualitative analysis procedure was questionable because the grounded theory approach, which is inductive coding, produced the same codes as the factors in the MALQ. This may have prevented thick, rich, and possibly interesting new findings of the participants' perception of their listening strategies. Therefore, while the study design, the methodology, and the results were relevant to the first review question, the evidence is rather limited to provide new insights into the question.

2. Cai and Lee (2012)

Cai, W., & Lee, B. P. H. (2012). Processing unfamiliar words: Strategies, knowledge sources, and the relationship to text and word comprehension. *Canadian Journal of Applied Linguistics / Revue Canadienne de Linguistique Appliquee*, 15(1), 122–145.

<https://search.proquest.com/docview/1033568066?accountid=13042>

The scholars explored strategies and knowledge sources to process unfamiliar words in relation to successful word comprehension and listening comprehension. Twenty Chinese second-year university students participated in individual interviews, where they listened to nine science texts including one novel word. A native speaker of American English produced the recording by reading aloud the texts. After the listening task, the participants were asked to answer the questions about the meaning of novel words with and without retrieval cues (i.e., replaying the sentence including the novel word). Then, they were asked to recall the content of the texts. Their verbal reports were recorded and transcribed, which were coded for strategies for processing unfamiliar words and knowledge sources for inferring word meaning. Their answers to the questions were scored for lexical inferencing (by the

resemblance to the native speakers' answers) and text comprehension (the total number of correct propositions). The chi-square analysis of the frequencies of strategies suggested that there was a statistically significant difference in the frequencies of the lexical inferencing strategy (60.6%), the lexical ignoring strategy (15%), and "no attention" (24.4%; $X^2 (2, n = 180) = 62.433, p < .000$). As for text comprehension, a significant difference was found among the three strategies ($F = 14.258, p < .001$). The Tukey's post hoc test suggested that the listeners' comprehension was better with inferencing than ignoring ($t = 3.5994, p < .05$) or no attention ($t = 5.5916, p < .05$). There was also a statistically significant difference in the frequencies of the use of the knowledge sources, $X^2 (6, n = 92) = 50.1304, p < .000$. Semantics 1 (semantics of words in the local co-text) combined with extra-textual knowledge (i.e., general knowledge beyond the text; 35.9%) and Semantics 3 (semantics of the overall co-text; 23.9%) were used more than the expected value, whereas other knowledge sources were used but below the expected value. Qualitative data provided a detailed account of how the most frequently used knowledge sources were employed. Lastly, although an ANOVA analysis showed a significant difference ($F = 13.892, p < .001$), the Tukey's post hoc test showed no significant difference across the lexical inferencing scores with

the different knowledge sources, except the use of morphology combined with extra-textual knowledge and the use of semantics of specific words in the global co-text (Semantics 2) combined with extra-textual knowledge. The two types of knowledge sources led to significantly lower lexical inferencing scores. Qualitative descriptions of the two such cases followed, explaining how the listeners who used those knowledge resources failed to make successful inferences based on incorrect information on morphology or salient but irrelevant words. Based on the results, the authors argued that lexical inferencing was the most frequently used strategy, which could be associated with successful listening comprehension. While the participants relied more heavily on semantics, the flexible use, rather than selective use, of different knowledge sources seemed important for accurate lexical inferences. Despite the opposing views held by some researchers (e.g., Ridgway, 2000), the findings indicated that listeners do employ complex strategies, such as the lexical inferencing strategy in listening. It was also suggested extra-textual knowledge is often used in inferencing, although the misuse of extra-textual knowledge occurs when basing it on inappropriate textual information. The study is a mixed-methods convergent design and effectively exploited the quantitative data to give detailed accounts of the quantitative findings to

answer their research questions. However, although the authors claim that “flexible use of the knowledge sources” is important, the results do not explain or provide evidence for what the flexible use means. As the study did not investigate metacognitive awareness, the relationship between metacognitive awareness and the flexible use needs to be explored empirically to test this claim. Moreover, neither their population nor sampling strategy was not explicitly explained, although it can be assumed that their population was “L2 learners” and the sampling strategy was either “convenience” or “purposive” strategy. The study provided strong evidence for the first review question, although the generalization of the findings need caution due to the unclear sampling strategy and the use of the artificial listening material and task. While the study collected data on lexical inferencing and listening performance, the second review question cannot be fully answered because the study did not incorporate experimental designs or statistical techniques that can determine the causal relationship. In addition, the study also not control for linguistic knowledge, which also made it difficult to identify the effectiveness of strategy

3. Cai and Lee (2010)

Cai, W., & Lee, B. P. H. H. (2010). Investigating the effect of contextual clues on the processing of unfamiliar words in second language listening comprehension. *Australian Review of Applied Linguistics*, 33(2), 18.1-18.28. <https://doi.org/10.2104/aral1018>

The authors explored how novel words with different contextual clues affect strategies and knowledge sources L2 learners use in listening comprehension. Twenty Chinese university students listened to nine academic texts, each of which included one novel word. The novel words were categorized according to three clue types: local co-text clues (i.e., clues from the sentence in which the target word occurs), global co-text clues (i.e., clues from an overall understanding of the text), and extra-textual clues (i.e., clues from background knowledge or experience regardless of the involvement of textual information). In individual interviews, the participants were asked to listen to nine texts and answer questions about the meaning of novel words with or without retrieval cues and how they got the meaning of them. Then, the participants were asked to recall the content of the texts. Their verbal reports were recorded and transcribed, which were coded for strategies for processing unfamiliar words and knowledge sources for inferring word meaning.

