L1-mediated strategies in primary and secondary English grammar classrooms: A systematic review

Chi LAM

MSc in Applied Linguistics and Second Language Acquisition, 2025

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION



1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.

2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].

3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.

4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.

5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Chi LAM
Date:	1 October, 2025

L1-mediated strategies in primary and secondary English grammar classrooms: A systematic review

Chi Lam

Department of Education, University of Oxford
Wolfson College
Supervisor: Dr. Faidra Faitaki



Dissertation submitted in partial fulfilment of the requirements for the degree of
MSc in Applied Linguistics and Second Language Acquisition

Abstract

Following García (2009)'s reassertion of the pedagogical value of translanguaging, research has increasingly recognised the role of bi/multilingual learners' first language (L1) in supporting language development. Grammar instruction is no exception: many teachers across contexts report incorporating learners' L1 to explain complex syntactic structures despite believing immersion is essential for language acquisition. This belief-practice gap highlights a key pedagogical tension under Communicative Language Teaching (CLT).

Focusing on English learning as a second language (ESL) and foreign language (EFL), this systematic review examines how learners' L1 has been integrated into explicit grammar instruction at primary and secondary levels. Specifically, it identifies the types of L1-mediated strategies described in empirical research and collates their reported effects using a narrative approach.

A total of seven studies met the eligibility criteria. Despite the small sample size and exclusive EFL focus, a range of L1-based approaches was identified, namely translanguaging, pedagogical translanguaging, codeswitching, and 'use of L1'. Although the language gains associated with these strategies were not always significant, there was no evidence of L1 use impeded grammar development.

A key finding of the synthesis concerns inconsistencies surrounding the strategies featured, reflecting the jingle-jangle fallacy: identical labels were applied to different practices (jingle), while similar practices were given different labels (jangle). The incongruity raises questions about comparability and generalisability of findings. A second inconsistency relates to study quality: Variability in reporting and measuring classroom own-language use makes it difficult to draw confident conclusions about the effects of these approaches. This review calls for clearer conceptualisation and more rigorous and systematic measurement of L1-facilitated instructional practices across different strategies.

Acknowledgements

This dissertation records the intellectual journey I undertook over the past year. I did not see it as a mere assignment, but a testament of my personal and academic growth. I could not have come this far without the support of the following people:

First and foremost, credit must go to my supervisor, Dr. Faidra Faitaki, who has offered me countless advice and encouragements throughout the project, and has inspired me with her passion for the field. Her patience and insightful questions provided the confidence and direction necessary when I shifted the study towards a systematic review. My thanks must also go to Janice, Xinyi and Nic, without whose help I would not have been able to complete several essential parts of my study.

I am, of course, indebted to my family and friends for their unwavering love and support throughout the past year.

And finally, I thank myself for staying true to the promise I made at the beginning.

