Department of Education, University of Oxford

A Systematic Review of Research on Adult L2 Language Learning Strategies and Strategy Instruction

John Joseph Kwong Ho Wong

Trinity Term 2022

Word Count: 17,343

Chapter 1 Introduction

1.1 Rationale

Second language (L2) learning strategies and strategic instruction have been a key research focus over the last three to four decades. While much research focuses on theory, the empirical support for language learning strategy use and strategy instruction have not reached conclusive results (Plonsky, 2011). Questions about methodological flaws and murky definitions have hindered the progress of research in this field (Hassan et al., 2005). Building upon previous related research such as Hassan et al. (2005) and Plonsky (2011), this present review seeks to systematically search for up to date, published empirical research (after 2011) which explore language learning strategies and strategic instruction in relation to an adult L2 language learning context.

A key impetus for this present review involves the author's personal motivations as a teacher-researcher. As an L2 language instructor of English for adult learners in Canada, the author is primarily interested in empirical research concerning language learning strategy and strategy instruction which could be actionable and applicable in the classroom. Within the author's own teaching context, the majority of adult L2 language learners come from a variety of nationalities, L1 backgrounds, and are looking for strategies and approaches to not only maximize language acquisition during class, but also utilize those skills to develop their language ability independently outside of the classroom. With a focus on pragmatic pedagogy and application, the findings of this review may support fellow educators of adult L2 learners and other relevant stakeholders in helping students achieve their language acquisition goals as effectively and efficiently as possible.

1.2 Outline

This present review begins by exploring an overview of relevant literature relating to language learning strategy and strategy instruction, especially within the adult L2

language learner context. Methodology is then presented, detailing the eligibility criteria, search strategies, screening process, quality assessment and results synthesis. Following the results section which summarizes the findings of included studies, the discussion portion of the review will synthesize the findings to answer specific research questions. Finally, limitations of the present review, recommendations for future research as well as final concluding thoughts will be discussed.

To ensure high quality reporting standards and methodological quality, this systematic review attempted to follow the guidelines stated in PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) (Moher et al., 2009; Page et al., 2021).

Chapter 2 Literature Review

This chapter discusses the current topic of language learning strategies (LLS) and strategy instruction (SI) in relation to available research and identifies three review questions. First, research and controversy about language learning strategies over the past four decades is explored. The discussion then introduces new innovations and understandings concerning L2 language learning strategies over the last two decades (such as context and complexity) which underpin contemporary research. Next, exemplary studies are reviewed to highlight previous findings and limitations which both contextualize and explain the need for the present study. Finally, based on current knowledge gaps in the field of language learning strategies and strategy instruction, the research questions of the current study will be addressed.

2.1 Defining language learning strategies in the face of self-regulation

From the field's early days in the 1970s (Rubin, 1975) to over four decades of cumulative research, "definitional fuzziness" (Rose, 2015) has both plagued and

galvanized research on language learning strategies and strategy instruction. Over the years, the focus shifted from “good” language learners (Rubin, 1975) to L2 learning strategies, with Oxford’s Strategy Inventory of Language Learning (SILL) (Oxford, 1990) kickstarting an abundance of literature and becoming the most “widely used instrument in language learner strategy research” (White et al., 2007).

The explosion of language learner strategy research from the 1990s resulted in mounting criticism of the field in the early 2000s. Two prime examples include Macaro (2006) and Dörnyei’s (2005) landmark papers which questioned the lack of definitional consensus and theoretical rigour (Macaro, 2006) in the field. Dörnyei (2005) took the radical position of abandoning language learning strategy altogether in favour of self-regulation. Until that point, language learning strategy research had been operating in a vacuum within applied linguistics. Research in the field involved overly simplified catch-all terms and approaches, resulting in “decades of conflicting research findings and difficulties in defining the characteristics of a strategy and the relationships between strategies [which] opened the field up to much criticism” (Rose, 2015). Macaro (2006) and Dörnyei (2005) re-shaped the discourse of language learning strategies by re-focusing research to higher scientific standards while situating language learner strategy research within the wider fields of psychology and education.

Macaro’s (2006) summary of problems in language learning strategy research included questions concerning whether learning strategies occurred in or outside the brain, what these strategies consisted of, how to classify them, whether they survived across all contexts and situations, as well as whether they were integral or simply addictive to language processing abilities. A cursory glance at Google Scholar (n.d.) shows that Macaro (2006) was cited 898 times, demonstrating its impact through how much of subsequent language learning strategy research over the next two decades was driven toward answering Macaro’s (2006) poignant criticisms of the field.

While Macaro (2006) shone a spotlight on the lack of consensus in language learning strategy research, Dörnyei (2005) nearly abandoned the field altogether.

Dörnyei (2005) argued that L2 language research should follow the example of educational psychology, first marginalizing the term “learning strategy” and later abandoning it in “favor of the more versatile concept of *self-regulation*” (Dörnyei, 2005, p. 170). Under this argument, focusing on self-regulation would allow for a more “dynamic concept than learning strategy” (Zimmerman & Risemberg, 1997, p. 105), while shifting the focus of research from the “*product* (strategies) to the *process* (self-regulation)” (Dörnyei, 2005, p. 191). While the concept of self-regulation resulted in seminal works and new psychometric instruments such as Tseng et al. (2006), many researchers viewed abandoning learning language strategies entirely as “throwing out the baby with the bathwater” (Rose, 2011)

In the decades following Macaro (2006) and Dörnyei (2005), the field of language-learner strategy has incorporated theoretical criticisms and concepts of self-regulation into multiple avenues of research. While full consensus is still hard to come by, Rose et al. (2018)’s systematic review of language learner strategy research from 2010-2016 painted an optimistic picture for future research through three distinct avenues. Firstly, studies such as Ardasheva and Tretter (2013) showed how amendments could be made to improve existing strategy measures while considering critiques. Secondly, studies focused on vocabulary learning have utilized self-regulation constructs in addition to centring around strategic learning (Rose et al., 2018). These studies like Ranalli (2012) and Zimmerman and Schunk (2011) introduce new theories from self-regulation research to help explain the strategic language of language learners (Rose et al., 2018). Finally, studies like Teng and Zhang (2016) utilize theory from both self-regulation and language learner strategies to investigate specific language skills such as “L2 listening, reading, spoken communication, and vocabulary learning” (Rose et al., 2018). Moving on from the incessant infighting within the language learning strategy space over the last two decades, state-of-the-art research and reviews like Rose et al. (2018) indicate that synthesis of disparate ideas and theories is possible, and that “at the end of the day, we are all working towards a common goal: To make learning a second or foreign language as efficient and effective as possible through the use of learning strategies” (Zhang et al., 2019).

It is under this controversial research backdrop that a definition of language learning strategies (in consideration of self-regulation) will be proposed. Owing to its careful consideration of 33 definitions spanning over four decades in a content-analytic study of strategy definitions, I have chosen to use Oxford (2017)'s encompassing definition of L2 learning strategies to inform this present study. While the definition is hefty, I believe Oxford (2017) is currently the best definition of language learning strategies in lieu of the definitional fuzziness that has dogged the field since the 1970s:

> "L2 learning strategies are complex, dynamic thoughts and actions, selected and used by learners with some degree of consciousness in specific contexts in order to regulate multiple aspects of themselves (such as cognitive, emotional, and social) for the purpose of (a) accomplishing language tasks; (b) improving language performance or use; and/or (c) enhancing long-term proficiency. Strategies are mentally guided but may also have physical and therefore observable manifestations. Learners often use strategies flexibly and creatively; combine them in various ways, such as strategy clusters or strategy chains; and orchestrate them to meet learning needs. Strategies are teachable. Learners in their contexts decide which strategies to use. Appropriateness of strategies depends on multiple personal and contextual factors."

2.2 Innovations: understanding L2 learning strategies and strategy instruction in context and complexity

Over the last two decades, key innovations involving context and complexity have shaped our understanding of L2 learning strategies and strategy instruction. With the advent of social media and increased media availability through the internet, a proliferation of culture-mixing in different L2 learning contexts have enriched L2 learner strategy instruction (Oxford 2017). Specifically, in a culturally heterogenous L2 learning environment, strategy instruction can never be fully uniform across all

cultures and contexts. With regards to practitioner application, Oxford (2017) suggests that teachers and learners of L2 languages can draw from a wide range of tools acknowledging differentiation in strategy instruction, such as attention to sensory preferences, cognitive styles, current strategy use, personal interests, and learners' goals. Rather than applying a one-size-fits-all approach toward strategy instruction, the suggestions of differentiating strategy instruction illustrated in Oxford (2017) may result in more personalized applications of strategy instruction.

Recently, researchers and teachers have reframed language learning strategies and strategy instruction from understanding how to do various unconnected tasks towards a new purpose, "with an aim toward autonomy or the transformation of fixed mindsets into growth mindsets" (Oxford, 2017). According to Dweck (2022), Entity theory (or a fixed mindset), is the belief that "qualities such as intelligence and learning ability are static, given, and fixed" (Oxford, 2017). In other words, innate language talent (or lack of) determines L2 language learning success and failure (Williams et al., 2016). On the other hand, incremental theory (or a growth mindset) will result in language learners who see themselves "as strategic agent[s] who actively develops his or her abilities" (Oxford, 2017). While Oxford (2017) and Williams et al., (2016) note that growth mindsets will better benefit L2 language learners, recent studies such as Williams et al. (2016) point out that fixed mindsets can be modified. This opens the door toward effective language learning strategy use and strategy instruction to all L2 language learners, regardless of their mindset at the onset of their language studies.

Additionally, hope-based interventions seem to be a fruitful avenue of research in both language learning strategies and strategy instruction. Oxford (2017) notes how "it stands to reason that learners would not use strategies unless they had at least some small hope that the strategies would be helpful." Crucially, the absence of hope results in a "learned helplessness" that is interwoven with being stuck with a fixed mindset (Oxford, 2017). Without hope and a growth mindset, it seems unlikely that any amount of strategy instruction and/or knowledge of effective language learning strategies will matter to a student who has given up ever becoming a

successful learner. In other words, by utilizing hope-based interventions in language research, these new puzzle pieces may better unlock effective language learning than the continual re-hashing of language learner strategy research's many "pitfalls of the past" (Rose et al., 2018).

Finally, recent research has emphasized the need to capture strategic flexibility in the face of strategies not fitting into standard categorizes (Oxford, 2017). Because strategy roles (like "cognitive" strategies, or "metacognitive" strategies) may be more flexible than standard categories suggest, Oxford (2017) proposes that language learning strategies should be considered along L2 skill areas instead. These skill areas, like "listening or writing, or language subsystems, such as pronunciation or vocabulary" (Oxford, 2017, p. 314), form the key organizational framework of the present study, as this study's purpose is to explore the deployment, effectiveness, and development of language learning strategies in adult L2 language learners. Therefore, rather than being limited to any individual strategic framework, this study will review a comprehensive search of related studies along the lines of select L2 skill areas.

2.3 Findings and limitations of previous studies, highlighting key knowledge gaps

Hassan et al.'s (2005) rigorous systematic review of strategy training in language learning forms a solid foundation with which to situate the present study. Hassan et al. (2005) sought to explore the effectiveness of strategy training through analyzing 38 studies out of an initial 567 potential references. At the time, the study found "sufficient research evidence to support claims that training language learners to use strategies is effective" (Hassan et al., 2005, p.4) and that "the evidence for its effectiveness is stronger for adult and higher education learners" (Hassan et al., 2005, p.66). Though broad, Hassan et al.'s (2005) affirmation of strategy training's effectiveness as well as its effectiveness for adult and higher education learners provides suitable justification for the present study's research focus on learner

strategies in an adult context, especially when considering the subsequent research supporting Hassan et al. (2005), as documented in Oxford (2017).

In terms of weaknesses and limitations of Hassan et al. (2005), there were questions over the longevity of post-intervention effects. There was also a question mark over the specifics behind whether strategy training was directly responsible for effective language learning, or if improved awareness of different strategies was responsible in self-motivating the learner (Hassan et al., 2005). Additionally, while Hassan et al. (2005) organized their systematic review with L2 language skills in mind (speaking, reading, writing, listening, vocabulary, and overall language ability), they noted that more research may be needed concerning the effectiveness of strategy training on specific areas such as speaking, writing, and vocabulary. This present study seeks to address these knowledge gaps using research done from 2011-2021 to further contribute to current discourse on language learning strategy and strategy instruction.

Another important meta-analysis exploring language learning strategies and strategy instruction was Plonsky (2011). In analyzing 95 samples from 61 primary studies of strategy instruction for L2 learners, Plonsky (2011) explored the effectiveness of L2 strategy instruction. Encouragingly, Plonsky's (2011) findings show that the study "carries important implications for language practitioners, such as empirical justification for integrating learner training programs into L2 curricula" (p. 1013). In short, Plonsky (2011) advocates that strategy instruction should hold greater importance within the wider context of L2 acquisition (p.1013).

Like Hassan et al. (2005), Plonsky (2011) noted similar weaknesses, limitations, and gaps to the study. A key limitation was the lack of delayed posttests, stating that "the value of SI depends greatly on whether and to what extent its effects last over time, yet only eight of the [61] studies contributing to this meta-analysis included delayed post-tests." Moreover, in line with assertions by Macaro (2006) and Rose (2015), Plonsky (2011) notes the challenge of choosing appropriate instruments and well as a need for more thorough explanations, scientific rigour, and better statistical reporting in studies on strategy instruction.

2.4 Research questions

In considering the current knowledge gaps in the present literature, this review aims to answer the following review questions:

1. What empirical research was carried out on adult L2 language learning strategies and strategy instruction between 2011 and 2022?
2. What evidence is available for the effectiveness of language learning strategies and/or strategy instruction for adult L2 language learners in terms of language proficiency?
3. What evidence is available for how language learning strategy use has changed over time for adult L2 language learners?

Chapter 3 Methodology

This chapter explains the methodology in conducting this current systematic review. The aims and principles of the review are first explored. Then, the systematic review process will be expanded upon. Search strategy and review protocol in the current review will be discussed. Next, inclusion and exclusion criteria pertaining to the systematic review will be identified. After multiple screenings with a second reviewer to limit bias and follow systematic review principles listed by Gough et al. (2012) and Macaro et al. (2018), data will be extracted using a data extraction table (study information, methodology, data analysis, results etc.) based on Macaro et al. (2018). A weight of evidence framework based on Gough (2007) and used by Macaro et al. (2018) will also be utilized as part of the data extraction process. Risk of bias across included studies will be further assessed using critical appraisal tools such as the Mixed Methods Appraisal Tool (MMAT); Hong et al., (2018). This tool is appropriate in systematic mixed studies and versatile in appraising qualitative, quantitative, and mixed methods studies. Reporting attempts to follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines (Moher et

al., 2009; Page et al., 2021). Data is then summarized and synthesized using a narrative approach (Gough et al., 2012), a suitable choice in systematic reviews when the studies considered contain diverse methodologies, theoretical underpinnings, and utilize differing types of qualitative and quantitative data.

3.1 Aims and principles of the review

Systematic reviews are particularly relevant in the fields of applied linguistics and second language acquisition (SLA) due to concerns about the reliability of research evidence (Gough et al., 2012) as well as the large body of research in areas such as adult L2 language learning strategies and strategy instruction (Plonsky, 2011). By synthesizing multiple studies, benefits, and actionable insights for stakeholders such as “students, teachers, and educational policy makers” can be explored (Macaro, 2019). As teacher-researchers in an adult SLA context, the current review opens the possibility for new research-informed pedagogical insights which could be directly applied to the classroom. The educational and occupational background of the researchers involved in the current study are important as the relevant perspectives of the research team pertains to pragmatic applications in the classroom.

Unlike a meta-analysis which focuses on studies that with statistical data that can be pooled, a systematic review is particularly useful in exploring the totality of existing research without solely focusing on quantitative data favoured by a meta-analysis (Macaro, 2019).