As a result, a statistically significant difference was found in the use of the three strategies across the three conditions, $X^2 (4, n = 180) = 25.871, p < .000$. It was also found that the participants used the inferencing strategy much more frequently with Condition 2 (global co-text clues; 68.3% of their strategy use) and Condition 3 (extra-textual clues; 76.7%) than with Condition 1 (local co-text clues; 36.7%). On the other hand, the ignoring strategy was used more frequently with Condition 1 (local co-text clues; 30.0%) than with Condition 2 (global co-text clues; 10.0%) or Condition 3 (extra-textual clues; 5.0%). A qualitative description explained how the meaning of globally defined novel words could be reinforced with multiple co-text clues. The use of knowledge sources was also found to be different across three conditions, $X^2 (6, n = 96) = 95.234, p < .000$. It was found that the participants used the knowledge sources (e.g., semantics in the local co-text, the semantics of the overall co-text, and extra-textual knowledge) according to the corresponding clue types (local, global, and extra-textual clues). Based on the findings, the authors argued that priority should be given to teaching the inference strategy by global and extra-textual cues and that L2 learners' ability to infer the meaning of novel words can be used for vocabulary acquisition through comprehensible oral input. The study revealed how the inferencing strategy

can be effectively used from global co-text clues and background knowledge by L2 learners. Nevertheless, since the participants were English majors, the findings of this study may be generalizable only to relatively higher proficiency L2 learners.

4. Carrier (2003)

Carrier, K. A. (2003). Improving high school English language learners' second language listening through strategy instruction. *Bilingual Research Journal*, 27(3), 383–408. <https://doi.org/10.1080/15235882.2003.10162600>

The researcher explored the effectiveness of listening strategy instruction on listening performance. Seven high school ESL students who study various academic subjects in English in the US received six-week listening strategy instruction on a variety of listening strategies based on Vandergrift's Interactive-Constructivist model of listening (1999). Pre- and posttests examined discrete (bottom-up) listening skills (i.e., the ability to discriminate sounds, syllable number, syllable stress, contractions and reductions, word stress, sentence meaning, and thought groups) and video (top-down) listening skills (i.e., the ability to selectively attend to, comprehend, and record information from oral input supported by visual cues). Discrete listening skills were measured by the number of correct responses to prepared questions, whereas video listening skills were evaluated by the number of correct facts included in their notes taken during the listening. It was shown that posttest scores on the discrete listening tests were statistically significantly larger than

pretest scores, $W = 1$, $p < .025$, one-tailed. Similarly, posttest scores on the video listening tests were statistically significantly larger than pretest scores, $W = 0$, $p < .01$, one-tailed. Thus, this study showed that explicit listening strategy instruction helped this group of high school ESL students improve their discrete listening ability and their video listening and note-taking abilities. Based on the findings, the author argues for balanced and explicit listening practice on bottom-up and top-down listening strategies. This study, however, has several disadvantages. The effectiveness of explicit strategy instruction is unclear because the experimental design included neither randomization, a control group, nor a delayed posttest. Moreover, although the intervention covered a range of top-down listening strategies, only the note-taking strategies were featured as top-down listening strategies; the mere teaching of note-taking strategies might have yielded a similar result. Since there was no evaluation of listening strategies before and after the intervention, the effectiveness of listening strategy instruction may have resulted from other factors, such as increased vocabulary knowledge. Moreover, the small sample size and the use of audio and video recordings in the listening activities make it difficult to generalize the findings to a wider population as well as listening strategies in L2 lecture comprehension.

5. Chand (2007)

Chand, R. K. (2007). Same size doesn't fit all: Insights from research on listening skills at the University of the South Pacific (USP).

International Review of Research in Open and Distance Learning, 8(3), 1–22. <https://doi.org/10.19173/irrodl.v8i3.383>

The author explored five study skills teachers' and 19 former students' perceptions of listening skills and strategies in a distance study skills course at the University of the South Pacific (USP). The context in which the teachers taught and the students learned was reported in detail, describing the varied experiences and access to resources and facilities. While some teachers taught at the same location with adequate facilities made available to them, other teachers had to travel to remote areas for tutorials. Similarly, some students commuted to USP centers to attend satellite tutorials, whereas other students only studied with the course materials delivered to them and never met their tutors. In the course, the students listened to recordings of lectures and seminar presentations, practiced note-taking and attending to transitional words, and engaged in pre- and post-listening activities. The author interviewed the teachers and students about their experience with the listening skills section of the course. The interview data were transcribed

and coded using NVivo based on a grounded theory approach. The data revealed discrepancies between the teachers' and students' views. Due to the limitation of time, the lack of contact with the students, and the lack of resources and facilities, the teachers tended to underestimate their students' listening skills, the use of strategies, and self-regulation. In contrast, the students reported the use of various listening strategies proposed by O'Malley, Chamot, Stewner-Manzanares et al. (1985), and the development of their listening skills during the distance course. The author also pointed to the gap between research and practice, based on the finding that the teachers' insufficient knowledge of listening strategies expressed during the interviews. It was also noted that research so far had been inclined toward listening in classroom settings and that there was a need to gain more insights into listening skills and strategies in distance learning settings. This study is unique in investigating listening skills and strategies in a distance study skills course and offering detailed descriptions about the environment and experiences that the teachers and students had. Nevertheless, the study relied on the existing listening strategy framework to organize the emergent themes, reducing the richness of students' account of their experience regarding the use of listening skills and strategies. Although NVivo was used to increase the

transparency of the data collection and analysis, the validity and reliability of the procedures were unclear because of the non-participation of their participants or her colleagues. The author claims that teachers undervalued their students' listening skills and strategies without sufficient knowledge of listening strategies, although the teachers' accounts rather reflected the problems the students might have actually experienced. In this regard, the students' accounts were not convincing due to the lack of thick descriptions and opposing views on the issue. Another weakness of this study is related to the fact that the data did not reflect the views of the current students. Moreover, since the study mainly concerned listening strategies in non-classroom learning situations when listening to recordings of academic speech, the findings of this study may not be generalizable to listening strategies in L2 lecture comprehension.