Table of Contents

Abstract *5*

Acknowledgements *6*

List of abbreviations *10*

Chapter 1 Introduction *11*

1.1 Grammar and grammar teaching **11**

1.2 Grammar instruction in CLT **12**

1.3 Benefits of the use of L1 ..... **13**

1.4 Explicit grammar in primary and secondary school contexts **15**

1.5 Rationale and outline of the dissertation **16**

Chapter 2 Literature Review *17*

2.1 Operationalising different plurilingual pedagogical approaches **17**

2.1.1 Translanguaging and pedagogical translanguaging 17

2.1.2 Pedagogical translanguaging (PTL) 18

2.1.3 Code-switching 19

2.1.4 Pedagogical translation 19

2.1.5 Use of the L1 20

2.2 'L1-mediated strategy' as the umbrella term **20**

2.3 Existing research **22**

2.4 Review questions **27**

Chapter 3 Methodology *28*

3.1. Systematic Review **28**

3.2 Procedures **29**

3.2.1. Initial search 29

3.2.2. Protocol registration 29

3.2.3. Eligibility criteria 29

3.2.4. Search strategy 33

3.3. Data collection **34**

3.3.1 Title and abstract screening 34

3.3.2 Full-text screening 34

3.3.3 Chain citation 34

3.3.4 Quality assurance 35

3.4. Data extraction **35**

3.5. Critical quality appraisal **35**

3.5.1 Risk of bias (RoB) 36

3.5.2 Data synthesis 37

3.5.3 Confidence of Evidence (CoE) 37

3.6. Summary **38**

Chapter 4 Results ***39***

4.1 Search results **39**

4.1.1 Initial screening 39

4.1.2 Full-text screening 39

4.1.3 Snowballing 40

4.2 Data extraction **42**

4.2.1 Summary of studies characteristics 42

4.2.2 Study and demographic characteristics 42

4.2.3 Outcome characteristics 43

4.2.4 Risk of bias (RoB) in studies 46

4.3 RQ1: Types of L1-mediated strategies **46**

4.3.1 Prescribed labels 46

4.3.2 Classroom implementation 47

4.3.3 Purpose of L1-use 48

4.3.4 Amount of L1 48

4.4 RQ2: Effects of grammar instructions using L1 ..... **49**

4.4.1 Linguistic outcomes 49

Translanguaging 50

Pedagogical translanguaging (PTL) 50

Codeswitching 52

Use of L1 53

4.4.2 Metalinguistic awareness 55

4.4.3 Non-linguistic outcomes 56

Perceived faster rate of learning 56

Perceived slower rate of learning 57

4.5 Chapter Summary **57**

Chapter 5 Discussion ***59***

5.1 RQ 1: L1-mediated strategies **59**

5.1.1 Jangle fallacy 60

5.1.2 Jingle fallacy 61

5.1.3 Function and extent of L1 use 63

5.2 RQ2: Effects on grammar development **64**

5.2.1 Linguistic outcomes 64

5.2.2 Metalinguistic awareness 66

5.2.3 Non-linguistic outcomes - Perceived rate of learning 67

5.3 Contributions of the included studies **67**

5.3.1 Addressing knowledge gaps 67

5.3.2 Limited generalisability 68

5.4 Chapter summary **70**

Chapter 6 Conclusion ***71***

6.1 Summary of findings 71
6.2 Limitations of this review 72
6.3 Pedagogical and research implications 73
References 75
Appendix 1 IDESR Registration Record 82
Appendix 2 Search terms used for specific databases and results returned 89
Appendix 3 Data Extraction Table 91
Appendix 4 JBI Critical Appraisal Tools for Risk of Bias (RoB) Assessment 93
Appendix 5 Reference of all included studies 101
Appendix 6 Results from chained-citation and reasons for exclusion 102
Appendix 7 Excluded studies that are highly relevant but deemed unsuitable 105
Appendix 8 Example entry from data extraction table 106
Appendix 9. Summary table of RoB appraisals for all seven included studies 111
Appendix 10 Sample of a completed RoB Appraisal Form 113

List of tables

Table 1. Inclusion and exclusion criteria. 30
Table 2. Boolean search terms used. 33
Table 3. Summary of studies included. 44
Table 4. Labels and implementations reported. 48
Table 5. Cross-over design adopted in He (2017). 53

List of figures

Figure 1. PRISMA 2020 Flow Diagram (From Page et al, 2021). 41

List of abbreviations

Abbreviations	Full form
CLT	Communicative Language Teaching
EFL	English as Foreign Language
ELL	English Language Learners
ELT	English Language Teaching
ESL	English as Second Language
FFI	Form-focused Instructions
FonF	Focus on Form
FonFS	Focus on Forms
FL	Foreign Language
GTM	Grammar Translation Method
L1	First Language
L2	Second Language
MOI	Medium of Instructions
PTL	Pedagogical Translanguaging
SLA	Second Language Acquisition

Chapter 1 Introduction

Second and foreign language teachers of English often report facing two persistent dilemmas: whether to employ learners' first language (L1) in the classroom (Blackledge & Creese, 2010), and whether to teach grammar explicitly (Ellis & Shintani, 2014a; Ellis, 2016; Pawlak, 2022). These dilemmas are most evident in the use of the L1 to support explicit grammar instruction. Both tensions stem largely from the dominance of the pedagogical approach known as Communicative Language Teaching (CLT), which adopts a naturalistic and immersive approach to language learning (Ellis & Shintani, 2014b). CLT frames L1 use, particular in grammar instruction, as illegitimate (Blackledge & Creese, 2010).