3.2 The systematic review process

To ensure that the current review answers the research questions in a systematic way, the five features which “differentiate systematic reviewing from more traditional narrative reviewing” (Macaro et al., 2012) will be used as guidelines in this review. Based on the features in both Gough et al. (2012) and Macaro et al. (2018), these features are as follows:

1. Systematic reviews are always carried out by more than one reviewer.
2. Transparent procedures must be used from throughout the review process from start to finish. An initial protocol or agreement among the review team members specifies how the review will be conducted.
3. Studies must be included through comprehensive and reliable searching.
4. Reviewer bias should be reduced as much as possible.
5. Syntheses detailing the reliability of the evidence reviewed should be produced.

As argued by Macaro et al. (2018) and Macaro (2019) doctoral theses shall be considered for the current review as it is a “feature of systematic reviewing that often differentiates it from narrative reviews”. Crucially, as stated in Macaro (2019) the absence of unpublished doctoral work in most research efforts means that “the null hypotheses might too often go unreported”, as doctoral work might not have definite/significant outcomes favoured by peer-reviewed journals. While an argument can be made against the peer-reviewed nature of unpublished doctoral work, Macaro (2019) highlights the fact that examiners are sometimes more rigorous than peer reviewers and are often subsequently published in peer-reviewed journals. Hence, with the right considerations, doctoral work can be relied upon as a valuable source of information during the systematic review process.

A key consideration in a systematic review is how to begin searching for literature while simultaneously eliminating bias. Macaro (2019) provides two viable methods to achieve this aim:

1. Rely on the review team’s existing knowledge of the field to formulate targeted, relevant research questions in service of the systematic review’s objectives.
2. Begin with keywords which can be inserted into reliable search engines to see what broad search appears.

In response to these considerations on eliminating bias in the systematic review process, the team in this systematic review are all teacher-researchers within SLA or the L2 Language space. Moreover, the formulation of the current review’s research

questions is based on past systematic review literature such as Hassan et al. (2005) and Plonsky (2011). Following a similar rationale to Macaro et al. (2018), the current review focuses on keywords and search terms that are present in relevant literature. This resulted in the inclusion of field-specific terms such as “language learning strategy”, “strategy instruction”, and “self-regulation”.

3.3 Search strategy and review protocol

A protocol document was created (Appendix A) following the PRISMA-P (Moher et al., 2015) which details how the review will be conducted. The protocol includes sections such as the introduction (rationale and objectives), and methods (eligibility criteria, information sources, search strategy, selection process, data collection process, data items, study risk of bias assessment, and data synthesis). This protocol was approved by my supervisor. The protocol was also reviewed by a second reviewer, a fellow master’s student studying the Applied Linguistics for Language Teaching program at the University of Oxford.

3.3.1 Inclusion/exclusion criteria

Eligibility criteria (inclusion and exclusion) was based on systematic reviews such as Rose et al. (2020) which explored a similar area in language pedagogy.

Inclusion criteria:

- Must contain primary empirical research.
- Must be articles been published between 2011 and 2021.
- Must be about language learning strategies (LLS) or strategy instruction (SI), specifically within L2 acquisition.
- Must be topically related to one of the research questions mentioned above in the objectives section (language learning strategy usage, effectiveness of

learning strategies, strategy usage over time, and effective strategy instruction).

Exclusion criteria:

- Focuses on topics unrelated to language learning strategies, strategy instruction, or L2 acquisition.
- Population does not include adult L2 learners.
- Theoretical articles or reports on practice without research methodology.
- Although research in multiple languages is a form of high-quality systematic reviewing (Moher et al., 2003), due to time constraints, only sources written in English will be considered.

After discussions with a second reviewer during and after first and second screenings, a final amended inclusion-exclusion criteria was used for this review, as illustrated below in Table 3.3.1:

Table 3.3.1 Eligibility Criteria

Item	Eligibility Criteria	Rationale
Bibliographic Information	**Include 1:** Studies with a full reference or sufficient information.	Retrieval of studies is unfeasible without sufficient bibliographic information.
	Exclude 1: Studies with insufficient bibliographic information.	
Date of Publication	**Include 2:** Studies published between 2011-2022	The current study continues in the vein in the work done by Plonsky (2011) and Hassan et al. (2005). Hence, only
	Exclude 2: Studies published before 2011	

		research after 2011 will be considered.
Language of Publication	**Include 3:** Papers written in English	Although research in multiple languages is a form of high-quality systematic reviewing (Moher et al., 2003), due to time constraints, only sources written in English will be considered.
	Exclude 3: Papers that are not written in English.	
Topic Focus	**Include 4:** Focuses on language learning strategies (LLS), strategy instruction (SI), specifically within L2 acquisition.	The focus of the current study is about research on adult L2 language learning strategies and strategy instruction. Anything unrelated to that research will be excluded.
	Exclude 4: Focuses on topics unrelated to language learning strategies, strategy instruction, pedagogy, or L2 acquisition. Topics dealing with language policy/curriculum rather than the actual teaching/acquisition of language will also be excluded. To further clarify, the study must have some sort of pedagogical merit for language teachers and/or some form of learning benefit for language learners. After reading the study, can something be changed/used in the language learning process? If no, exclude the study.	

Participants	**Include 5:** Participants are adult L2 language learners (university students are fine).	The current study is interested in only adult L2 language learners. Adults who are not engaged in L2 language learning and/or other populations will not be considered.
	Exclude 5: Non-adult participants. Non-adult L2 language learners. Adults who are not L2 language learners and/or not involved with language acquisition. Essentially, participants who are not adult L2 language learners and/or language instructors/teachers will be excluded.	
Intervention and Methodology	**Include 6:** Studies that contain primary empirical research. This research should have either qualitative or quantitative aspects of language learning strategy/strategy instruction effectiveness as it relates to one of the three research questions in the study.	As a systematic review, this current study seeks to explore new research and innovations in the field rather than re-hash what is already widely known in other reviews. Additionally, due to the considerations for future research in studies like Hassan et al. (2005) and Plonsky (2011), only studies with delayed posttests and/or longitudinal studies will be considered. This is in order to better explore the effect of language learning strategies and/or strategy instruction over
	Exclude 6: Theoretical articles, systematic reviews or reports on practice without research methodology (no primary empirical research.)	
	Include 7: Studies which contain delayed posttests and/or deal with longitudinal effects/interventions.	
	Exclude 7: Studies with no delayed posttests and/or studies which do not explore or observe the effect of language learning strategies/strategy	

	instruction over time (non-longitudinal studies or studies without repeated observations).	time, which is an area under-represented in literature.

3.3.2 Information sources

Information sources as defined by Moher et al. (2009) include all databases, journals, citations, and contact with study authors that were used as sources of evidence. For this study, based on similar research such as Plonsky (2011) and the recommendations by my supervisor, the following relevant databases were consulted:

- ProQuest
- Scopus
- EBSCO
- Web of Science (Core Collection)
- Linguistics and Language Behavior Abstracts (LLBA)
- Educational information Research Center (ERIC)
- Google Scholar

An initial search was conducted on November 11, 2021.

3.3.3 Electronic search

Brunton et al., (2012) notes that the ideal search locates as many relevant results as possible while at the same time limits the number of irrelevant results. To facilitate this ideal search, search terms were piloted and organized according to key relevant themes to the current study on adult language learning strategies and strategy instruction. Namely, themes concerning “skill”, “strategy”, “age”, “content learning” and “second language”. In following Brunton et al.’s (2012) recommendations, searches concerning “skill”, “strategy”, “age”, and “content learning” were limited to title and abstract (resulting in a more precise search) while searches concerning

“second language” was searched anywhere in the document to find the as many relevant results as possible. This search strategy (and subsequent narrowing based on the inclusion/exclusion criteria) reduced the number of retrieved results from over 10,000 studies to a few hundred – a more manageable amount.

Table 3.3.3 Search terms used in the current study

Search Terms								
Skill		**Strategy**		**Age**		**Content Learning**		**Second Language**
title and abstract	**AND**	**title and abstract**	**AND**	**title and abstract**	**AND**	**title and abstract**	**AND**	**anywhere in the document**
language-learning		strateg*		adult		listening		additional-language
language		instruction		young-adult		reading		EAL
language-learn		SI		university		writing		EFL
language-skill*		LLS		college		speaking		ESL
language-learning-strategy				adult*		vocabulary		foreign-language
						grammar		L2
						pronunciation		second-language
						spoken-communication		EAP
						proficiency		ESP
						post-test		ESOL
						post-intervention		bilingual
						self-regulation		multilingual
						empirical		
						vocab		
						vocab*		
						intervention		
						delayed-post-test		
						longitudinal		

A Boolean search string expressing the above search strategy resulting in 465 articles was conducted on February 9, 2022, in the Scopus database (limitations on publication years and/or languages was added to reduce the amount of non-relevant studies, as expressed in the eligibility criteria in Table 3.3.1):

TITLE-ABS (language-learning OR language OR language-learn OR language-skill* OR language-learning-strategy OR strategy-based-instruction OR sbi) AND TITLE-ABS (strateg* OR instruction OR si OR lls) AND TITLE-ABS (adult OR young-adult OR university OR college OR adult*) AND TITLE-ABS (intervention OR delayed-post-test OR longitudinal OR listening OR reading OR writing OR speaking OR vocabulary OR vocab OR vocab* OR grammar OR pronunciation OR spoken-communication OR proficiency OR post-test OR post-intervention OR self-regulation OR empirical) AND ALL (additional-language OR eal OR efl OR esl OR foreign-language OR l2 OR second-language OR eap OR esp OR esol OR bilingual OR multilingual OR strategy AND based AND instruction OR sbi) AND NOT ALL (child*

) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011)) AND (LIMIT-TO (LANGUAGE , "English") OR EXCLUDE (LANGUAGE , "Spanish") OR EXCLUDE (LANGUAGE , "Turkish") OR EXCLUDE (LANGUAGE , "Chinese") OR EXCLUDE (LANGUAGE , "Japanese") OR EXCLUDE (LANGUAGE , "Malay")) AND (EXCLUDE (EXACTKEYWORD , "Child*")) AND (EXCLUDE (SUBJAREA , "MEDI")) AND (EXCLUDE (SUBJAREA , "HEAL")) AND (LIMIT-TO (SUBJAREA , "ARTS"))

After removing duplicates (n=19) and coming up with bibliographic data for a final collection of n=419 articles, these articles were then exported as an RIS file and uploaded to Rayyan, an online platform and tool which helps facilitate systematic reviews (Ouzzani et al., 2016).

3.3.4 Study selection

To reduce the risk of systemic bias and ensure that inclusion-exclusion (eligibility) criteria was applied in a consistent manner, study selection took place in two stages. There was a first screening which looked at titles and abstracts, and a second screening which looked at the full text of the study (Higgins et al., 2019). A title and abstract screening saves time and allows reviewers to sort through a greater quantity of studies while a full text screening helps narrow down the studies further so that only relevant articles to the review are kept.

In accordance with Gough et al.’s (2012) features of a systematic review, a second reviewer was recruited, and she participated in both the first and second screening. Like the author of the study, the second reviewer was also a teacher-practitioner who was studying a master’s degree in Applied Linguistics for Language Teaching at

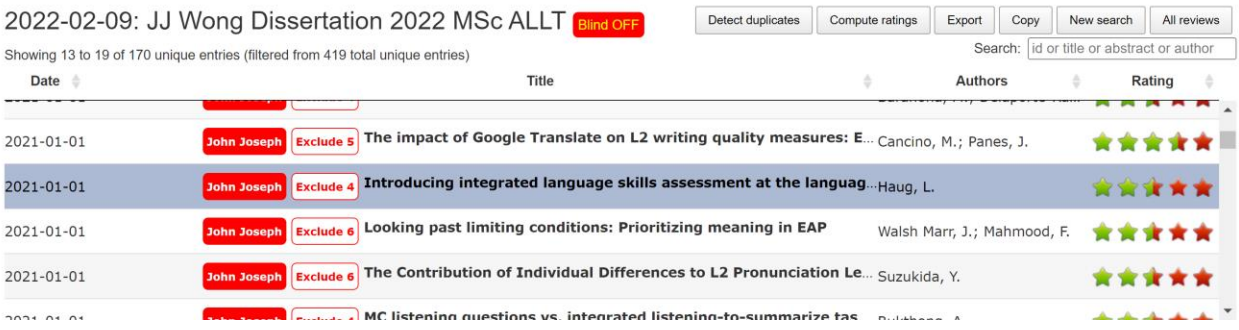
the University of Oxford. She first read the protocol document and discussed eligibility criteria and clarifications with the author during separate meetings. Although not researching the exact same topic, the second reviewer was also writing a dissertation on a similar topic dealing with language learning pedagogy. Hence, she was both knowledgeable in the applied linguistics and suitable as a collaborator.

3.3.5 Title and abstract screening (first screening)

Rayyan was used by both reviewers to screen the titles and abstracts of potential studies to be included in the current systematic review. Studies were originally searched in February 2022 and subsequently screened for the first time on June 18, 2022. These studies were included or excluded based on the Eligibility Criteria (listed in Table 3.3.1). After removing n=19 duplicate studies, a total of n=419 studies were screened. Of this number, n=233 was included and n=186 was excluded. Excluded studies were tagged with reasons on Rayyan, such as not fulfilling Criteria 5, dealing with the wrong population of participants.

To ensure inter-rater reliability, 10% of the studies considered in the first screening were randomly assigned to the second reviewer through a spreadsheet. The second reviewer then utilized the same eligibility criteria as the author during her first screening.

Screenshot 1: Rayyan Platform with Exclusion Reasons



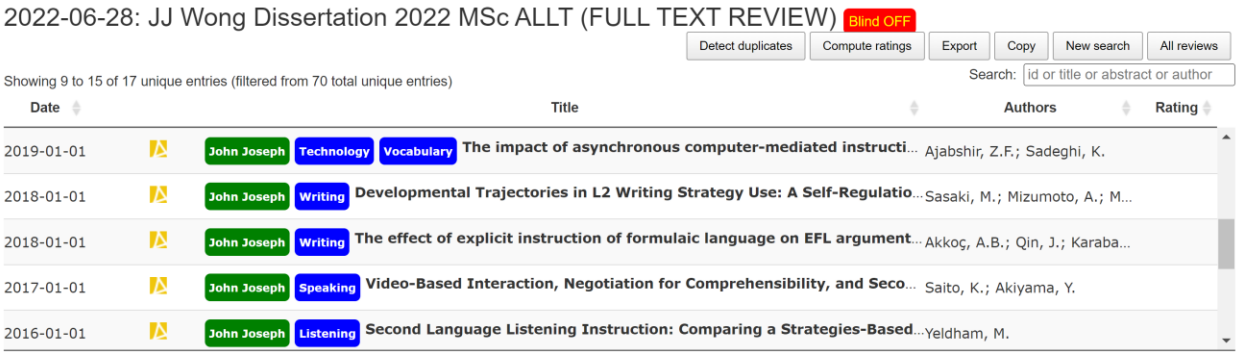
Any disagreements were discussed over virtual meetings, and the eligibility criteria was edited to further reduce and limit ambiguity. After reaching agreement with

regards to eligibility standards, the reviewers in the current study yielded a Cohen's kappa (Cohen's k) of 1 for the title and abstract first screening, indicating a high degree of agreement as well as inter-rater reliability between both reviewers.

3.3.6 Full-text screening (second screening)

For the second screening, full texts were retrieved through the University of Oxford's Bodleian Library online system. The second screening was performed on June 25, 2022. In an effort to focus on Hassan et al.'s (2005) and Plonsky's (2011) calls for more empirical studies involving delayed posttests and/or a longitudinal focus, as well as due to feasibility considerations, the reviewers agreed to further exclude studies which did not explicitly involve posttests and/or were not longitudinal in nature (n=163 more studies were excluded). As a result, n=70 studies were ultimately considered for the second screening.