6. Chou (2015)

Chou, M. H. (2015). The influence of topics on listening strategy use for English for academic purposes. *English Language Teaching*, 8(2), 44–54. <https://doi.org/10.5539/elt.v8n2p44>

The researcher explored listening strategies used in listening activities in English for general academic purposes (EGAP) and English for specific academic purposes (ESAP) courses among 92 Taiwanese university undergraduates. They were intermediate level learners and participated in the EGAP course for 18 weeks during the first semester, followed by the ESAP course for the same period in the second semester. The listening activities for the EGAP course involved everyday life situations, whereas the activities for the ESAP course concerned academic subjects. After listening to those materials, the learners answered comprehension questions in a test format. A questionnaire was newly developed based on listening strategies (Vandergrift, 1997) and test-taking strategies (Cohen, 2006), which was used at the end of each course. More detailed insights into how strategies were used were sought from 20 students who agreed to participate in retrospective interviews. Retrospective interviews were conducted in the library

discussion rooms to gain detailed insights into how strategies were used by the interviews. The verbal data were translated into English with the help of two bilingual students, although the data were transcribed and coded into similar categories by the author. The questionnaire survey indicated that the participants agreed with the use of some metacognitive strategies (i.e., attending to the main ideas and details) and some cognitive strategies (i.e., using mental imagery, making a mental summary, and taking notes) more for the EGAP listening task than for the ESAP listening task. On the contrary, the learners agreed to have used two cognitive strategies (i.e., translating English to Chinese and using knowledge about Chinese) and test wiseness strategies to a greater extent for the ESAP listening task than the EGAP listening task. The learners' verbal data revealed that listening to main ideas were found to be useful by the learners for both the EGAP and ESAP listening tasks. The learners believed that strategies that involved their prior knowledge (e.g., using imagination, making inferences, using knowledge about Chinese, guessing from the tone, making a mental summary) were most useful in the EGAP listening while taking notes of keywords and test wiseness strategies were perceived to be useful for the ESAP listening. Metacognitive strategies, such as monitoring comprehension or making decisions as to

whether the words heard matched the context, were considered as least helpful during the EGAP listening, possibly due to their cognitively demanding nature. For the ESAP listening task, the learners found guessing by the tone of voice, using imagination, and translation least helpful, although the survey results suggested that translation was found to be used more for the ESAP task than the EGAP task. Lastly, the perceived effectiveness of the note-taking strategy and the lexical inferencing strategy differed among the learners. The author concluded that familiarity with the content affected the effectiveness of listening strategies that made use of the listeners' prior knowledge. This study effectively combined the quantitative descriptive study and a qualitative description and provided the rich accounts of how the effectiveness of strategies differed depending on the listening task and why they did so. However, while the questionnaires used were valid and reliable, no strategies were reported to increase the validity and reliability of the qualitative method. There was an issue with the study design because the listeners participated in the ESAP course after taking part in the EGAP course. Therefore, participation in the latter may have affected the use of listening strategies in the former. The interpretation of the scales was also problematic because the author made claims about the frequency of strategies based on agree-

disagree scales. The evidence for the effectiveness of the listening strategies in this study is weak because it is based on the learners' perception. Finally, the population and the sampling strategy were unclear. Although the findings may be relevant to intermediate EFL learners in general, the study cannot make a strong claim about the population. Moreover, the environment (e.g., the library discussion room) and the listening task (e.g., listening to recordings and answering test items) reduced the ecological validity of the findings. Therefore, the findings may not be generalizable to listening strategies in L2 lecture comprehension.

7. Du (2018)

Du, F. (2018). Question type, language proficiency and listening strategy use: An exploratory study. *Asian EFL Journal*, 20(9), 193.

https://search.proquest.com/docview/2153600344?accountid=13042%0Ahttp://oxfordsfx.hosted.exlibrisgroup.com/oxford?url_ver=Z3

9.88-

2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Allba&atitle=Question+Type%2C+Language+Proficie
ncy+

The scholar explored what listening strategies EFL learners use depending on their listening proficiency and linguistic knowledge in response to different listening question types. Nine university students were recruited at a university in China and were divided into high and low listening proficiency (i.e., HLP and LLP) and linguistic knowledge groups (i.e., HLK and LLK). After listening to listening passages taken from academic English textbooks, the participants answered multiple-choice and open-ended questions for five question types. Retrospective verbal data were collected while they answered the question. The data were coded into listening

strategies based on existing taxonomies using NVivo, although the author also attempted to identify new strategies. In response to questions concerning the global understanding of the passage (global questions), HLP learners made use of selective attention (to notice discourse markers), whereas LLP learners relied on elimination deduction and matching lexis heard to lexis. However, one LLP learner was successful in answering a global question by combining the strategies with his prior knowledge about the topic. Moreover, HLK students answered questions that require the integration of pieces of information within a few sentences (local questions) using elaboration. In contrast, LLK and LLP students seemed to prefer monitoring to answer those questions. Interestingly, HLK students had trouble decoding form-meaning relationships in speech despite their sufficient linguistic knowledge. In general, all the participants used elaboration, comprehension monitoring, and real-time assessment of input most frequently, although the use of these strategies did not necessarily lead to successful comprehension. The author claimed that HLK learners often used general deduction strategy (i.e., a deduction based on general information) and hypothesis confirmation strategy (i.e., confirming the correctness of interpretation or hypothesis), whereas LLK learners tended to rely on the word level meaning of the text

and used strategies, such as elimination deduction (i.e., a deduction based on a process of elimination) and monitoring for sense (i.e., checking whether interpretation intuitively makes sense), both of which compensated for the lack of linguistic knowledge.