Screenshot 2: Second Screening on Rayyan Platform



As in the first screening, 10% of the studies considered for the second screening were randomly assigned to the second reviewer to ensure inter-rater reliability. To better facilitate the screening process, PDF files of the studies were directly uploaded to Rayyan. There was initially only moderate agreement at 71.43% with a Cohen's k of 0.46, but after discussing and resolving disagreements caused by differing interpretations of the inclusion/exclusion criteria, full agreement was reached between the two reviewers.

Ultimately, n=15 studies were selected from the original n=70 for inclusion in this systematic review. These studies were then retained and used for data collection.

3.3.7 Data collection process and data items

For data collection and data items, a data extraction tool was used to increase both the consistency and transparency during the data extraction process (Li et al., 2019). To extract and organize relevant data from multiple studies, a data extraction tool adapted from a similar systematic review in Macaro et al. (2018) which followed the guidelines of Cochrane's good practice data extraction form (Cochrane…, 2017) was used. To adapt the data extraction form for the current study, references to EMI (English as a Medium of Instruction) were replaced by criteria regarding language learning strategies and/or strategy instruction.

The form used in the current study was used to collect relevant information from the studies ultimately chosen for data extraction after the first and second screening. This information included the context of the study, rationale, relevant literature review, research questions, methodology, data collection instruments, results, and findings, as well as an initial weighting of evidence (see Appendix C for a completed example).

3.3.8 Risk of bias (quality assessment) in individual studies

Boutron et al. (2019) notes how systematic reviews requires risk of bias assessments due to the ease at which systematic review results and conclusion may be swayed by biased results from the included studies. To counteract this risk, quality appraisal tools were used in this study such as a Weight of Evidence (WOE) framework following guidelines set by Gough (2007). Specifically, the Weight of Evidence framework utilized in Macaro et al.'s (2018) systematic review was adapted and applied in the current study due to the compatible subject material and

processes of both Macaro et al (2018) and the current study.

Table 3.3.8a: *Weight of Evidence framework adapted from Macaro et al. (2018)*

Weight of evidence		**High /Medium/Low**
WOE: **Relevance**	of particular **focus of the study** for addressing the question or sub-questions of this specific systematic review. Please add the number of the Review Question(s)	
WOE: **Appropriateness**	of research **design and analysis** for addressing the question, or sub-questions, of this specific systematic review.	
WOE: **Trustworthiness**	Taking account of all quality assessment issues, can the study findings be trusted in answering *the study* question(s)?	
WOE: **Contribution**	of the study to answer the question/s of this specific systematic review	

Another quality appraisal tool used in the current study was the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018). The MMAT was chosen as it was designed for the appraisal stage of systematic mixed reviews containing studies that may be qualitative, quantitative and/or mixed methods studies. Additionally, as the current review only includes empirical studies, the MMAT is a good fit as it was designed for primary research and cannot be used for “non-empirical papers such as review and theoretical papers” (Hong et al., 2018). Lastly, the ease of use of the MMAT, thoroughness in methodological quality criteria, and division of study designs into five separate categories (qualitative, quantitative randomized controlled trials, quantitative non-randomized, quantitative descriptive, and mixed methods) further enhances the appeal of the MMAT as the primary quality appraisal tool in the current study.

Table 3.3.8b: Part I of the *Mixed Methods Appraisal Tool (MMAT), version 2018 (Hong et al., 2018)*

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?				
	S2. Do the collected data allow to address the research questions?				
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
1. Qualitative	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?				
2. Quantitative randomized controlled trials	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?				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?				
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?				
	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 administered (or exposure occurred) as intended?				
4. Quantitative descriptive	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?				
5. Mixed methods	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 outputs of the integration of qualitative and quantitative components adequately interpreted?				
	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?				

3.3.9 Summary and synthesis of results

To synthesize results, reporting in the current study attempted to follow PRISMA reporting guidelines (Moher et al., 2009), which have been used in systematic reviews and meta-analyses to obtain higher methodological quality as well as better reporting standards (Panic et al., 2013). Because the data included in the current study could not be easily extracted and presented in the way suggested by Page et al. (2021b), the PRISMA reporting guidelines were adapted accordingly, and a narrative approach (Gough et al., 2012) was utilized to report data when appropriate. Both the PRISMA flow diagram (see Chapter 4: Results) and the PRISMA checklist (Appendix B) were used and adapted when needed in the current study.

Miake-Lye et al. (2016) notes that there is no best practice for how evidence maps should be reported in a systematic review. Hence, research contexts, participants, and methodological designs of included studies will be presented in a tabular format in Chapter 4 Results to best convey the aforementioned information. Results of the MMAT used to appraise the included studies will also be presented in a tabular format.

Additionally, key findings of the review will be presented in a narrative fashion to synthesize both qualitative and quantitative data, in accordance with suggestions by Boland et al. (2017); Snilstveit et al. (2012); and Thomas et al. (2012). Although a narrative approach has its critics and drawbacks, when used in conjunction with WOE frameworks and a trusted critical appraisal tool (MMAT), a narrative approach can be an invaluable part of a systematic review executed in a rigorous and transparent manner.

Finally, results will be synthesized and organized according to the research questions of the current review, so that the data from the included studies can be synthesized and integrated toward answering the research questions.

Chapter 4 Results

The results of the systematic review are reported in this chapter. The process of selecting the final included studies is shown through a modified PRISMA (2020) flow diagram (Page et al., 2021). Descriptions of the included studies are presented. Finally, overall quality of studies and methodological quality of studies are assessed.

4.1 Included studies

Figure 4.1: (modified) PRISMA (2020) Flow Diagram for Searching and Screening (Page et al., 2021)

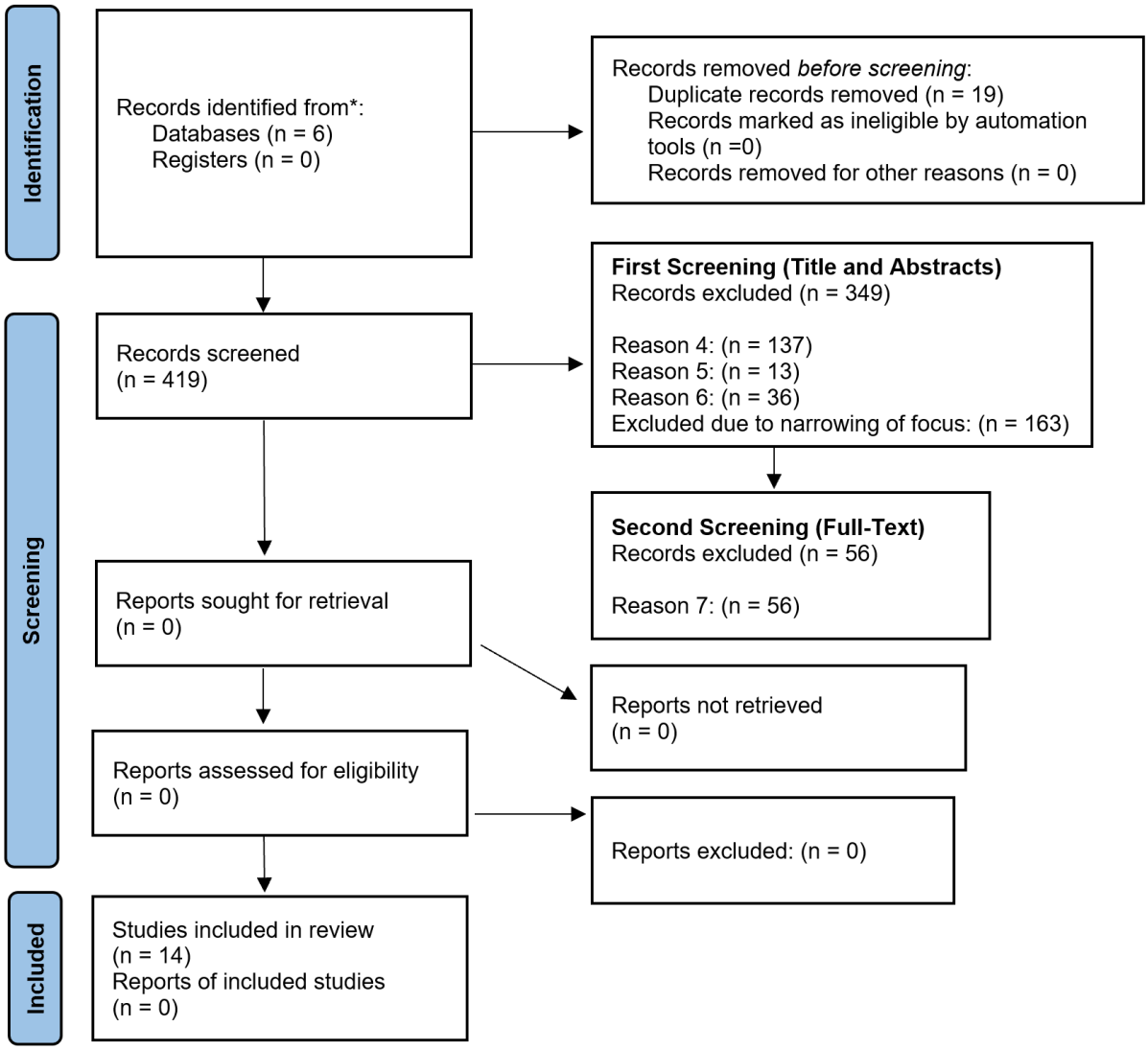


Figure 4.1 summarizes the results of searching and screening at each stage of the systematic review process. Initially, electronic searching with consultation from multiple databases resulted in 438 records. Of these, 19 records were removed as duplicates, leaving 418 records for the initial first screening of titles and abstracts. Further exclusion of 349 records resulted in 70 records considered for the second full-text screening. During this full-text screening, the texts were found using online university library access, and 56 records were excluded. Ultimately, 14 studies were chosen for data extraction, summary, and synthesis in this present systematic review.

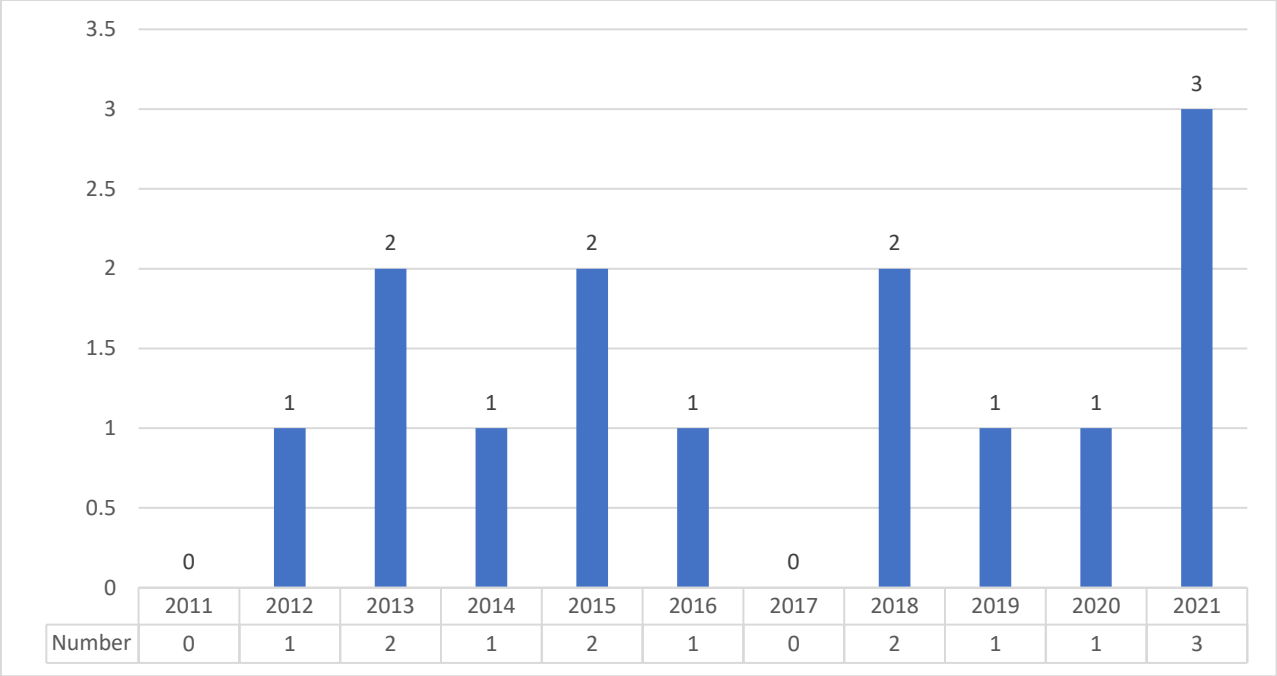
4.2 Characteristics of included studies

4.2.1 Publication details

Due to the scope of the present study, only studies published after 2011 were considered. Figure 4.2.1 shows the studies per year of the included studies in this present systematic review. While the numbers may be skewed due to the present study's insistence on focusing on studies with delayed posttests and/or longitudinal focuses (as recommended by previous reviews such as Plonsky (2011)), it is clear that the data overall suggests a greater interest in language learning strategies and strategic instruction for adult L2 learners.

Crucially, it appears that during the publication period of 2011-2021, there is a general increase in both longitudinal studies and studies attempting to utilize delayed posttests to analyze the effects of language learning strategies and/or strategic instruction over extended periods of time.

Figure 4.2.1 Studies per Year (n=14)



While the present study may have benefited from exploring grey literature, Table 4.2.1 clearly indicates that only journal articles made up the final included studies. The insistence of only considering empirical, non-theoretical publication types surely contributed to these findings. Additionally, the electronic search strategy may also have minimized the possibility of finding publication types other than journal articles.

Table 4.2.1 Studies by Publication Type (n = 14)

Publication Type	**Number of Studies**
Journal Article	*14*
Dissertation or Thesis	0
Conference Report	0
Book Chapter	0

4.2.2 Research contexts

While all studies explored language learning strategies and/or strategic instruction for adult populations, most studies consisted of young adult participants ranging from 18-21 as shown in Table 4.2.2a. This can be explained due to the fact that a

majority of studies were conducted in or funded by universities, meaning that undergraduate students were the easiest population to obtain for research.

While having most studies include participant populations aged 18-21 is fine for the present study, this could prove to be a limitation as well as an avenue for future research. Most studies included utilized participants who were current university students. This may make their learning and/or life contexts different from older adults who are no longer engaged in full-time education or have been away from education for a while.

Future research could explore language learning strategies and strategic instruction on participants who are older or have been out of school for a while. This population group may be overlooked in literature, but due to their L2 language needs for career, family, and/or immigration purposes, they would greatly benefit from the aforementioned research.

Table 4.2.2a Studies by Age (n = 14)

Age	**Number of Studies**
18 - 21	9
21 - 24	1
25 – 30	2
30 +	1
Unknown	1

Within the 14 included studies, most studies (n = 9) utilized a mixed participant population of both males and females, as shown in Table 4.2.2b. While this does not guarantee the elimination of gender as a potential variable, it does indicate that the conclusions drawn from the present study can be broadly (albeit cautiously) applicable to both male and female second language learners.

Table 4.2.2b Studies by Participants' Gender (n = 14)

Gender	**Number of Studies**

Not Stated	5
Mixed	9
Male	0
Female	0

Regarding countries where research takes place, the studies included were scattered fairly evenly among countries primarily in Asia and North America as shown in Table 4.2.2c. Taiwan, Canada, Japan, and the USA was where most of the studies were conducted.

This result may be explained by the fact that “Northern America” (Canada and the USA) has “historically been one of the principal destination regions of international migrants worldwide” (*Migration Data in Northern America*, 2021). Hence, a pragmatic need for education and research on second language acquisition.

On the other hand, second language acquisition, particularly for English, is essential in the Asian context as various Asian nations assert themselves as part of a globalized world (Chang, 2011). Specifically, acquiring English within the Asian context is seen as a “way of speeding up national development, understanding other cultures, and as a tool for international communication” (Chang, 2011).

Future research may benefit from exploring language learning strategies and strategy instruction in other regions around the world such as Africa, Europe, South America, and Oceania.

Table 4.2.2c Studies by Country (n = 14)

Country	**Number of Studies**
Taiwan	3
Canada	2
Japan	2
USA	2

UK	1
China	1
Iran	1
Turkey	1
Vietnam	1

As the present study focuses on second language acquisition without narrowing down to a specific L1 or L2 language, it is important that different L1 languages were represented by participants among the included studies. Table 4.2.2d shows the different L1 languages represented. Note that some studies had participants who spoke different L1 languages, which explains why there are more instances of L1 languages represented than the total number of included studies.

Mandarin Chinese seems to be the most represented L1 language within the included studies, followed by English and Japanese. Overall, it appears that most of the world's largest language families are represented, such as Sino-Tibetan, Indo-European, and the Afro-Asiatic language families. Only one Austronesian language (Malay) and one Niger-Congo language (Shona) is represented, indicating that future research could recruit participants whose L1 has been underrepresented in language learning strategy and strategy instruction research.

Table 4.2.2d L1 Languages Represented in Included Studies (n = 14 studies, but some studies have participants with different L1 languages)

L1 Language	**Number of Studies Where the L1 is Represented Among Participants**
Mandarin Chinese	8
English	3
Japanese	3
French	2
Turkish	2
Tagalog	2

Ukrainian	2
Russian	2
Arabic	2
Spanish	1
Italian	1
Portuguese	1
Shona	1
German	1
Mongolian	1
Greek	1
Dutch	1
Cantonese	1
Malay	1
Croatian	1
Vietnamese	1
Polish	1
Farsi	1

Table 4.2.2e shows the studies by L2 language among the included studies. Most studies are concerned with learning English as an L2. While this may in part be due to English's influence as a global language as well as the electronic search strategy (studies not written in English were excluded from the search), it would be important for future research to explore other L2 languages to ensure that language learning strategy and pedagogical approaches to strategy instruction are not language dependent.

Table 4.2.2e Studies by Second Language (L2) (n = 14)

L2 Language	**Number of Studies**
English	12
Korean	1

German	1

When looking at L2 proficiency level among the participants in the included studies, Table 4.2.2f shows that there are a wide range of proficiency levels represented, from beginner to advanced. With 3 studies dealing with upper-intermediate to advanced learners, Plonsky's (2011) call for more studies investigating advanced learners has been answered.

Only a few studies in this review, such as Akkoç et al. (2018) and Zhang (2021) used internationally recognized language tests such as the IELTS and TOEFL to quantify language proficiency scores. This means that while study conclusions may be applicable to different proficiency levels, it cannot be said with certainty that all the included studies view language proficiency with the same scale (such as in the studies which deal with languages other than English). Nonetheless, the wide range of language proficiencies included in this review helps alleviate the fear the outcomes of the studies are only applicable to language learners with specific proficiency levels.

Table 4.2.2f Studies by Second Language (L2) Proficiency (n = 14)

L2 Language	**Number of Studies**
Beginner	3
Lower-Intermediate	2
Intermediate	3
Upper-Intermediate	1
Advanced	2
Not Stated	3

Participants in the included studies come from a wide range of diverse disciplinary backgrounds. Unfortunately, as shown in Table 4.2.2g, the majority of studies do not indicate the primary discipline of participants. On a positive note, it appears that among the studies which did indicate participants' disciplines, no discipline (perhaps

with the exception of English-related disciplines) was overly represented, indicating that the conclusions drawn from this review may be applicable to various disciplines.

Table 4.2.2g Studies by Discipline of Participants (n = 14 studies, but some studies have participants with different disciplines)

Discipline	**Number of Studies**
Non-English Majors	1
Psychology	1
English-Majors (Interpretation)	1
Department of Library Science and Textile Management	1
Health Sciences	1
English Language Teaching	1
British and American Studies	1
Not Stated	8

Among the included studies, a large number of them had fairly large participant samples, with eight studies consisting of over 41 participants, and four studies consisting of over 60 participants, as shown in Table 4.2.2h.

Studies with between 0-10 participants tended to be qualitative and deliberately small such as Yeldham and Gruba (2013), as well as Schneider (2021). Schneider (2021) and Sasaki et al., (2018) (which had more participants, n = 37) were noteworthy as they were both longitudinal studies on writing strategies which covered the longest timeframe (four years) among included studies.

A positive of the larger participant size of most studies included in this review is that sampling size should not be a primary limitation when considering study results. While samples could always be larger, especially when dealing with such an over-arching topic like language learning strategies and strategy instruction, it would be a greater concern if the majority of studies (especially quantitative studies) in this review had small participant sizes.

Table 4.2.2h Participants per Study (n = 14)

Number of Participants	Number of Studies
Not Stated	1
0 – 10	2
11 – 20	0
21 – 30	1
31 – 40	2
41 – 50	3
51 – 60	1
60+	4

The topic focus of the various studies included in this review is of particular interest when considering previous relevant literature such as Plonsky (2011) and Hassan et al. (2005). Table 4.2.2i indicates that Grammar, Speaking, Writing, and Listening were all key topic focuses. This corresponds with Plonsky's (2011) earlier calls for future research, arguing that more research must be done on strategy instruction's effects on writing, listening, pronunciation, and grammar. Additionally, the present review also adds to Hassan et al. (2005) which indicated limited evidence for improving speaking through language learning strategies and/or strategic instruction.

Table 4.2.2i Studies by Topic Focus (n = 14)

Topic Focus	Number of Studies
Grammar	4
Speaking	3
Writing	3
Listening	2
Learner Autonomy	1
Other (Orthography)	1
Reading	0

The breakdown of studies by methodology type is shown in Table 4.2.2j. The majority of studies in this review utilized quantitative, and/or mixed-methods methodology. Descriptive statistics were applied during data analysis of the majority of studies included in this review, which is in line with both Plonsky's (2011) and Hassan et al.'s (2005) calls for more thorough reporting of data, as well as higher quality reporting through basic descriptive statistics, sample sizes and measurement reliability.

Table 4.2.2j Studies by Methodology Type (n = 14)

Methodology Type	**Number of Studies**
Quantitative	8
Mixed-Methods	5
Qualitative	1

A major finding of previous reviews in language learning strategy and strategy instruction was the noteworthy absence of "long-term, post-intervention testing or follow-up" (Hassan et al., 2005). Similarly, Plonsky (2011) echoes these sentiments by calling for more studies which have "additional measurements of the persisting effects of strategic instruction". In response, this present study narrowed its focus to only include empirical research which either contained a delayed posttest within its methodology and/or were longitudinal in nature.

Table 4.2.2k shows the methodology of included studies with regards to their use of delayed posttests or longitudinal observations. As shown below, there was a fairly even mix of studies which used delayed posttests (n = 7) and longitudinal studies (n = 7).

Among studies with a delayed posttest, the minimum interval between intervention and delayed posttest was 2 weeks, yet the most popular approach was issuing a delayed posttest one month after the intervention, with five of the included studies doing so. It is heartening that more delayed posttests exist are being carried out in research, supporting Plonsky's (2011) insistence that the benefit of learner

strategies “depends greatly on whether and to what extent its effects last over time”. However, future research may benefit from incorporating delayed posttests at longer intervals after intervention (such as a year or longer). This would be crucially important in terms of pedagogic effectiveness as it would provide key support for policy decisions regarding “the likelihood of long-term benefits of strategy training” (Hassan et al., 2005).

Seven studies included in the review contained longitudinal methodologies. The minimum timeframe among longitudinal studies was 0.5 years (1 semester), while the longest time frame was over 4 years. Among these studies, the two studies with the longest timeframes (Schneider (2021) and Sasaki et al (2018)) were noteworthy in being non-experimental in nature, responding to Hassan et al.’s (2005) call for non-experimental research which “is particularly important in order to understand the full detail of processes in action during strategy training and learner strategy development”.

Table 4.2.2k Studies with Delayed Posttests and/or Longitudinal Methodologies (n = 14)

Inclusion of Delayed Posttests and/or Longitudinal Methodology	**Number of Studies**
Delayed Posttest (2 Weeks After)	2
Delayed Posttest (1 Month After)	5
Longitudinal Study (0.5 – 1 Year)	3
Longitudinal Study (2 Years)	2
Longitudinal Study (4 Years)	2

Concerning strategy elicitation methods used in the included studies, Table 4.2.2l shows that questionnaires and semi-structured interviews were the most popular strategy elicitation methods. “Other” methods included observations, diaries, journals, and other types of interviews such as retrospective and unstructured interviews were also represented in the strategy elicitation methods present.

Table 4.2.2l Studies by Strategy Elicitation Method (n = 14 total studies, but some studies are represented more than once as multiple strategy elicitation methods were used)

Strategy Elicitation Method	Number of Studies
Questionnaire	5
Semi-Structured Interview	5
Other	4
N/A	2
Retrospective Interview	2
Unstructured Interview	1
Naturally Occurring Conversations	1
Verbal Reports	1
Think-Aloud Protocol	0
Stimulated Recall	0

A range of task types were used in the included studies. Table 4.2.2m illustrates how writing tasks such as writing letters of application (Reynolds & Kao, 2021) and speaking tasks like visual-based conversation (Saito & Akiyama, 2016) were the most used task types in the studies.

Listening tasks such as transcription and dictation (Yeldham, 2015), reading tasks like reading and matching (Rahimi et al., 2020), and grammar tasks such as error correction (Smart, 2014) were also represented in this present study.

Table 4.2.2m Studies by Task Type (n = 14)

Main Task Type	Number of Studies
Writing Task	4
Speaking Task	3
Listening Task	2
Reading Task	2

Grammar Task	2
N/A	1

4.2.3 Summary of characteristics of included studies

In summary, when considering the different publication details and research contexts of the included studies relating to language learning strategy and strategy instruction research for L2 adult learners, it can be said that:

- From 2011 to 2021, while the numbers do not increase dramatically, research in this area which includes either a longitudinal focus and/or delayed posttests in this field has overall been trending upwards.
- Most studies have participants within the age range of 18-21, however, there are studies with participants ranging in ages from 21-30 and even 30+.
- Most studies have mixed-gender participant populations, indicating that little evidence exists concerning gender as a potential variable in language learning strategy or strategic instruction.
- Regarding the country where research takes place, Taiwan, Canada, Japan, and the USA are most frequently represented. Most of the research seems to be conducted in North America and Asia. Europe, Africa, South America, and Oceania are not well represented.
- Participants represented 23 different L1 languages. Of these, Mandarin Chinese was the most widely spoken. Most participants had a Sino-Tibetan, Indo-European or Afro-Asiatic L1 language, with only one Austronesian language (Malay) and one Niger-Congo language (Shona) represented. Little evidence exists concerning the relationship between L1 and language learning strategy or strategy instruction, but future research could explore more underrepresented participant L1 languages.
- Among L2 languages, the vast majority (twelve out of fourteen) of studies researched English. While other L2s like Korean and German are represented, more research needs to be conducted on other L2s.

- In terms of L2 language proficiency, the studies in this review consisted of an even spread of different proficiency levels ranging from beginner to advanced.
- Participants studied a wide range of academic disciplines. Although most participants studied a major that was related to English, other disciplines like health sciences and library science and textile management were also represented.
- Except for qualitative and mixed-methods studies focusing on small participant samples, most studies had 41 or more participants, with the 60+ participant category being the most common.
- Language learning strategy and strategy instruction related to grammar, speaking, and writing were the most common topics studied. Listening, learner autonomy, and orthography were also studied. Interestingly, no studies included in the review explicitly focused on reading.
- Most studies used a quantitative methodology. Mixed-methods studies were also well-represented. Only one study used a qualitative methodology.
- Among studies that used a delayed posttest, having the delayed posttest conducted a month after intervention was most common, followed by having the delayed posttest conducted two weeks after intervention.
- Among longitudinal studies, three studies had a time frame of between half a year to one year, two studies had a time frame of two years, and two studies had a time frame of four years.
- Questionnaires and semi-structured interviews were the most common strategy elicitation method. Other methods included retrospective interviews, unstructured interviews, and naturally occurring conversations.
- Writing, speaking, and listening tasks were the most common task types in the studies. Reading and grammar tasks were also represented.

4.3 Quality assurance of included studies

To assess the overall quality of included studies, a Weight of Evidence (WoE)

framework utilized in previous systematic reviews exploring similar topics (Macaro et al., 2018) was adapted and incorporated in this review.

Table 4.3 shows the WoE assessment for quality assurance of included studies (n = 14). Regarding the first question of “relevance” to this specific systematic review, the review’s research questions are listed again:

1. What empirical research was carried out on adult L2 language learning strategies and strategy instruction between 2011 and 2022?

2. What evidence is available for the effectiveness of language learning strategies and/or strategy instruction for adult L2 language learners in terms of language proficiency?

3. What evidence is available for how language learning strategy use has changed over time for adult L2 language learners?

While there was a mix of studies with different levels of quality assurance, it appeared that most studies were ranked high regarding relevance, appropriateness, trustworthiness, and contribution to the present review.

Table 4.3 Weight of Evidence (WoE) Assessment for Quality Assurance of Included Studies (n = 14)

Study	**WoE: <u>Relevance</u> (of particular focus of the study for addressing this specific systematic review. Relevant Review Question(s) are identified)**	**WoE: <u>Appropriateness</u> (of research design and analysis for this present review)**	**WoE: <u>Trustworthiness</u> (can the study findings be trusted?)**	**WoE: <u>Contribution</u> (of the study to answer the question(s) of this specific systematic review)**

(Akkoç et al., 2018)	Medium (1, 2,)	Medium	Medium	Medium
(Brown, 2012)	High (1, 2, 3)	High	High	High
(Heift, 2019)	Medium (1, 2)	High	Medium	Medium
(Munro et al., 2015)	Medium (1, 2)	Medium	Low	Low
(Nguyen & Gu, 2013)	High (1, 2)	High	High	High
(Rahimi et al., 2020)	High (1, 2)	High	High	High
(Reynolds & Kao, 2021)	High (1, 2)	High	Medium	Medium
(Saito & Akiyama, 2016)	Low (1, 2)	Low	Low	Low
(Sasaki et al., 2018)	High (1, 2, 3)	High	Medium	High
(Schneider, 2021)	High (1, 2, 3)	High	High	Medium
(Smart, 2014)	High (1, 2)	Medium	Medium	Medium
(Yeldham & Gruba, 2013)	High (1, 2, 3)	Medium	Medium	High
(Yeldham, 2015)	High (1, 2)	High	Medium	High

(Zhang, 2021)	High (1, 2)	High	Medium	High

4.3.1 Weight of evidence assessment for high overall quality studies in relation to the present review

Most included studies had high overall quality regarding its WoE assessment in relation to the present review.

4.3.1a Brown (2012) – Teaching Korean hangul through mnemonics

Brown (2012) explored the teaching of Korean Hangul to 84 participants from a variety of L1 backgrounds, using visual/verbal and physical mnemonics. The study was noteworthy for being one of the few studies which explored an L2 language other than English, and it was also the only study which investigated Korean as the target L2. The study had high relevance to every research question in the present review as it focused on mnemonics as a form of language strategy and a pedagogical tool.

Results showed that utilizing mnemonics resulted in clear statistically significant advantages during immediate recall (p <.0001), and statistically significant advantages during posttest delayed recall 4/5 weeks after instruction. During posttest delayed recall, the advantage that the mnemonics group had over the “design principles” group (a mainstream pedagogy for teaching Korean Hangul) was smaller, but still had a mean total recall/production rate of 90.6% compared with the 76.1% of the design principles group.

Brown (2012) contributes greatly to the present study by demonstrating the effectiveness of mnemonics as both a language learning strategy and/or pedagogical tool which has been demonstrated to be effective not only immediately after instruction, but also after a delayed period.