Although higher linguistic knowledge was observed to lead to the employment of the strategies that were regarded as more effective, HLK learners were not always successful in making use of strategies effectively (e.g., wrong inferences based on prior knowledge).

As such, the author argued for the importance of teaching how to use listening strategies to learners of English. In general, the study's data collection and analysis methods were valid and reliable, although the poor reporting quality lowered the overall validity and reliability of the findings. Furthermore, the findings of this study may not reflect what occurs in listeners' minds in second lecture comprehension because of the low ecological validity of the findings. In addition to the use of recordings and retrospective interviews, the author decided to explore listening strategies in response to listening questions to observe how task types affect strategy use. Although this was achieved, listeners in classroom settings are not exposed to such questions in listening to a lecturer.

8. Flowerdew and Miller (1992)

Flowerdew, J., & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal*, 23(2), 60–80. <https://doi.org/10.1177/003368829202300205>

The researchers investigated the perceptions, problems, and strategies in second language lecture comprehension among 30 university students who studied TESL in Hong Kong. The study employed an ethnographic approach, arguing that the findings derived from a learner strategy approach can only show listeners' cognitive speech processing and disregard other important dimensions involved in second language lecture comprehension. Data were collected in various forms using observation, questionnaires, diaries, and interviews during a ten-week lecture course. Five strategies were revealed. Many students reported the usefulness of pre- and post-reading and help received from peers and tutors, although they were shy and avoided asking questions during the lecture. While students tried hard to maintain their concentration, some found it easier to do so when they were able to understand lectures. Students tended to mark the relevant sections of their textbook rather than taking notes. Opposing views were

expressed by the students in terms of the benefits of note-taking. This study effectively combined different data collection methods and offered insights into strategic behaviors contextualized in Chinese cultures in terms of how students sought help and employed marking instead of taking notes. Although this study has high ecological validity and revealed realistic insights into what and how strategies are used, some of the findings may not be generalizable to learners in other cultural contexts. Moreover, the study did not define listening strategies nor provided a theoretical framework to understand them.

9. Fung and Macaro (2019)

Fung, D., & Macaro, E. (2019). Exploring the relationship between linguistic knowledge and strategy use in listening comprehension.

Language Teaching Research, 00(0), 1–25. <https://doi.org/10.1177/1362168819868879>

The authors explored the relationship between linguistic knowledge (LK) and listening strategies among 646 secondary school students in Hong Kong. The questionnaire specifically developed for quantification of strategies to listen to a teacher was used.

Linguistic knowledge was measured for their vocabulary and grammatical knowledge via existing tests. The questionnaire and the two tests were piloted and validated, although the reliabilities were not checked in this study. The results revealed that recall of prior knowledge, selective attention on difficult words, and summarization were the most-often used strategies. The learners with higher LK (i.e., HLK) used strategies, such as selective attention on difficult words or segments, recall of prior knowledge, auditory representation and imagery, and evaluation of understanding, more frequently than the learners with lower LK (i.e., LLK). On the other hand, LLK learners seemed to employ more translation strategies compared to HLK learners. When the two LK learner groups

were subdivided into higher strategic (i.e., HS) or lower strategic (i.e., LS) groups, both similarities and differences were observed among the four groups of learners. The two highly strategic groups (HLK/HS and LLK/HS) showed similarities in the use of most strategies, as did the two less strategic groups (HLK/LS and LLK/LS). However, their behavior differed in terms of three strategies. Firstly, the two LLK groups were characterized by more use of translation than the two HLK groups regardless of HS or LS. Secondly, the LLK/LK group used hide and seek strategy as frequently as the two HS groups. Finally, the HLK/HK group used more selective attention on difficult words or segments than the LLK/HS group. These findings led the authors to conclude that linguistic knowledge can promote or hinder some strategies (e.g., selective attention), which were impacted by the availability of their cognitive resources freed up with higher linguistic knowledge. This explains why HLK learners can use more strategies than LLK learners in general. However, it was also observed that despite varied linguistic knowledge the two HS groups shared the use of many strategies, as the two LS groups did. This was clearly illustrated in the HLK/LK group, which used fewer strategies in general and did not feature any particular strategy use. The study provided insights into the relationship between linguistic knowledge and listening

strategies in a realistic context, which made a significant contribution to understanding the nature of strategies. However, the findings may be only generalizable to EFL secondary school learners in Asian contexts.

10. Hosogoshi (2016)

Hosogoshi, K. (2016). Effects of captions and subtitles on the listening process: Insights from EFL learners' listening strategies. *JALT CALL Journal*, 12(3), 153–178. <https://eric.ed.gov/?id=EJ1125240>

The scholar investigated the effects of captions and subtitles on EFL learners' listening strategies. To do so, she conducted a quasi-experimental study where 114 Japanese EFL learners were divided into three homogeneous groups according to their listening proficiency and listened to a lecture video with no-text, captions, or subtitles. After taking a listening comprehension test (to ensure the engagement in listening), the learners filled in the listening strategy questionnaire, which was developed specifically for the purpose of this study. On the listening comprehension test, the subtitle group obtained scores significantly higher than the no-text group and the caption group. The subsequent one-way ANOVA tests revealed that the subtitle group imagery and summarization strategies significantly more often than the other two groups. On the contrary, the correlational analyses suggested that the subtitle group tended to use strategies in isolation, whereas the no-text group and the caption group seemed to orchestrate cognitive and

metacognitive strategies. Based on the results, the author argued that the subtitle group utilized more imagery and summarization strategies related to parsing and utilization more often than the other two groups but not the strategies concerning perceptual processing because L1 subtitles facilitated the perceptual processing of J. R. Anderson's (1985) three-phase model of comprehension. Moreover, it was also claimed that no-text and caption conditions approximate real-life listening because of more use of strategies in combination. However, as the author notes, the correlations that were used to make claims about the combination do not provide information on whether the correlated strategies are actually used in combination. In addition, while it was argued that the use of subtitles scaffolds perceptual processing and promote learning of strategies used in the later stages, such as parsing and utilization, it is questionable if perceptual processing that occurred in L1 helps learners to acquire relevant L2 listening strategies in those stages. The validity and reliability of the collection tools should be further examined using other samples to establish the trustworthiness of the findings of this study. Moreover, the ecological validity of the findings is questionable because the participants listened to a lecture on a computer screen with or without captions or subtitles.