4.3.1b Nguyen and Gu (2013) – Strategy-based instruction, writing proficiency and learner autonomy

Nguyen and Gu (2013) investigated how strategy-based instruction helps develop writing proficiency and learner autonomy (LA), which has been studied for over three decades in language learning. While many claims have been made about benefits of LA, there was a gap in literature for evidence concerning the “effectiveness of initiatives designed to help learners become more autonomous” (Benson, 2007). As such, Nguyen and Gu (2013) was highly relevant to this present study as it empirically explored how or whether metacognitive self-regulation strategies affected written English proficiency and LA development.

91 participants at a university in Vietnam were part of the study. They were divided into two control groups (normal writing lessons) and one experimental group (taught metacognitive self-regulation strategies). The mixed methods design and inclusion of a delayed posttest was appropriate for the present review. Through writing tests such as a comparison and contrast essay, and an argumentative essay, the study found that training with strategic instruction (metacognitive self-regulation strategies) resulted in enhanced self-regulation and enhanced writing performance in a way that was sustainable over six weeks after instruction.

While all groups demonstrated gains in their writing scores, descriptive statistics and ANOVA comparing the mean scores of the post-writing test indicated that the experimental group received the greatest mean score of 7.348, with an increase of 1.515 from the pretest. This is compared with the scores of control group 1 (5.7760) and control group 2 (6.500), which both achieved increases of 0.0640 from their respective pretests (All p values for the experimental group and the two control groups were $<.05$). Although all groups suffered a drop in score during the delayed test, the experimental group outperformed and maintained their edge over both control groups. As the delayed test used a different type of writing test (argumentative essay rather than a comparison and contrast essay), the results indicate that teaching metacognitive self-regulation strategies is not only effective

over time for improving specific writing tasks but may also have its beneficial effects transferrable to other forms of writing proficiency.

Regarding learner autonomy, students who received the strategic instruction "enhanced their ability to plan, monitor, and evaluate a writing task" (Nguyen & Gu, 2013), with planning being the most used skill. These results were corroborated using both descriptive statistics and group interviews. Trustworthiness and contribution of Nguyen and Gu (2013) to the present review was high due to its systematic methodology, descriptive analysis, and relevance in illuminating the benefits of teaching self-regulation strategies within the context of writing proficiency and learner autonomy.

4.3.1c Rahimi et al. (2020) – Jigsaw vs. input tasks and grammar development

Rahimi et al. (2020) investigated how pedagogy such as jigsaw (unfocused tasks) versus input flood (focused tasks) would affect grammar development in terms of accuracy and durability of L2 learners. This study was highly relevant to the present review due to its pragmatic application in the classroom. Participants involved 62 Iranian university students aged 19-21. A quantitative methodology with a delayed posttest was appropriate with regards to the present review.

The study was highly trustworthy as it utilized data collection instruments such as the Oxford Placement Test (OPT) and the Untimed Grammatically Judgment Test UGJT which were both validated in literature. Multiple statistical tests such as MANOVA, MANCOVA, ANCOVA, and Levene's test were utilized to limit potential influence of confounding variables, again adding credence to the study's trustworthiness.

Results indicated that learners in the experimental group (input flood tasks, or focused tasks) did better than the control group (jigsaw tasks, or unfocused tasks), although both groups improved in the immediate posttest. Interestingly, in the delayed posttest 2 weeks later, the jigsaw task group showed better long-term recognition than the input flood group.

Overall, the study concludes that teaching with both jigsaw and input flood tasks would best benefit students. However, one limitation of the study was its decision to measure learner production by measuring the improvement or usage of only a specific grammar form (regular past tense /-ed/). Future research may incorporate written production tests as a gauge of improved proficiency, and may also benefit from utilizing participants who studied a wide range of disciplines (all participants in this study were psychology majors).

4.3.1d Sasaki et al. (2018) – Writing strategy use and self-regulation

Another study which looked at writing strategy use, and self-regulation was Sasaki et al. (2018). This study was relevant with regards to every research question in the present review. It was a longitudinal study conducted over 4 years with 37 Japanese university students aged 18 when the study began. The study's choice of explanatory sequential mixed methods design (Creswell, 2014) was appropriate in answering its research questions concerning L2 writing strategy use, strategy use changes over time, and/or participant's emic (insider) accounts of their strategy use.

Results indicated that as learners became more proficient, their use of Global Planning (GP) as a writing strategy increased. Interestingly, while there was no real pattern, high L2 writing ability positively related to Local Planning (LP), which does not match findings of previous studies that indicated LP being related to mainly low-proficiency writers (Manchón et al., 2009). This result may be explained by students with higher writing proficiency more effectively using both GP and LP.

However, this discrepancy between Sasaki et al. (2018) and literature impacts its trustworthiness, although it could be a fruitful area for future research. Finally, there was a positive effect of both GP instruction and study abroad experiences, which reinforces key-findings of other self-regulation studies like Schunk (2001). Sasaki et al. (2018) contributes to the present review by providing long-term evidence from a self-regulation theory perspective regarding the importance of teaching and developing L2 writing strategies such as global planning.

4.3.1e Schneider (2021) – L2 learners' writing strategies as acts of identity

Schneider (2021) also looked at L2 learners' writing strategies, but as acts of identity rather than in the context of self-regulation. This study was a longitudinal study over four years done with a qualitative methodology and only seven advanced university L2 English learners of different nationalities (four by the end of the study).

The relevance and appropriateness of Schneider (2021) was clear because of its naturalistic, longitudinal observation of L2 student writing strategy use over time. Its trustworthiness could also be assured as the data collection and analysis was trialed by previous studies such as Charmaz's (2006) grounded-theory approach.

Results of Schneider (2021) showed that all participants used growth, investment, and survival strategies at some stage over the four years. These strategies were chosen to align with the students' goals and identities, which was a new finding in research. Importantly for instructors and institutions, Schneider (2021) illustrated how important education contexts are for learning. Hence, the responsibility of creating positive pedagogical conditions which encourage the use of growth and investment strategies (rather than survival strategies) falls on the educational institution as well as instructors.

Schneider (2021) contributed to the present study by indicating how simply instructing students to utilize a particular strategy was not enough, and that the educational incentives must be aligned with student goals and identities. Further research must be made into how to practically apply this knowledge, hence explaining the medium contribution of Schneider (2021) to the present review.

4.3.1f Yeldham (2015) – Strategy-based vs. interactive, strategies/bottom-up skills approaches to listening instruction

A strategy-based approach was compared with an interactive, strategies/bottom-up skills approach (strategies plus bottom-up skills) with regards to listening instruction

in Yeldham (2015). 77 Taiwanese lower-intermediate to intermediate listening proficiency English L2 university students participated in the longitudinal study over one and a half semesters.

Yeldham (2015) answered most of the research questions in this review and utilized a quantitative methodology that was appropriate in answering the study's research questions as well as the questions of this present review. While this was one of two included studies without explicit research questions, the research questions were implied through the study's hypotheses. Namely, that the interactive course would develop listening comprehension more than the strategies course, strategies and affect-related learner characteristics would gain more from the strategies course than the listening course and bottom-up skills would develop more in the interactive course than the strategies course.

With the participants split between a strategy instruction group and an interactive instruction group, statistical analysis of data revealed that contrary to expectations, the strategies group outperformed the interactive group in listening comprehension. Regarding strategic ability and learner characteristics, the strategy group outperformed the interactive group. Finally, the interactive group outperformed the strategy group in developing bottom-up listening skills such as recognizing segments, identifying grammatical functions, accentual intonation, and recognizing words in a text.

One limitation of the study which made its trustworthiness "medium" rather than "high" was the absence of a lack of control group to compare with both the strategies instruction group and the interactive instruction group. More precise data may also have been acquired with the use of additional instruments such as the Metacognitive Awareness Listening Questionnaire (MALQ; Vandergrift & Goh, 2012). Despite its limitations, Yeldham (2015) offered a high contribution to this present review due to its pedagogical implications in offering empirical evidence to support more focus on strategy instruction over an interactive approach when it comes to L2 listening, particularly for lower-intermediate level learners.

4.3.1g Zhang (2021) – Pragmatics, computer-mediated communication and data-driven instruction

The development of pragmatics (specifically compliment responses) with regards to computer-mediated communication (CMC) and data-driven instruction was explored in Zhang (2021). The study filled a gap in literature concerning the dearth of empirical data on computer-mediated communication. 59 Chinese university sophomores who were English L2 learners participated in the study. The study utilized a mixed methods approach to obtain both statistical data as well as qualitative data to expand upon the quantitative data collected. The study was methodologically sound, with trained linguistics PhD raters to further increase coding reliability.

An experimental group received two sessions of data-driven instruction (designed with guidelines from Taguchi (2015) about L2 pragmatics, including conscious-raising activities, meta-pragmatic explanations, and corrective feedback.) with text-based CMC Skype interaction with native English speakers. The control group received text-based CMC Skype interaction with native English speakers and regular instruction based on the university's curriculum.

Results indicated that CMC plus data-driven instruction was more effective than just CMC with the regular curriculum, and that the effects were long-term, as confirmed by delayed posttests four weeks after the intervention in the form of retrospective interviews.

One limitation and potential confounding variable may be that the experimental group's improvements might have been due to them spending more time specifically practicing CRs than the control group. Though the study's trustworthiness was rated medium, its contribution to the present review was high. This was because Zhang (2021) provided empirical evidence that raising students' pragmatic awareness as part of a data-driven approach was a form of effective pedagogy that could be used to teach pragmatics, especially in learning scenarios which lack a naturalistic L2 environment.

4.3.2 Weight of evidence assessment for medium overall quality studies in relation to the present review

4.3.2a Akkoç et al. (2018) – Explicit teaching of formulaic language and writing quality

Akkoç et al. (2018) explored “explicit teaching of formulaic language” and how it would affect L2 argumentative writing quality. The study investigated 85 Turkish students aged 18-21 and used pretests, posttests, and delayed posttests during data collection.

Findings indicated that the argumentative writing scores of those who were taught formulaic language through explicit instruction increased significantly. Even though there was a slight decrease in mean scores during the delayed posttest a month later, mean scores were still higher than the pretest, indicating a lasting effect of explicit formulaic language instruction on argumentative writing quality.

While the results of the study had some implications and merits regarding language pedagogy and strategy instruction (in this case, more about the process of teaching), the study seemed to focus on topics tangential to the present systematic review, such as formulaic language rather than any particular language strategy or form of strategic instruction. Moreover, the trustworthiness of Akkoç et al. (2018) may be questioned due to its findings of significant difference in the quality of written essays between students retrieving treatment and those who did not contradicts previous literature such as Čolović-Marković (2012).

4.3.2b Heift (2019) – Meta-linguistic vs. repetition feedback approaches for German grammar

Heift (2019) was the only included study which investigated German L2 learners. In the 2-year long longitudinal study, two technology-mediated feedback types were examined, meta-linguistic feedback (which identified learner mistakes and provided error explanations without revealing the correct answer) and repetition (which prompted learners to correct mistakes without any error explanation), with regards to error correction and performance of German L2 learners. As the study focused on feedback, it had a medium rating of relevance to the present review, even though it did have pedagogical implications which could be applied during strategy instruction or language instruction in general. 42 university German L2 students participated in the study from a wide range of different nationalities and were randomly given either meta-linguistic or repetition style feedback.

Results of the study as shown through analysis of variance using two-way repeated measures ANOVA indicated that meta-linguistic feedback led to significantly higher correct responses than repetition. The methodology and data analysis of Heift (2019) was deemed sufficiently appropriate for the present review. Regardless of feedback type, both groups scored higher in the introductory German courses than in the intermediate German courses. However, more errors appeared in German 201 rather than German 202, even though German 202 was more complicated and challenging. This result was not well explained by the researchers and, in conjunction with the lack of consideration for other potential confounding variables such as learner attitudes and proficiency levels, resulted in a “medium” rating for trustworthiness.

In terms of contribution to the present study, Heift (2019) provided empirical evidence for the merits of both meta-linguistic feedback and repetition-feedback, with meta-linguistic feedback having more effectiveness. However, due to the lack of further explanation about these pedagogical implications as well as further investigation required, the contributions of Heift (2019) to the present study can only be “medium” as further investigation is required for this topic to be effectively implemented in teaching pedagogy.

4.3.2c Smart (2014) – Data-driven learning, deductive corpus-informed instruction, and traditional grammar instruction

Different approaches to L2 grammar instruction such as data-driven learning, deductive corpus-informed instruction, and traditional grammar instruction were explored in Smart (2014). The topic matter and focus on grammar pedagogy was highly relevant to the present study.

49 English L2 participants with Chinese or Arabic L1s were divided into three instructional treatment groups regarding grammar instruction on the passive voice in English: a data-driven learning treatment based on guided induction principles (DDL), a deductive instructional treatment with corpus-informed teaching methods (DCI), and a traditional grammar instruction treatment (TGI). The study utilized a quantitative methodology and delayed posttests to observe the change in grammar ability over time. While the methodology was sufficient, it may have been more appropriate to utilize a mixed methods approach to obtain qualitative participant data, as the authors themselves admit the limitation of utilizing a “narrow view of what a grammar structure is” (Smart, 2014).

Results of the study via ANOVA tests showed that prior to instruction there were no significant differences between the three groups. The RMANOVA showed that the DDL group was the only group with statistically significant increased mean scores between pre-tests and posttests. Moreover, while the DCI group showed more gains than the TGI group, both groups did not have statistically significant gains. Finally, the delayed posttests two weeks later indicated that the DDL group maintained its gains over time.

Smart (2014) had the potential to contribute greatly to the present study due to its empirical evidence in support of a guided induction based strategic DDL approach to grammar instruction. However, the limitations of the study limited both its trustworthiness and contribution to the present review, as Smart (2014) admits that the passive voice as a grammatical construction may be best suited to learn using DDL. Therefore, further research must be made on different grammatical forms and

teaching methods in order for the conclusions Smart (2014) to be a major contributor in the present study.

4.3.2d Reynolds and Kao (2021) – Grammar acquisition through digital game-based instruction, teacher instruction, and direct focused written corrective feedback

Digital game-based instruction, teacher instruction, and direct focused written corrective feedback in relation to grammatical accuracy of English articles was explored by Reynolds and Kao (2021). The study was relevant to the present review as it had pedagogical implications dealing with strategy instruction. A quantitative methodology was used, and a pretest-posttest-delayed posttest design with written tasks was implemented. 45 Taiwanese university students who were English L2 learners participated in the study. They were randomly and equally divided into two experimental groups (digital game-based instruction using the game *English Extras In Business with A, An and The* with focused error correction and teacher instruction with focused error correction) or focused error correction.

The study explored how different instruction would affect the learning of English articles. Results showed that error correction worked best when the instruction was designed explicitly for certain language features, as displayed by both the digital game-based instruction and teacher instruction groups. The data gleaned from the delayed posttest a month after the immediate posttest indication that digital game-based instruction paired with focused written corrective feedback (provided by the teacher) may be most effective in the long-run. Moreover, the learning effect in the game-based instruction group was shown for generative use of the targeted grammatical structure. In other words, learners successfully learned grammar in one context (game) and applied it another context (writing job application letters).

Overall, Reynolds and Kao (2021) provided empirical evidence for the merits of utilizing digital games to supplement teacher instruction while teaching grammar. Both teacher instruction with focused error correction and/or digital game-based

instruction with focused error correction improved students' accuracy in using English articles. Questions remained over whether these effects remain the same for other grammatical structures, which explained why trustworthiness and contribution were rated a "medium", despite the high relevance of topic focus and appropriateness of study design.

4.3.2e Yeldham and Gruba (2013) – Effective instructional approaches for second language listening development

Yeldham and Gruba (2013) investigated effective instructional approaches for developing second language listening. Participants were all intermediate English L2 learners and consisted of six 18-year-old native Mandarin speakers. All students participated in a bottom-up skills course and data was collected through longitudinal multi-case studies over one and a half semesters.

For most listeners, the findings of the study showed that a bottom-up listening skills course was inadequate for developing listening ability. Perhaps more importantly, the study showed that an interactive approach between bottom-up and top-down listening processes was perhaps the optimal method. Regardless of their predispositions to a particular listening process as determined by the initial verbal report, participants in the study sought or achieved an interactive listening process style over time to compensate for the processing that they originally "lacked".

Implications and contributions of the study were highly relevant to this present review, as Yeldham and Gruba (2013) provided evidence toward instructing students with both top-down and bottom-up listening skills to develop an interactive listening approach. On the other hand, the appropriateness and trustworthiness of Yeldham and Gruba (2013) was deemed a "medium" rating because of the small sample size as well as the lack of limitations discussed in the study in light of issues such as the small sample size. Future research would benefit from investigating the empirical effectiveness of an interactive listening approach as suggested in Yeldham and Gruba (2013).

4.3.3 Weight of evidence assessment for low overall quality studies in relation to the present review

4.3.3a Munro et al. (2015) – Longitudinal study for pronunciation instruction and English segmentals

Munro et al. (2015) was a longitudinal study over a timeframe of two years which investigated adult L2 pronunciation instruction, specifically with regards to segmentals in English. The study had some degree of relevance to the present review due to its focus on pedagogical practices for pronunciation instruction. 17 Mandarin and 23 Slavic speakers in Canada participated in the study. They were all English L2 learners and were given speaking tests at different points over two years to observe how their speaking proficiency would develop.

Relevant to this review, the results of the study only indicated that individualized pronunciation training was needed as different L1 languages may result in different English L2 pronunciation problems. However, no further implications of this fact were expanded upon, which decreased the contribution of this study. While the study had potential to explore more about pedagogy or learning strategy, much of the study was dedicated to other factors such as exploring differences between Mandarin and Slavic-language groups or specific pronunciation features of English such as the /w/ and final /l/. Hence, the study was considered low with regards to trustworthiness and contribution to the present review.

A limitation of the study was the question about applicability of the study's findings in the real world since Munro et al. (2015) only considered single-word elicitations of 21 targets in an artificial setting during the speaking tests. Again, this contributed to its "low" rating for trustworthiness and contribution as more research must be made to consider other aspects of pronunciation, like "vowels and prosody, which are known to have an impact on intelligibility" (Munro et al., 2015).

4.3.3b Saito and Akiyama (2016) – L2 speech learning regarding video-based instruction

Another low overall quality study in relation to the present review was Saito and Akiyama (2016), which explored L2 speech learning regarding video-based instruction and negotiation for comprehensibility. The study's main relevance to the present review was simply that it offered speaking to native speakers as a "learning strategy", which could be considered of low relevance to the present review overall since no other strategy or pedagogy was mentioned. The study consisted of 20 Japanese undergraduate students who were learning English as an L2. Over one academic semester, the participants were divided into either an experimental group (with task-based conversation activities with native speaker partners over video), and a comparison group (weekly individual vocabulary/grammar activities without native speaker interaction).

Results indicated that overall, the experimental group "significantly improved in global, temporal, and grammatical qualities of their L2 speech" (Saito & Akiyama, 2016). At the same time, pronunciation and accented-ness did not seem to improve much in the experimental group. The study concluded that L2 interaction with native speakers would be particularly beneficial during the early stages of learning how to speak in an L2, but more research must be conducted over longer interaction periods, as well as other forms of pedagogy and/or language strategy to determine the most efficient learning approach. In short, while Saito and Akiyama (2016) had potential to dive deeper into L2 speaking pedagogy, it ultimately only scratched the surface of the subject area.

The study admitted to important limitations such as a very specific target population, lack of qualitative data, and issues of adequate length of speech samples. Taken together, Saito and Akiyama (2016) was an overall low quality study in relation to the present review due to its shallow implications and multitude of limitations which required further investigation.

4.4 Methodological quality of included studies

The methodological quality of studies in this review were evaluated using the 2018 version of the MMAT (Hong et al., 2018), a critical appraisal tool specifically designed for systematic mixed studies reviews with empirical research that utilizes different methodologies. In the present review, all included studies used either a quantitative non-randomized study methodology, a mixed methodology, or a qualitative methodology.

Overall, as shown in Tables 4.4.1, 4.4.2, and 4.4.3, the methodological quality of the included studies was high. Most studies satisfied the MMAT's requirements to qualify as containing high quality methodology within their respective categories of study designs. This would mean that that any study's value to this present review had more to do with its Weight of Evidence (relevance, appropriateness, trustworthiness, and contribution) than to its methodological quality.

4.4.1 Methodological quality of quantitative non-randomized studies

Most studies in this present review (n = 8) used a quantitative non-randomized methodology. This methodology is defined as "any quantitative studies estimating the effectiveness of an intervention or studying other exposures that do not use randomization to allocate units to comparison groups (Higgins & Green, 2008). Heift (2019) claimed to randomly select feedback types for different students, but provided no evidence to how they arrived at this "randomness", thereby failing the standards set out for randomness by the MMAT. Most studies included in this review chose to keep pre-existing classes intact without manually altering participant groups. While this may be argued as a form of randomization, according to the MMAT, it is insufficient because true randomization requires at minimum a simple allocation by following a predetermined plan/sequence, such as using a published list of random

numbers (Hong et al., 2018). None of those steps were taken in these studies, hence their inclusion under the category of quantitative non-randomized studies.

Overall, the methodological quality of included quantitative non-randomized studies in the present review were fairly high, as shown in Table 4.4.1. Five studies did not directly account for or discuss confounding variables throughout the study; Akkoç et al., (2018), Heift (2019), Munro et al. (2015), Rahimi et al. (2020), and Reynolds and Kao (2021). Moreover, Yeldham (2015) was interesting as it did not explicitly present review questions, but instead implied those questions through stating its hypotheses. While this does not mean that the studies should be abandoned entirely, it does provide evidence that future research must explicitly state its research questions and/or account for confounders in design or analysis, to ensure that the methodological quality of research in this area remains at a high standard.

Table 4.4.1 Methodological Quality of Included Quantitative Non-Randomized Studies (n = 8) Evaluated Using the MMAT

Study	Are there clear research questions?	Do the collected data allow to address the research questions?	Are the participants representative of the target population?	Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Are there complete outcome data?	Are the confounders accounted for in the design and analysis?	During the study period, is the intervention administered (or exposure occurred) as intended?
(Akkoç et al., 2018)	Yes	Yes	Yes	Yes	Yes	No	Yes
(Heift, 2019)	Yes	Yes	Yes	Yes	Yes	No	Yes
(Munro et al., 2015)	Yes	Yes	Yes	Yes	Yes	No	Yes
(Rahimi et al., 2020)	Yes	Yes	Yes	Yes	Yes	No	Yes
(Reynolds & Kao, 2021)	Yes	Yes	Yes	Yes	Yes	No	Yes

(Saito & Akiyama, 2016)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Smart, 2014)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Yeldham, 2015)	Can't tell	Yes	Yes	Yes	Yes	Yes	Yes

4.4.2. Methodological quality of mixed methods studies

Among the mixed methods studies (n = 5) in the present review, most of them were of high methodological quality according to the MMAT. Yeldham and Gruba (2013) was the only exception, as the study did not explicitly list its research questions. Instead, it was implied that Yeldham and Gruba (2013) wanted to explore how bottom-up skills instruction would affect L2 listening development over time. Besides this study, all of the other mixed methods studies included in the review had clear research questions and collected data which addressed the research questions. Adequate rationales for utilizing mixed methods were provided. The different studies effectively integrated various components to answer their respective research questions. Qualitative and quantitative components were adequately interpreted, and any divergences or inconsistencies were addressed. Finally, the different components in each mixed methods study adhered to the quality expected of the methods used.

Table 4.4.2 Methodological Quality of Included Mixed Methods Studies (n = 5) Evaluated Using the MMAT

Study	**Are there clear research questions?**	**Do the collected data allow to address the research questions?**	**Is there an adequate rationale for using a mixed methods design to address the research question?**	**Are the different components of the study effectively integrated to answer the research question?**	**Are the outputs of the integration of qualitative and quantitative components adequately interpreted?**	**Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?**	**Do the different components of the study adhere to the quality criteria of each tradition of the**

							methods involved?
(Brown, 2012)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Nguyen & Gu, 2013)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Sasaki et al., 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(Yeldham & Gruba, 2013)	Can't Tell	Yes	Yes	Yes	Yes	Yes	Yes
(Zhang, 2021)	Yes	Yes	Yes	Yes	Yes	Yes	Yes

4.4.3. Methodological quality of qualitative studies

Schneider (2021) was the only qualitative study included in the present review. As a study which utilized a longitudinal approach to observe the writing development of seven students over the course of four years, Schneider (2021) was methodologically sound and satisfied the requirements of the MMAT within the category of qualitative studies. Research questions were clearly stated, and the large amounts of data (54 interviews and 105 digital files of written work) sufficiently addressed the aforementioned questions. A qualitative data collection method was a good approach for addressing the research question, although the authors did hint that future research may benefit from incorporating and quantifying other sources of data such as students' written tests. This would mean that future research may potentially benefit from a more mixed methods approach. The findings of Schneider (2021) adequately derived from the data, interpretations of results were substantiated by the data, and the qualitative data sources, collection, analysis, and interpretation were all coherent and explained in a meaningful manner.

Table 4.4.3 Methodological Quality of Included Qualitative Studies (n = 1) Evaluated Using the MMAT

Study	**Are there clear research questions?**	**Do the collected data allow to address the research questions?**	**Is the qualitative approach appropriate to answer the research question?**	**Are the qualitative data collection methods adequate to address the research question?**	***Are the findings adequately derived from the data?***	**Is the interpretation of results sufficiently substantiated by data?**	**Is there coherence between qualitative data sources, collection, analysis and interpretation ?**
(Schneider, 2021)	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Chapter 5 Discussion

5.1 Summary of evidence

This chapter discusses the results of the included studies as they relate to the research questions of the present review. Specifically, focuses on learner and pedagogical implications and actionable steps will be discussed. Limitations of the findings are then explored, followed by recommendations for future research. Finally, conclusions based on the present review will be discussed.

5.2 First review question: What empirical research was carried out on adult L2 language learning strategies and strategy instruction between 2011 and 2022?

In general, based on the data and trends mentioned in section 4.2 concerning characteristics of included studies, empirical research in this area which includes delayed posttests and/or longitudinal focuses are trending in a positive direction with more research being conducted over time (Figure 4.2.1). While there was more

empirical research carried out on adult L2 language learning strategies and strategy instruction between 2011 and 2022 (as shown in Figure 4.1) the parameters of including only studies including delayed posttests and/or longitudinal tests was important in answering previous calls by Hassan et al. (2005) and Plonsky (2011) for more empirical research with delayed posttests and/or observing effects of language strategy and strategy instruction over time. Furthermore, as shown by the methodological quality of included studies in section 4.4 of the current review (Table 4.4.1, Table 4.4.2, Table 4.4.3), empirical research between 2011 and 2022 seem to have on the whole better methodological quality and includes better reporting in ways such as through descriptive statistics and statistical analysis, especially when compared with the research of previous years which lacked basic statistical reporting and often did not follow rigorous methodologies (Plonsky, 2011). The included studies between 2011 and 2022 showed that quantitative and mixed methods methodologies were most commonly used in research.

While most of the research (12 out of 14 studies) explored English as an L2, participants across the studies came from a variety of different L1 backgrounds. Grammar, speaking, and writing were the most common research focuses. Listening, learner autonomy, and orthography was also studied. No included study explicitly focused on reading. While most studies had some pragmatic pedagogical applications which could be applied by language learners or instructors, research in this area still appear to have a lot of room for further investigation. Specifically, stakeholders would have benefitted from research into how to pragmatically apply the conclusions of each study in clear actionable steps. Additionally, more research into different language contexts, participant populations, and longer time frames for delayed posttests (rather than just two to four/five weeks after intervention, as shown in the included studies) need to be investigated.

5.3 Second review question: What evidence is available for the effectiveness of language learning strategies and/or strategy

instruction for adult L2 language learners in terms of language proficiency?

The effectiveness of language learning strategies and/or strategy instruction for adult L2 language learners in terms of language proficiency was discussed in every included study in the present review.

Regarding grammar, Heift (2019) found empirical evidence supporting both meta-linguistic feedback, which identified learner mistakes and provided error explanations without revealing the correct answer and repetition-feedback, which prompted learners to correct mistakes without any error explanation. Meta-linguistic feedback was found to be more effective, but this conclusion still needs to be further investigated. Rahimi et al. (2020) also explored grammar pedagogy and concluded that utilizing both input flood tasks (focused tasks) and jigsaw tasks (unfocused tasks) together resulted in accurate and durable grammar-acquisition effects. One pedagogical approach that was explored was utilizing digital games with direct focused corrective feedback, which resulted in sustained grammar improvements according to Reynolds and Kao (2021). Finally, Smart (2014) found that grammar instruction resulted in positive learning outcomes particularly through using a data-driven learning treatment based on guided induction principles (DDL).

Writing was another language area focused on by a few of the included studies. Akkoç et al. (2018) found that raising awareness strategies through explicit teaching were effective in improving students' use of formulaic language. Crucially, continuity of explicit instruction was found to be important, and instruction should be supported by examples and techniques, not just explanation. In a longitudinal study over four years of writing strategy use through a self-regulation perspective, Sasaki et al. (2018) concluded that with more writing proficiency, global planning strategies were utilized more often. The study also indicated a positive effect of global planning instruction, being taught self-regulation techniques, and of study abroad experiences for a learner's writing development. Schneider (2021), another four-year longitudinal study, found that students all used growth, investment, and survival writing strategies at some stage over four years, in accordance with their goals and

identities. Implications of Schneider (2021) suggested that educational contexts are highly important for learning, and that was the instructors' and institutions' responsibility to create environments which encouraged the use of growth and investment strategies rather than survival strategies. Crucially to the present review, Schneider (2021) demonstrated that simply applying a particular strategy was not adequate for learner development. Instead, educational environments and incentives must be aligned with a learner's goals and identities.

A popular topic focus was in L2 speaking. Saito and Akiyama (2016) showed that task-based interaction could help inexperienced language learners become more successful communicators. However, an extensive amount of L2 interaction may be needed, and further research on the topic was required. Zhang's (2021) investigation on English L2 learners' development of complement responses using computer-mediated communication (CMC) and data-driven instruction showed that raising learners' pragmatic awareness could help their speaking development. The combination of CMC with data-driven instruction was shown to be particularly useful in situations where L2 learners may not have access to a naturalistic L2 learning environment. A longitudinal study on segmental priorities for English L2 learners was conducted by Munro et al. (2015). The study concluded that L2 learners would benefit from individualized pronunciation instruction, as learners from different L1 backgrounds may not equally benefit from particular "catch-all" pronunciation training. Further research into learners from other L1 backgrounds may be needed, Munro et al. (2015) only explored participants from Mandarin language or Slavic language L1 backgrounds.

Studies like Yeldham (2015) and Yeldham and Gruba (2013) explored L2 listening. Yeldham (2015) compared a strategies-based approach with an interactive, strategies/bottom-up approach for language L2 listening instruction. The study found that the strategies group outperformed the interactive group in listening comprehension. The strategy group also outperformed the interactive group with regards to strategic ability and learner characteristics. Finally, the interactive group outperformed the strategies group in developing bottom-up skills. The key

implications of Yeldham (2015) were in recommending more strategy instruction rather than implementing an interactive approach for lower-intermediate level L2 listeners. In a similar vein, Yeldham and Gruba (2013) also investigated pedagogical approaches to teaching L2 listening. In this study, it was found that teaching both top-down and bottom-up listening processes to develop an interactive listening approach was most beneficial for L2 learner. There was a need for future research in this area, as the results of Yeldham (2015) appeared to contradict the implications of Yeldham and Gruba (2013).