11. Jin and Xu (2017)

Jin, B., & Xu, W. (2017). Strategy use awareness in academic listening practices relative to L2 motivation among Chinese tertiary students. *Journal of Language Teaching and Research*, 8(4), 722. <https://doi.org/10.17507/jltr.0804.11>

The scholars explored the correlations between second language learning motivation and listening strategies. 79 second-year English major students at a Chinese university were recruited in the study. The listening tests conducted prior to the study indicated that their listening proficiency was at an upper-intermediate level. Following a TOEFL lecture listening practice and comprehension questions, the students filled in the questionnaire developed to elicit L2 motivation and listening strategies related to the listening task. The strategies were framed in O'Malley and Chamot's (1990) categorization of listening strategies. The correlation analyses suggested that teacher feedback positively correlated with self-evaluation, elaboration, inference, and after-class interactions. According to the authors, this indicated the potential benefit of strategy instruction. The significant correlations between commitment to the task and metacognitive strategies, such as planning, selective attention, self-monitoring, and self-evaluation suggested the

importance of clarifying their goals. While attribution and low-self efficacy were negatively correlated with metacognitive, cognitive, and affective strategies, none of them were found to be statistically significant. The significant correlation was also found between task significance and strategies, such as inference and after-class interactions, although the authors did not find any theoretical account for the relationships. Based on the findings, the study went on to discuss how EFL learners can be motivated to develop their listening strategies. However, the results should not be interpreted this way because the correlations do not suggest causation. Moreover, while the study employs a number of correlation analyses, confounding variables were never controlled for in the analyses. Although the results may apply to Chinese EFL learners, it is questionable if they can be generalizable to a wider population. In addition, the ecological validity of the findings is low because the listening task investigated in this study involved listening to a lecture video.

12. Lee and Cai (2010)

Lee, B. P. H., & Cai, W. (2010). The Effects of Language Proficiency on Unfamiliar Word Processing in Listening Comprehension. *Hong Kong Journal of Applied Linguistics*, 12(2), 61–82.

The researchers were interested in the effects of language proficiency on the use of strategies to process unfamiliar words and knowledge sources in listening comprehension. Twenty second-year English majors at a university in Beijing, who learn English as a second language, participated in the study. The data were collected through individual interviews, where they listened to nine texts (from a science textbook) and answered questions about how they made sense of the meaning of novel words with or without retrieval cues. At the end of the listening task, the participants were asked to recall the content of the texts. Their verbal reports were recorded and transcribed, which were coded for strategies for processing unfamiliar words and knowledge sources for inferring word meaning. The participants were divided into 10 high- and low- proficiency (HP and LP, respectively) levels based on the results on listening tests and the instructor's assessment of their listening ability. The results of a chi-square analysis revealed a statistically

significant difference in the use of strategies across proficiency groups, $X^2 (2, n = 180) = 20.550, p < .000$. While HP (16.6% of the use of strategies) and LP learners (13.3%) used the ignoring strategy similarly, differences were found in the way the inferencing strategy (73.3% in HP learners vs. 47.8% in LP learners) and the “no attention” strategy (10.0% in HP learners vs. 38.9% in LP learners) were used. Another chi-square analysis also found a statistically significant difference in the use of knowledge sources across proficiency groups, $X^2 (3, n = 82) = 38.375, p < .000$. While HP and LP learners both used Semantics 1 combined with extra-textual knowledge frequently (42.3% and 36.7% respectively), the stark differences were observed in the use of Semantics 3 (40.4% and 3.3% respectively) and morphology combined with extra-textual knowledge or Semantics 2 combined with extra-textual knowledge (1.9% and 56.7% respectively). The following qualitative description was further detailed that LP learners tended to fail in using correct clue words (both in local and global co-text) and morphological clues, which led to the overreliance of their extra-textual knowledge. Although HP learners also misused clues, they were able to identify the dissonance between the inappropriate use of knowledge sources and the overall understanding of the text (i.e., Semantics 3). Based on this argument, the authors argued for the

importance of teaching L2 learners to monitor the use of strategies and knowledge sources for its coherence with the text as well as enhancing LP learners' linguistic processing ability, such as distinguishing minimal pairs. Although the findings of this study suggest that proficient and less proficient L2 learners use strategies and knowledge sources differently due to the difference in their linguistic knowledge, their linguistic knowledge was not measured.

13. O'Bryan and Hegelheimer (2009)

O'Bryan, A., & Hegelheimer, V. (2009). Using a mixed methods approach to explore strategies, metacognitive awareness and the effects of task design on listening development. *Canadian Journal of Applied Linguistics*, 12(1), 9–38.