Brown (2012) was the only study which focused on orthography. The study explored pedagogical approaches for teaching Korean Hangul and concluded that visual/verbal and physical mnemonics were noteworthy techniques that resulted in significant learning advantages over time when compared with traditional teaching methods. Brown (2012) found that mnemonics were not only quantitatively effective, but based on qualitative data, was also enjoyed by students.

Finally, Nguyen and Gu (2013) was the only study focused on learner autonomy. The study concluded that strategy-based instruction was an important and worthwhile pedagogical approach as it improved self-regulation skills while also resulted in significant and lasting gains in writing scores.

5.4 Third review question: What evidence is available for how language learning strategy use has changed over time for adult L2 language learners?

A few of the included studies explored how language learning strategy use changed over time for adult L2 language learners. In Brown's (2012) study on Korean Hangul, language learning strategy use of mnemonics was found to be utilized long-term by the students, even after intervention. After four or five weeks, there was qualitative evidence in Brown (2012) indicating that students continued to utilize mnemonics outside of class. Additionally, Brown (2012) also found that some students had adapted the principles behind class-instructed mnemonics to create their own

mnemonics to aid their orthographic learning process. In other words, utilizing mnemonics was one language learning strategy that was retained long-term for students, and was also beneficial to students in inspiring alternative self-created mnemonics for other Korean characters not covered during the intervention.

Sasaki et al. (2018) found that with regards to L2 writing strategy, four years of observations revealed that learners naturally leaned toward more global planning strategies as their writing proficiency improved. Additionally, the positive effect of study abroad experiences and career-related experiences seemed to positively impact the development of self-regulated learning over time. To aid in the development of language strategy usage, Sasaki et al. (2018) suggested that teaching should include both language learning strategies and self-regulation options. Overall, Sasaki et al. (2018) provided evidence for Oxford's (2017) conception of language learning strategy as a "complex and dynamic phenomenon based on interaction between learners with agency and their environments".

In another longitudinal four-year study, Schneider (2021) demonstrated that over time, learners would adopt specific writing strategies (growth, investment, or survival strategies) based on their individualized goals and identities. Schneider (2021) illustrated the power and responsibility that instructors and institutions have in creating a learning environment conducive to learner goals and identities. Strategy usage was determined by the learning environment and its relation to learner motives. Schneider's (2021) important contribution to the present study was that over time, a learner may prioritize alternative strategy choices to suit their needs. In other words, the usage and application of language learning strategies was dynamic, and instructors and institutions hold the power in encouraging these strategic applications.

Regarding listening strategies over time, Yeldham and Gruba (2013) found that it was inadequate to only instruct students using bottom-up skills. Naturally over time, L2 listeners tended to gravitate and acquire the skills (be it bottom-up or top-down) that they "lacked". In other words, regardless of a learners' initial propensity toward employing bottom-up or top-down listening skills, Yeldham and Gruba (2013) found

that instructing and encouraging both bottom-up and top-down listening processes would be most beneficial to L2 learners.

Chapter 6 Conclusion

6.1 Limitations

Limitations of the included studies have been discussed previously in chapters 4 and 5 of the present review. In summary, a key limitation of many included studies was the absence of considering potential confounding variables and/or discussing limitations in the study's conclusion(s) (such as in the five studies shown in Table 4.4.1). By not addressing these confounding variables, research conclusions may be thrown into doubt as there may be other factors responsible for the results found in the studies that were unaccounted for. Another limitation was whether or not the conclusions of studies focusing on specific language aspects could be extrapolated to other language areas. An example may be the four studies investigating grammar (Heift, 2019; Rahimi et al., 2020; Reynolds & Kao, 2021; Smart, 2014) which all mentioned how future research needed to explore other grammar aspects to reinforce their findings. Publication bias (Song et al., 2010) may be another limitation of the present review. Specifically, as only peer-reviewed journal articles were considered in the review, none of the considered studies confirmed a null hypothesis. Potential under-reporting may have affected the landscape of available research. Another limitation mentioned in some of the included studies was the absence of a control group, which would have better framed the results found in intervention groups. Finally, a major limitation of the findings is that most of the delayed posttests utilized in the studies occurred two to four/five weeks after intervention or immediate posttest. From an instructor's or language learner's perspective, more research on language strategy effects over a longer period of time (one year or more) may be beneficial. In other words, it may be too quick to claim "long-term" effects from any of the included studies, as the delayed posttests may not have been far enough removed from the related interventions.

6.2 Recommendations for future research

While it is heartening that more empirical research has followed the calls of Hassan et al. (2005) and Plonsky (2011) to adhere to higher methodological and quality standards, there remains plenty of room for future research. At present, the increasing usage of longitudinal approaches, delayed posttests, and mixed methods methodology appears to be moving this research area into a positive direction.

Many of the quantitative studies in the included review called for future research with qualitative components to better understand participant processes and perspectives from an emic standpoint. Delayed posttests may also be conducted at later intervals (a few months or a year or longer) to better substantiate claims of any “long-term” benefits of language learning strategy usage and/or strategy instruction.

From a pragmatic standpoint, more research into actual application of the recommended language strategies and/or pedagogical approaches (especially within a natural learning environment) could provide more actionable implications for both language learners and instructors. Finally, as mentioned in a few of the included studies, exploring the application of language learning strategies and strategic instruction utilizing other language aspects (different grammar points, different L1 groups, different L2 languages etc..) may prove fruitful in strengthening the reliability and applicability of any strategy and/or pedagogical approach.

6.3 Final thoughts

The present review has built on previous literature by exploring new approaches and empirical research in exploring the usage, effectiveness, and changes of language learning strategies and strategy instruction for adult L2 learners. At present, there seems to be ample evidence suggesting the pragmatic and pedagogical applicability and benefits of language learning strategy usage and strategy instruction within an

adult L2 language learning context, but questions remain as to which strategies to prioritize and how to apply them within learning and instructional settings.

References

Akkoç, A. B., Qin, J., & Karabacak, E. (2018). The effects of explicit instruction of formulaic language on EFL argumentative writing quality. *Indonesian Journal of Applied Linguistics*, 8(2). https://doi.org/10.17509/ijal.v8i2.13282

Ardasheva, Y., & Tretter, T. R. (2013). Strategy Inventory for Language Learning-ELL Student Form: Testing for Factorial Validity. *The Modern Language Journal*, 97(2), 474–489. https://doi.org/10.1111/j.1540-4781.2013.12011.x

Benson, P. (2007). State-of-the-art article. *Language Teaching*, 40, 21–40.

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

Brown, L. (2012). The use of visual/verbal and physical mnemonics in the teaching of Korean Hangul in an authentic L2 classroom context. *Writing Systems Research*, 4(1), 72–90. https://doi.org/10.1080/17586801.2011.635949

Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2012). Finding Relevant Studies. In David; Gough, S. Oliver, & J. Thomas (Eds.), *An Introduction to Systematic Reviews* (2nd ed., pp. 93–122). London: Sage Publications.

Chang, B-M. (2011). The roles of English language education in Asian context. *Journal of Pan-Pacific Association of Applied Linguistics*, 15(1), 191-206

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Thousand Oaks, CA: SAGE.

Cochrane Effective Practice and Organisation of Care (EPOC). (2017). Data collection form. Retrieved from https://epoc.cochrane.org/resources/epoc-resources-review-authors

Čolović-Marković, J. (2012). The effects of explicit instruction of formulaic sequences on second-language writers (Unpublished Doctoral dissertation). The University of Utah, the USA.

Cresswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: SAGE.

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. Mahwah, NJ: Erlbaum.

Dweck, C. S. (2022). MINDSET: NEW PSYCHOLOGY OF SUCCESS. Ballantine Books.

Google Scholar. (n.d.). Google Scholar. https://scholar.google.ca/scholar?hl=en&as_sdt=0%2C5&q=macaro+2006&btnG=

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-systematic-reviews/book245742

Hassan, X., Macaro, E., Mason, D., Nye, G., Smith, P., & Vanderplank, P. (2005). Strategy training in language learning: a systematic review of available research. In: Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. Retrieved from: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=297

Heift, T. (2019). A longitudinal observation of technology-mediated feedback for L2 learners of German. *Technology-Mediated Feedback and Instruction*, 170(2), 154–179. https://doi.org/10.1075/itl.19009.hei

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, UK: Wiley Online Library.

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

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

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

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

Macaro, E. (2019). Systematic reviews in applied linguistics. In McKinley, J., & Rose, H. (Eds.). *The Routledge Handbook of Research Methods in Applied Linguistics* (Routledge Handbooks in Applied Linguistics) (1st ed.). Routledge.

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

Manchón, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes: Implications for theory. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 102–129). Bristol, UK: Multilingual Matters.

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

Migration Data in Northern America. (2021, October 7). Migration Data Portal. https://www.migrationdataportal.org/regional-data-overview/migration-data-northernamerica#:%7E:text=As%20of%202020%2C%20a%20total,in%20Canada)%20(Ibid).

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

Moher, D., Pham, B., Lawson, M. L., & Klassen, T. P. (2003). The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment*, 7(41). https://doi.org/10.3310/hta7410

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). https://doi.org/10.1186/2046-4053-4-1

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teachin*g, 53(1). https://doi.org/10.1515/iral- 2015-0002

Nguyen, L. T. C., & Gu, Y. (2013). Strategy-based instruction: A learner-focused approach to developing learner autonomy. *Language Teaching Research*, 17(1), 9–30. https://doi.org/10.1177/1362168812457528

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Oxford, R.L. (1990). Language learning strategies: What every teacher should know. Boston: Heinle & Heinle. Google Scholar

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews*. BMJ, n71. https://doi.org/10.1136/bmj.n71

Panic, N., Leoncini, E., Belvis, G. De, Ricciardi, W., & Boccia, S. (2013). Evaluation of the Endorsement of the Preferred Reporting Items for Systematic Reviews and Meta- Analysis (PRISMA) *Statement on the Quality of Published Systematic Review and Meta-Analyses*. PLoS One, 8(12), 1–7. https://doi.org/10.1371/journal.pone.0083138

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

Rahimi, S., Ahmadian, M., Amerian, M., & Dowlatabadi, H. R. (2020). Comparing Accuracy and Durability Effects of Jigsaw Versus Input Flood Tasks on the Recognition of Regular Past Tense /-ed/. *SAGE Open*, 10(2), 215824402091950. https://doi.org/10.1177/2158244020919505

Ranalli, J. (2012). Alternative models of self-regulation and implications for L2 strategy research. *Studies in Self-Access Learning Journal*, 3(4), 357e376.

Reynolds, B. L., & Kao, C. W. (2019). The effects of digital game-based instruction, teacher instruction, and direct focused written corrective feedback on the grammatical accuracy of English articles. *Computer Assisted Language Learning*, 1–21. https://doi.org/10.1080/09588221.2019.1617747

Rose, H. (2011). Reconceptualizing Strategic Learning in the Face of Self-Regulation: Throwing Language Learning Strategies out with the Bathwater. *Applied Linguistics*, 33(1), 92–98. https://doi.org/10.1093/applin/amr045

Rose, H. (2015). Researching Language Learner Strategies. In B. Paltridge & A. Phakiti (Eds.), *Research Methods in Applied Linguistics: A Practical Resource* (pp. 421–438). New York, NY: Bloomsbury.

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

Rose, H., McKinley, J., & Galloway, N. (2020). Global Englishes and language teaching: A review of pedagogical research. *Language Teaching*, 54(2), 157–189. https://doi.org/10.1017/s0261444820000518

Rubin, J. (1975). What the “Good Language Learner” Can Teach Us. *TESOL Quarterly,* 9(1), 41. https://doi.org/10.2307/3586011

Saito, K., & Akiyama, Y. (2016). Video-Based Interaction, Negotiation for Comprehensibility, and Second Language Speech Learning: A Longitudinal Study. *Language Learning*, 67(1), 43–74. https://doi.org/10.1111/lang.12184

Sasaki, M., Mizumoto, A., & Murakami, A. (2018). Developmental Trajectories in L2 Writing Strategy Use: A Self-Regulation Perspective. *The Modern Language Journal*, 102(2), 292–309. https://doi.org/10.1111/modl.12469

Schneider, J. (2021). Writing Strategies as Acts of Identity. *TESOL Quarterly*, 56(1), 230–253. https://doi.org/10.1002/tesq.3061

Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 119–144). Mahwah, NJ: Lawrence Erlbaum.

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. ReCALL, 26(2), 184–201. https://doi.org/10.1017/s0958344014000081

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.; Parekh, S.; Hooper, L.; Loke, Y. K.; Ryder, J.; Sutton, A. J.; Hing, C.; Kwok, C. S.; Pang, C.; Harvey, I. (2010). "Dissemination and publication of research findings: An updated review of related biases". *Health Technology Assessment*. 14 (8): iii, iix–xi, iix–193. doi:10.3310/hta14080. PMID 20181324.

Taguchi, N. (2015*). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going*. Language Teaching, 48, 1–50.

Teng, L. S., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal*, 100(3), 674e701.

Thomas, J., O’Mara-Eves, A., Kneale, D., & Shemilt, I. (2012). Synthesis Methods for Textual or Mixed Methods Data. In David Gough, S. Oliver, & J. Thomas (Eds.), *An Introduction to Systematic Reviews* (2nd Edition, pp. 181–210). London.

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, 78–102.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening*. New York, NY: Routledge.

White, C, Schramm, K & Chamot, A (2007), Research methods in strategy research, in AD Cohen & E Macaro (eds)*, Language learner strategies: thirty years of research and practice*, Oxford University Press, Oxford, pp. 93–116.

Williams, M., Mercer, S., & Ryan, S. (2016). Exploring Psychology in Language Learning and Teaching (Oxford Handbooks for Language Teachers) (Illustrated ed.). Oxford University Press.

Yeldham, M. (2015). Second Language Listening Instruction: Comparing a Strategies-Based Approach With an Interactive, Strategies/Bottom-Up Skills Approach. *TESOL Quarterly*, 50(2), 394–420. https://doi.org/10.1002/tesq.233

Yeldham, M., & Gruba, P. (2013). Toward an instructional approach to developing interactive second language listening. *Language Teaching Research,* 18(1), 33–53. https://doi.org/10.1177/1362168813505395

Zhang, L. J., Thomas, N., & Qin, T. L. (2019). Language learning strategy research in System: Looking back and looking forward. *System*, 84, 87–92. https://doi.org/10.1016/j.system.2019.06.002

Zhang, Y. (2021). Combining computer-mediated communication with data-driven instruction: EFL learners' pragmatic development of compliment responses. *System*, 103, 102624. https://doi.org/10.1016/j.system.2021.102624

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a Self-Regulated Writer: A Social Cognitive Perspective. *Contemporary Educational Psychology*, 22(1), 73–101. https://doi.org/10.1006/ceps.1997.0919

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). Handbook of self-regulation of learning and performance. New York: Routledge.

Appendices

Appendix A: Review protocol

TITLE

A systematic review of research on adult L2 language learning strategies and strategy instruction

INTRODUCTION

Rationale

Language learning strategies and strategy instruction has been researched in the field of second language acquisition for over four decades, to varying degrees of success. With unprecedented growth and research interest over the last decade (Gavriilidou & Mitits, 2021), the crucial role of language learning strategies in language learning has been established in diverse fields such as applied linguistics, psycholinguistics, and cognitive psychology, among others.

From the beginning in the 1970s focusing on “good” language learners (Rubin, 1975) to L2 learning strategies (Oxford, 1990), to the theoretical discussions and criticisms of Dörnyei (2005) and Macaro (2006), and to now incorporating other concepts in education like self-regulation into language learning strategy research (Rose et al., 2018), research on language learning strategies has had a variety of focuses and approaches.

As systematic reviews such as Hassan (2005) and Plonsky (2011) have pointed out, language learning strategies and strategy instruction have a potentially key role to play in the process of L2 acquisition. However, there were key limitations in both studies such as a lack of research in specific language areas like speaking, writing, and vocabulary (Hassan, 2005), lack of delayed post

tests in exploring strategy instruction (Plonsky, 2011), and challenges in choosing appropriate instruments and methodology (Hassan, 2005; Plonsky, 2011).