<https://search.proquest.com/docview/216987857?accountid=13042>

The authors conducted a longitudinal study to explore the use of listening strategies according to students' proficiency levels (low-intermediate to high-intermediate), the impact of repetition on listening strategies, and the development of students' metacognitive awareness. Four English as a second language (ESL) students attended a university-level strategy-based academic listening course over a 15-week semester, in which they learned listening strategies for lectures through classroom instruction and activities as well as 14 podcasts to listen to outside the classroom. The content of the podcasts covered the review of the strategies learned in the previous class, followed by lecture excerpts. The attempts were made to control for students' prior knowledge by choosing topics of general interest. A variety of data collection methods were used, such as verbal protocol, semi-structured interviews, the

Metacognitive Awareness Listening Questionnaire (MALQ), and student notes. During the think-aloud protocol, the participants listened to the podcast and verbalized their thoughts while taking notes. The verbal reports were then transcribed and coded by an open-coding approach. After the open coding, the coded strategies were classified into cognitive, metacognitive, and socio-affective strategies, according to the taxonomies proposed by O'Malley and Chamot (1990) and by Vandergrift (1997). In general, cognitive strategies especially summary strategies were used by all students more than metacognitive strategies. Unexpectedly, the student with the highest proficiency level used the least percentage of metacognitive strategies, indicating that the student was able to understand the listening text successfully. As for cognitive strategies, summarization is the strategy used most often by all students. In addition, higher proficiency students applied between-parts inferencing and personal elaboration to fill in the gap in knowledge to understand the text, whereas lower proficiency students used inferencing and elaboration strategies to process information at the local (word and sentence) level. In this study, the effect of repetition on listening strategies was also investigated by observing how the students used strategies while listening to the same text twice. The impact of repetition was more evident in one lower proficiency

student who failed to produce a complete summary in the first listening. Using bottom-up processing more heavily, the student's use of summary strategy led to a more coherent understanding of the listening text. Thirdly, this study addressed the development of metacognitive awareness throughout the semester. The results of the MALQ showed the largest increase in the use of problem-solving strategies (inferencing, elaboration, evaluation, monitoring, and summary). The MALQ also demonstrated the clear development of personal knowledge strategies in a high proficiency student, while a lower proficiency student stated in the interview that he became aware of his strengths and weaknesses through the participation in the course. Slight gains were observed in both directed attention strategies and mental translation strategies in the MALQ, although the increase in mental translation strategies contradicted the initial expectations. There was no change in planning and evaluation strategies. The significant contribution of this study is the combined use of verbal data and the MALQ in the attempt to provide a more comprehensive picture of the listening strategy use by the participants. Nevertheless, the findings of this study may not directly apply to other L2 learners due to the small sample size and the lack of statistical inferences. Moreover, the findings may not reflect the actual listening strategies during lectures

and lacked information on some strategies (e.g., socio-affective strategies) because strategies were elicited using podcasts and radio broadcast. While the study showed that some strategies, such as problem-solving strategies and personal knowledge strategies, may develop via explicit strategy instruction, the development of other strategies was not as clear.

14. O'Malley et al. (1989)

O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418–437. <https://doi.org/10.1093/applin/10.4.418>

The scholars investigated listening comprehension processes and strategies by effective and ineffective listeners. Eight high school students were selected and grouped into five effective and three ineffective listeners on the basis of their ESL teachers' holistic assessment. After being trained on thinking aloud using different language tasks, the students were asked to think aloud during the pause while listening to tape-recorded lectures. The verbal data were transcribed and coded for listening strategies by the researchers. The qualitative analysis suggested that the effective and ineffective listeners used listening strategies differently depending on each phase of listening comprehension. In perceptual processing, the effective listeners were able to maintain and direct their attention to the listening task, whereas the ineffective listeners often failed to redirect their attention when their concentration is disrupted. In parsing, the effective listeners listened for larger chunks, such as phrases or sentences, and attended

to intonation and pauses to use top-down processing. Bottom-up processing was used only out of necessity. On the other hand, the ineffective listeners attended to smaller units, such as words and sentences, and relied heavily on bottom-up processing. When the listeners found it difficult to parse the text, they used multiple strategies, such as inferencing, translation, and self-monitoring, to support their comprehension. In utilization, the effective listeners used elaboration with reference to their world knowledge, personal knowledge, and self-questioning. These listeners also inferred the meaning of unfamiliar words and phrases using their prior knowledge. The ineffective listeners, however, elaborated on the meaning of the text on fewer occasions and did not associate the information with their personal experiences. The study effectively identified listening strategies in the three phases of listening comprehension by effective and ineffective listeners. Nevertheless, the small sample size makes it difficult to generalize the findings. Since the study did not measure listening success objectively, the effectiveness of the listening strategies employed by the effective listeners was unclear. Moreover, it was not clarified in the study whether the amount of prior knowledge affected the use of elaboration in the students.

15. O'Malley, Chamot, Stewner-Manzanares et al. (1985)

O'Malley, J. M., Chamot, A. U., Stewner-Manzanares, G., Russo, R. P., Küpper, L., & Kupper, L. (1985). Learning strategy applications with students of English as a second language. *TESOL Quarterly*, 19(3), 557–584. <https://doi.org/10.2307/3586278>

The researchers investigated whether metacognitive, cognitive, and socio-affective learning strategies can be successfully taught and promote listening comprehension. 75 high school students were randomly allocated to two treatment and one control groups. For the period of eight days, the metacognitive group ($n = 27$) was taught one metacognitive strategy (selective attention), one cognitive strategy (note-taking), and one socio-affective strategy (cooperation), while the cognitive group ($n = 26$) received instruction on the cognitive and socio-affective strategies. The results from the pretests were treated as a covariate and controlled for in the subsequent analysis of covariance. Overall, the treatment groups outperformed the control group. According to the prediction, the metacognitive group performed better than the cognitive group in the first interim listening test. However, the cognitive group's performances were higher than the metacognitive group in the second and the third interim tests. The difference between the three

groups failed to reach significance in the last interim test as well as the posttest. As for the lack of effect of the intervention on metacognitive strategies in the latter listening tests, the author explained that the explicit cues were faded too quickly in the instruction for the metacognitive strategy to be transferred to the use in the latter listening tasks. Another possibility was that the students failed to use the metacognitive strategy effectively due to the increased task difficulty. Despite the limited period of intervention, the study presented one of the first evidence of the effectiveness of listening strategy instruction. However, the detailed procedure for the randomization was not reported. Therefore, it can be argued that the observed effectiveness of metacognitive and cognitive strategies exist, although the varying difficulty levels of the listening tests obscured the findings. Future studies should carefully match the intervention with students' progress and provide intervention for a sufficient period of time. Moreover, the detailed procedure for the randomization was not reported in the study. Therefore, it can be argued that the observed effectiveness of metacognitive and cognitive strategies exist, although the varying difficulty levels of the listening tests obscured the findings. Future studies should carefully match the intervention with students' progress and provide intervention for a sufficient period of time.