While Rose et al. (2018) focused more on best research practices in the field of language learning strategy, the present study seeks primarily to build upon the work of Hassan (2005) and Plonsky (2011), focusing on the role that language learning strategy and strategy instruction plays on adult L2 language learners.

Objectives

This study aims to identify and evaluate primary research on adult L2 language learning strategy and strategy instruction from 2011 to 2022 to gather, present, and explore the effectiveness of language learning strategies and strategy instruction for adult L2 language learners. The present study attempts to answer the following questions:

1. What empirical research was carried out on adult L2 language learning strategies and strategy instruction between 2011 and 2022?
2. What evidence is available for the effectiveness of language learning strategies in adult L2 language learners in terms of language proficiency?
3. What evidence is available for how language learning strategy use has changed over time for adult L2 language learners?

METHODS

Eligibility criteria for the review

Eligibility criteria (inclusion and exclusion) was based on systematic reviews such as Rose et al. (2020) which explored a similar area in language pedagogy.

Inclusion criteria:

- Must contain primary empirical research.
- Must be articles been published between 2011 and 2021.
- Must be about language learning strategies (LLS) or strategy instruction (SI), specifically within L2 acquisition.
- Must be topically related to one of the research questions mentioned above in the objectives section (language learning strategy usage, effectiveness of learning strategies, strategy usage over time, and effective strategy instruction).

Exclusion criteria:

- Focuses on topics unrelated to language learning strategies, strategy instruction, or L2 acquisition.
- Population does not include adult L2 learners.
- Theoretical articles or reports on practice without research methodology.
- Research on topics unrelated to learning language strategy or strategy instruction.
- Although research in multiple languages is a form of high-quality systematic reviewing (Moher et al., 2003), due to time constraints, only sources written in English will be considered.

Information sources

Following previous systematic reviews in a similar research area such as Plonsky (2011) as well as the recommendations by my supervisor Dr. Heath Rose at the University of Oxford, the current review will utilize the following relevant databases:

- ProQuest

- Scopus
- EBSCO
- Web of Science (Core Collection)
- Linguistics and Language Behavior Abstracts (LLBA)
- Educational information Research Center (ERIC)
- Google Scholar

An initial search was conducted on November 11, 2021.

Search strategy

While the initial search found around 3,383 articles, subsequent searches focused on specific topic-relevant keywords and utilized Boolean terms to focus the search according to the eligibility criteria of the study. Thematically, searches focused on themes such as “language learning strategy”, “strategy instruction”, “adult”, “second language”.

Search Terms								
Skill		Strategy		Age		Content Learning		Second Language
title and abstract	AND	title and abstract	AND	title and abstract	AND	title and abstract	AND	anywhere in the document
language-learning		strateg*		adult		listening		additional-language
language		instruction		young-adult		reading		EAL
language-learn		SI		university		writing		EFL
language-skill*		LLS		college		speaking		ESL
language-learning-strategy				adult*		vocabulary		foreign-language
						grammar		L2
						pronunciation		second-language
						spoken-communication		EAP
						proficiency		ESP
						post-test		ESOL
						post-intervention		bilingual
						self-regulation		multilingual
						empirical		
						vocab		
						vocab*		
						intervention		
						delayed-post-test		
						longitudinal		

While different databases utilize Boolean terms differently in their search queries, below is a query string from Scopus for a search conducted on February 9, 2022, which resulted in 465 articles:

TITLE-ABS (language-learning OR language OR language-learn OR language-skill* OR language-learning-strategy OR strategy-based-instruction OR sbi) AND TITLE-ABS (strateg* OR instruction OR si OR lls) AND TITLE-ABS (adult OR young-adult OR university OR college OR adult*) AND TITLE-ABS (intervention OR delayed-

post-test OR longitudinal OR listening OR reading OR writing OR speaking OR vocabulary OR vocab OR vocab* OR grammar OR pronunciation OR spoken-communication OR proficiency OR post-test OR post-intervention OR self-regulation OR empirical) AND ALL (additional-language OR eal OR efl OR esl OR foreign-language OR l2 OR second-language OR eap OR esp OR esol OR bilingual OR multilingual OR strategy AND based AND instruction OR sbi) AND NOT ALL (child*) AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011)) AND (LIMIT-TO (LANGUAGE , "English") OR EXCLUDE (LANGUAGE , "Spanish") OR EXCLUDE (LANGUAGE , "Turkish") OR EXCLUDE (LANGUAGE , "Chinese") OR EXCLUDE (LANGUAGE , "Japanese") OR EXCLUDE (LANGUAGE , "Malay")) AND (EXCLUDE (EXACTKEYWORD , "Child*")) AND (EXCLUDE (SUBJAREA , "MEDI")) AND (EXCLUDE (SUBJAREA , "HEAL")) AND (LIMIT-TO (SUBJAREA , "ARTS"))

Selection process

The current study will use Rayyan (Ouzzani et al., 2016) to manage search results. Identified studies will be assessed using the eligibility criteria for an initial/first screening. This first screening will look at the titles and abstracts of search results and a collaborator will screen 10% to increase reliability. This first screening will take place on June 18, 2022. A second screening consisting of a full-text screening will be performed with a collaborator on June 25, 2022. In both the first and second screening, if studies do not meet all criteria, they will be excluded from the study.

Data collection process

Data will be collected using a data extraction form modelled after the guidelines presented in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Thomas, 2022). Adjustments will be made to suit the purpose of this current review.

Data items

Data items extracted from each study based on the criteria in Higgins and Thomas (2022) will include:

- Information about data extraction from reports

- Eligibility criteria
- Study methods
- Participants
- Intervention
- Outcomes
- Results
- Miscellaneous

Study risk of bias assessment

Each study will be assessed for quality and trustworthiness based on the guidelines set out by Gough's (2007) Weight of Evidence Framework in conjunction with the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018) which will be used as a critical appraisal tool for the study.

Data synthesis

A narrative (qualitative) synthesis will be adopted, based on the recommendations and guidelines of Gough et al. (2019). This corresponds with the nature of the open research questions and allows the synthesis of various heterogeneous studies.

REFERENCES

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

Gavriilidou, Z., & Mitits, L. (2021). Situating Language Learning Strategy Use: Present Issues and Future Trends (Second Language Acquisition, 146) (Volume 146). Multilingual Matters.

Gough, D., Thomas, J., & Oliver, S. (2019). Clarifying differences between reviews within evidence ecosystems. *Systematic Reviews*, 8(1). https://doi.org/10.1186/s13643-019-1089-2 (Hassan, 2005)

Higgins, J., & Thomas, J. (2022). Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Training. https://training.cochrane.org/handbook/current

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

Moher, D., Pham, B., Lawson, M. L., & Klassen, T. P. (2003). The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment*, 7(41). https://doi.org/10.3310/hta7410

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

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021b). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, n71. https://doi.org/10.1136/bmj.n71

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

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

Rose, H., McKinley, J., & Galloway, N. (2020). Global Englishes and language teaching: A review of pedagogical research. *Language Teaching*, 54(2), 157–189. https://doi.org/10.1017/s0261444820000518

Rubin, J. (1975). What the “Good Language Learner” can teach us. *TESOL Quarterly*, 9(1), 41– 51. https://doi.org/10.2307/3586011

Appendix B: Prisma 2020 checklist (Page et al., 2021b)

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review.	1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	N/A
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	2-3
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	13-16
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	17
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	17-19
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	19-20
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	20-23
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	22-23
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	22-23
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	25-26
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	25-26
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	25-26
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	25-26
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	25-26
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	25-26
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	25-26
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	25-26
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	25-26
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	25-26

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Reported on page #
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	27
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	27
Study characteristics	17	Cite each included study and present its characteristics.	28-41
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	27-59
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	27-59
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	27-59
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	27-59
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	27-59
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	27-59
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	27-59
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	27-59
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	59-67
	23b	Discuss any limitations of the evidence included in the review.	59-67
	23c	Discuss any limitations of the review processes used.	59-67
	23d	Discuss implications of the results for practice, policy, and future research.	59-67
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Appendix C: Sample data extraction table (Rahimi et al., 2020)

Extraction Grid for In-Depth Review of Language Learning Strategies and Strategy Instruction

Reviewer 1: 1049104
Reviewed Item: Comparing Accuracy and Durability Effects of Jigsaw Versus Input Flood Tasks on the Recognition of Regular Past Tense /-ed/

IN-DEPTH REVIEW ITEM	CRITERIA FOR QUALITY	REVIEWER RESPONSE: yes/no/partly/unclear clearly stated/implied and/or brief comment/insertion/n.a. (not applicable)
STUDY INFORMATION	Date	June 2020
	Study ID / DOI	https://doi.org/10.1177/2158244020919505
	Author(s)	Soheil Rahimi, Moussa Ahmadian, Majid Amerian, Hamid Reza Dowlatabadi
	Title	Comparing Accuracy and Durability Effects of Jigsaw Versus Input Flood Tasks on the Recognition of Regular Past Tense /-ed/
	Year	2020
	Publication Type	Journal Article
	In-text Reference	(Rahimi et al., 2020)
	Full Reference	Rahimi, S., Ahmadian, M., Amerian, M., & Dowlatabadi, H. R. (2020). Comparing Accuracy and Durability Effects of Jigsaw Versus Input Flood Tasks on the Recognition of Regular Past Tense /-ed/. SAGE Open, 10(2), 215824402091950. https://doi.org/10.1177/2158244020919505
THE ABSTRACT	Does the abstract provide sufficient information to understand the study?	Yes
INTRODUCTION/ RATIONALE: DOES THE STUDY SAY.	When was the study carried out?	2020
	Why was it carried out at this point in time?	Attempt to gain insight on whether input flood tasks or Jigsaw tasks affect learners' recognition of L2 forms. Both input flood tasks and Jigsaw tasks have prior literature, but there is a dearth of studies investigating both types of tasks together.
	Why was it carried out with this particular group of people?	19–22-year-old Persian-speaking university students, homogenous regarding language and similar grammar proficiency levels.
	In which country was the study carried out?	Iran
	If the study was funded, and by whom?	No funding
THE LITERATURE REVIEW	Does the study state an actual recognisable theory/group of theories or constructs to which it is related? Which?	Yes. The study is related to literature focused on input flood tasks (focused tasks) as described by Nassaji and Fotos (2011), as well as Jigsaw tasks (unfocused tasks), as explored in studies like Pica (2005).

	How much empirical evidence is presented as part of the review?	Mostly narrative and qualitative evidence
	Is it mainly primary evidence, or mainly secondary evidence?	Secondary
	Does it end with a summary?	Yes
	Does the summary clearly invite the research questions that follow?	Yes
LANGUAGE LEARNING STRATEGY (LLS) / STRATEGY INSTRUCTION (SI) DEFINITION	Is there a definition? Is it compared with other definitions of LLS or SI?	No
THE TOPIC(S) OF THE STUDY	e.g.: pedagogy for teaching language learning strategies	Use of input flood tasks and Jigsaw tasks in promoting recognition of regular past tense /-ed/ in terms of accuracy and durability.
RESEARCH QUESTIONS	**RQs are clearly stated or implied?** Yes. 1. Is there any significant difference between the effects of instruction through input flood tasks versus instruction through Jigsaw tasks on Iranian EFL learners' accurate recognition of regular past tense /-ed/? 2. Is there any significant difference between the effects of instruction through input flood tasks versus instruction through Jigsaw tasks on Iranian EFL learners' durable recognition of regular past tense /-ed/?	
METHOD	What, broadly, is the methodology adopted? Quant/Qual/mixed? Purely descriptive; exploration of relationship among variables; experimental?[i]	Quasi-experimental design including a pretest, posttest, a delayed posttest, one experimental group, and one comparison group. Quantitative methodology used.
	prospective/longitudinal?	Longitudinal study
VARIABLES	Are there variables involved? (Where appropriate) what are the dependent and independent variable(s)?	Independent variables involved instruction via Jigsaw tasks or input flood tasks. Dependent variables were the recognition of the instructed language feature.
	What other variables are 'controlled for'? (Confounding variables?)	Participant language and grammar proficiency was homogenized through the Oxford Placement Test (OPT).
SAMPLING	Population; Nationality; L1;	Two Iranian university classes (62 students in total) ranging from age 19-22. L1 is Persian.
	sampling frame provided?	No
	Sampling procedure explained?	Yes

	What was the actual sample?	62 participants randomly assigned to experimental and comparison groups.
GROUPING	Are there recognisable groups in the sample?	Yes (intact classes)
	Was the sample divided into groups or did the groups already exist?	Groups existed (intact classes) but were assigned to different interventions randomly (coin toss).
	(Where relevant) did the groups know they were being divided up like this and for this purpose?	No
	What treatment <u>if any</u> did each group get?	The experimental group was instructed through input flood tasks. The comparison group was instructed through Jigsaw tasks.
	Were the groups aware of the treatment	No
ETHICS	Was participant consent sought?	Unclear

In-depth review item	criteria for quality	reviewer response: yes/no/partly or brief comment/insertion
DATA COLLECTION	Were the research instruments trialled or validated in some way?	Yes
	How/who was the data collected (does this seem a reliable way of collecting the data?)	Yes, data was collected via The Oxford Placement Test (OPT) and the Untimed Grammatically Judgment Test UGJT). Both of which seem to be validated in literature.
	Were there sufficient amounts of data collected?	Yes
DATA ANALYSIS	Are we told how the data was analysed?	Yes
	Does this seem like a valid way of analysing the data? (**how**?)	Yes, by utilizing multiple statistical tests such as a MANOVA, MANCOVA, ANCOVA, Levene's test, to limit potential influence of confounding variables.
	Does this seem like a reliable way of analysing the data? (**who**?)	Yes, by first checking for homogeneity between participants, and then running the same type(s) of analysis for both experimental and comparison groups.
	Does the analysis match the requirements of the research questions? (**sufficient**?)	Yes
RESULTS; FINDINGS; CONCLUSIONS ETC	**What are the actual results?** In answering the first research question, results indicated that the learners in the experimental group (input flood tasks) did better than the comparison group (Jigsaw tasks), though both groups improved in the immediate posttest. To answer the second research question, the comparison group (Jigsaw tasks) showed better long-term recognition than the experimental group (input flood tasks).	
	Are there any shortcomings in the reporting of the results?	n.a.
	Do their conclusions match your assessment of the findings/results?	Yes
	Are limitations of the study discussed? (e.g. confounding variables)	A more realistic/appealing result may have occurred if learners' production performance was also measured, such as through a writing test. Replicating the study with participants from different majors may also strengthen the study's findings.
	Are there implications? For teaching and learning?	The main implication for teaching and learning is that both input flood tasks and Jigsaw tasks (both focused and unfocused tasks) are beneficial when

		utilized together for the goal of accurate and durable effects when learning grammar.
	Do the implications match the study findings?	Yes
	Are there suggestions for further research	As seen in this study, some of the effects of some tasks (such as Jigsaw tasks) may not be observed directly in an immediate posttest, so delayed posttests should be used more in research on task-based instruction (TBI). Future studies may also benefit from more realistic learner production (a written test) and learners from different academic backgrounds (different majors).

INITIAL WEIGTHING BY REVIEWER

Weight of evidence		**High /Medium/Low**
WOE: **Relevance…**	of particular **focus of the study** for addressing the question or sub-questions of this specific systematic review. Please add the number of the Review Question(s)	High (Review Question 1, 2)
WOE: **Appropriateness…**	of research **design and analysis** for addressing the question, or sub-questions, of this specific systematic review.	High
WOE: **Trustworthiness**…	Taking account of all quality assessment issues, can the study findings be trusted in answering *the study* question(s)?	High
WOE: **Contribution..**	of the study to answer the question/s of this specific systematic review	High

Additional References:

Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative contexts*. Routledge.

Pica, T. (2005). *Classroom learning, teaching, and research: A task-based perspective. The Modern Language Journal*, 89(3), 339–352.