16. Olsen and Huckin (1990)

Olsen, L. A., & Huckin, T. H. (1990). Point-driven understanding in engineering lecture comprehension. *English for Specific Purposes*, 9(1), 33–47. [https://doi.org/10.1016/0889-4906\(90\)90027-A](https://doi.org/10.1016/0889-4906(90)90027-A)

The scholars were interested in why some students cannot comprehend the main points in academic lectures despite their high L2 proficiency. The authors conducted a qualitative investigation into the comprehension of a videotaped engineering lecture. 14 nonnative speakers of English who studied a variety of majors at graduate and undergraduate levels participated in the study. After listening to the videotape, the students summarized the listening content. Follow-up interviews were also conducted for several participants. The verbal data were transcribed and evaluated by the researchers for completeness and accuracy. While some students failed to understand the lecture due to the lack of proficiency, there were also students who could not understand the main points of the lecture despite their high level of English proficiency. The authors attributed the findings to the two types of listening strategies. According to the authors, a point-driven strategy is employed to understand the main (and subordinate) points of

discourse by taking into account the speakers' intentions and goals as well as other relevant contextual information. On the other hand, an information-driven strategy only concerns the acquisition of facts. By drawing on the distinction of the strategy types, the authors claim that the students who successfully understood the main points of the lecture used a point-driven strategy, while the students who did not succeed in comprehension relied on an information-driven strategy. The study points to the importance of increasing students' awareness of discourse-level pragmatics of academic lectures, such as the problem-solving nature of engineering lectures, in the case of the study. However, it should be noted that the study deduced the use of those listening strategies from the students' comprehension level of the lecture. The direct evidence was not reported in this study.

17. Rahimirad and Moini (2015)

Rahimirad, M., & Moini, M. R. (2015). The challenges of listening to academic lectures for EAP learners and the impact of metacognition on academic lecture listening comprehension. *SAGE Open*, 5(2), 215824401559060. <https://doi.org/10.1177/2158244015590609>

The authors provided Iranian learners of English with metacognitive strategy instruction for five weeks and looked at the change in listening comprehension, challenges experienced by them, and metacognitive strategies they perceived to use and to be effective.

15 adult learners who had had a master's degree were randomly assigned to treatment ($n = 8$) and control ($n = 7$) groups. The intervention consisted of five stages following Vandergrift (2003). Academic listening sections of the IELTS were administered in pre- and posttests. Individual interviews were held for all the participants soon after the pretest to elicit perceived challenges in lecture comprehension. After the posttest, only the participants in the treatment group were invited to individual interviews to gain insights into what metacognitive strategies they started using and found to be useful after the instruction. Independent t-tests were conducted to explore the group differences in the listening score for pre- and posttests. The significant difference was found in the posttest,

indicating the effectiveness of strategy instruction. The interview data suggested that the participants struggled with the fast delivery of a speech, the loss of concentration due to unfamiliar vocabulary, and the integral nature of listening to a lecture (e.g., note-taking, responding to questions). Furthermore, the participants who received strategy instruction stated that they increased the use of planning and directed attention most, followed by evaluation and monitoring. These findings further corroborated the effectiveness of instruction. On the other hand, the participants had no idea about how to selective attention appropriately nor its usefulness even after the strategy instruction. However, the study is limited in the small sample size and unclarity regarding the qualitative method and analysis procedures.

18. Rukthong and Brunfaut (2020)

Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53. <https://doi.org/10.1177/0265532219871470>

The scholars were interested in exploring the listening processes and strategies involved in integrated listening-to-summarize tasks, their effectiveness, and their relationship with the perceived listening difficulty. The authors recruited 72 Thai speakers of English at universities in the UK, who engaged in eight integrated tasks. In each task, the participants were asked to listen to a recorded academic lecture and summarize the content either in the form of writing or speaking. The summaries were assessed for the participants' comprehension of the content. The verbal data were collected from 12 participants through a stimulated recall, which provided insights into the participants' listening processes and strategies employed during the tasks. The rest of the students ($n = 60$) participated in a perception questionnaire, which tapped into several dimensions of perceived task difficulty. The stimulated recall data, notes taken by the participants while listening, and their oral or written summaries were coded for listening processes and

strategies. Five cognitive strategies and six metacognitive strategies were identified through the codes. As for cognitive strategies, the participants used prediction, fixation, inferencing, elaboration, and reconstruction as they listened to the recordings. They also utilized metacognitive strategies, such as preparing for listening, selective attention, directed attention, monitoring comprehension, real-time assessment of input, and comprehension evaluation. Furthermore, higher scorers in their summaries were characterized by their use of more types of metacognitive strategies. On the other hand, lower scorers relied more on directed attention and fixation strategies and used reconstruction strategies to compensate for the lack of understanding of the content. Lower scorers were also less successful in employing inferencing and elaboration. As for the listening tasks perceived to be more difficult, the participants used fixation, directed attention, and reconstruction strategies to recover the missed words and phrases. Inferencing and elaboration were used more for these tasks, although they were more effectively used in easier tasks. Comprehension monitoring was essential in organizing those cognitive strategies especially while listening to more difficult texts. The individuals' performance on the tasks also differed depending on the characteristics of the listening input. The study sheds light on the kind of listening strategies used for

integrated listening-to-summarize tasks. It also revealed that while the effective listeners tended to use metacognitive strategies and cognitive strategies, such as inferencing and elaboration, effectively, it is a challenge for them to adopt these strategies to the characteristics of the listening text. Although the study showed a range of listening strategies, the results may not represent the strategies used in real lectures due to the specific nature of the listening tasks.

19. Schroeder (2016)

Schroeder, R. M. (2016). The development and validation of the “Academic Spoken English Strategies Survey (ASESS)” for non-native English speaking graduate students. *Journal of International Students*, 6(2), 394–414.

<https://search.proquest.com/docview/1826533336?accountid=13042>

The researcher developed and validated the Academic Spoken English Strategies Survey (ASESS) to assess the listening and speaking strategy use of non-native English speaking (NNES) graduate students. It did so by the three rounds of testing and revising the questionnaire in reference to previous literature, experts’ insights, pilot studies, and the researcher’s journal. Then, exploratory factor analysis was run based on 384 participants’ responses to the questionnaire, resulting in five factors. Although the study provided new insights into listening strategies using an inductive approach, it lacked the theoretical explanations as to how strategies converged into factors. For instance, it is not clear why a strategy to “notice the speaker’s facial expressions, gestures, and voice changes” loaded onto “strategies for keeping on track while listening” (i.e., promoting the concentration and focus while listening)

rather than “strategies for clarifying understanding after listening”. The fifth factor consisted of only one strategy to “use laptops to check the meaning of words while listening”. The factor loadings and reliability score for each factor were not reported in the study (although the study states that they can be made available upon the contact with the author). Despite the author’s claim, these limitations reduced the validity and reliability of the questionnaire.

20. Smidt and Hegelheimer (2004)

Smidt, E., & Hegelheimer, V. (2004). Effects of online academic lectures on ESL listening comprehension, incidental vocabulary acquisition, and strategy use. *Computer Assisted Language Learning*, 17(5), 517–556.

<https://doi.org/10.1080/0958822042000319692>

The authors explored how a 15-minute authentic lecture video on horticulture (with overhead transparency notes and picture slides) affects listening comprehension and incidental vocabulary acquisition in 24 adult ESL learners. Along with the main interests, the study also investigated what listening strategies were employed in response to the listening activities with access to an online dictionary. Verbal data were collected from nine participants by retrospective interviews which were coded into metacognitive and cognitive strategies following Vandergrift's (1997) procedure. In general, cognitive strategies, such as linguistic and extralinguistic inferencing and summarization (22 out of 40 strategies), were used slightly more often than metacognitive strategies (18 out of 40 strategies). The strategy use varied depending on the learners' proficiency levels (i.e., self-reported TOEFL scores); high proficiency

students used both metacognitive and cognitive strategies and favored monitoring and inferencing, while low proficiency students used only cognitive strategies. Female learners used more strategies than male learners and tended to favor cognitive strategies over metacognitive strategies. On the other hand, male learners used slightly more metacognitive strategies than cognitive strategies and relied more on extra-linguistic inferencing. Finally, it was found that the learners who did not score well on the comprehension questions referred to the wrong media (i.e., overhead transparency notes and picture slides). Nevertheless, the study reported no measures to increase the validity and reliability of the coding of strategies in verbal data. Moreover, elaboration (i.e., use of world knowledge) was not reported by the learners, possibly due to the lack of relevant prior knowledge (an issue with relevance) or the validity issues inherent in the retrospective interviewing technique (i.e. forgetting and fabrication; Russo et al., 1989).

21. Weinberg et al. (2011)

Weinberg, A., Knoerr, H. H., Vandergrift, L., Knoerr, ne, & Vandergrift, L. (2011). Creating podcasts for academic listening in French:

Student perceptions of enjoyment and usefulness. *CALICO Journal*, 28(3), 588–605. <https://doi.org/10.11139/cj.28.3.588-605>

Weinberg et al. (2011) piloted podcasts for academic listening on Anglophone students who attended academic lectures in French in the US. Seventy-five university students in French as a second language classes (associated with four different academic disciplines: Political science, Anthropology, History, and Philosophy) filled in a questionnaire after they had watched a podcast episode (for academic listening in French) on a DVD in a class for seven weeks. At the end of the study, ten participants joined a focus group discussion to share their views about how they found the podcasts enjoyable and useful as well as whether there had been a change in their use of listening and note-taking strategies. In general, the students responded to the podcasts favorably; they seemed to like the contents but found room for improvement in terms of their presentation. The students found Podcasts # 2 (Preparing for the lecture) and # 7 (New technology and note-taking) most enjoyable and found Podcasts # 2 (Preparing for the lecture) and # 5 (The

Cornell note-taking system) most useful. On the other hand, the students found Podcasts # 3 (Monitoring and problem-solving during and after the lecture) and # 4 (Best practices in note-taking) neither enjoyable nor useful. However, the students revealed more favorable opinions about the podcasts except for Podcast #5 (The Cornell note-taking system) during the group discussion at the end of the course. Furthermore, the questionnaire collected data on the perceived change in participants' use of listening strategies before and after viewing podcasts. Student responses to a five-point Likert scale revealed notable changes in the use of listening strategies learned in the first (cognitive and socio-affective strategies), the second (planning), and the third (monitoring) episodes. However, the data on the perceived change in strategy use was collected at one point in time; no data were collected before watching each episode. While the focus of the podcasts (on listening to authentic lectures in French) increased real-world relevance, the study did not make explicit reference to the definitions or theories of listening strategies. In addition to no measures taken to increase the validity and reliability of the data collection and analysis procedures, a clear presence of the Hawthorne effect lowered the trustworthiness of the findings.