The prominence of CLT has undoubtedly shaped many teachers' beliefs about classroom language use. However, research consistently shows a belief-practice gap: despite the monolingual and implicit learning ideals of CLT, both learners' L1 and explicit grammar instruction remain staples in English as Second Language (ESL) and as Foreign Language (EFL) classrooms (Pawlak, 2022; Macaro, 2005). Teachers and students alike acknowledge the potential of L1 in facilitating grammatical development (Ellis & Shintani, 2014b; Schurz & Coumel, 2023; Sun & Zhang, 2021; Sturm et al, 2024).

To address this gap between pedagogical ideals and classroom realities, the present study adopts a systematic review approach to examine how the learners' mother tongue is used in grammar instruction and its effects. This chapter begins with a brief definition of grammar and common approaches to grammar instruction. Section 1.2 covers grammar instruction in CLT; Section 1.3 discusses the benefits of leveraging students' L1; Section 1.4 highlights the limited research on grammar instruction despite its importance, which forms the rationale for this dissertation. Section 1.5 outlines the organisation of this dissertation.

1.1 Grammar and grammar teaching

Grammar refers to the rules that bind linguistic units – ranging from individual words (e.g. inflections) to multiword units (e.g. formulaic expressions) – together for meaning making (Swan, 2005; Nunan, 2005). Learning grammar involves mapping the form, meaning and function of linguistic structures onto the broader system of a language (Richards & Schmidt, 2011). Accordingly, the role of grammar teaching is to draw learners' attention to specific structures, enabling them to develop metalinguistic awareness, which refers to the knowledge

of language beyond the immediate meaning of those structures in specific contexts, and to process them for comprehension and production (Ellis, 2006; Ellis & Shintani, 2014a). The ultimate goal is internalisation, whereby grammatical knowledge is proceduralised and features can be readily and accurately used in communication (Ellis, 2006).

There are three nuanced yet interrelated concepts under grammar teaching. First, form-focused instruction (FFI) refers to pedagogical events that draw students' attention to language structures (Spada, 1997). It contrasts with meaning-focused instruction, which prioritises fluency and gives little attention to accuracy (Spada, 1997). FFI can be explicit, via direct teaching of grammar rules, or implicit, in the form of an error correction during a communication (Spada, 1997). Second, explicit grammar instruction involves intentional and structured learning of grammar through rules, explanations, and examples (Ellis & Shintani, 2014a). Instruction often employs metalanguage (e.g. using labels such as 'agreement') and is usually followed by practices that help proceduralise knowledge (Ellis, 2006). By contrast, implicit instruction occurs incidentally, without direct explanation or focus on rules (Ellis & Shintani, 2014a). The final relevant concepts are Long (1991)'s focus-on-form (FonF) and focus-on-formS (FonFS) approaches. FonF emphasises teaching grammar in context, whereas FonFS examines the structure in isolation (Long, 1991).

While the terms may appear dichotomous, they should not be treated as binary. Rather, they exist on a continuum (Long, 1991; Ellis & Shintani, 2014a; Ellis, 2016; Pawlak, 2022). For instance, FFI approach can equip beginner learners with the basis of knowledge needed for meaning-focused tasks at later stages of language development (Ellis, 2006). Building on this idea, FonFS and explicit grammar teaching can be introduced as temporary digressions within meaning-focused language tasks, such as consciousness-raising activities (Ellis & Shintani, 2014a). Such instruction helps L2 learners notice grammar patterns, especially the less salient ones and those that cannot be acquired by exposure alone, such as adverb placement (Schmidt, 1990; Pawlak, 2022). Noticing facilitates the transformation of declarative (explicit) knowledge into proceduralised (implicit) and automatised knowledge (Ellis, 2016; Pawlak, 2022). According to Ellis (2016), explicit instruction optimises form-meaning mapping to one's long-term memory, facilitating acquisition.

1.2 Grammar instruction in CLT

Under CLT, the primary goal of English language teaching (ELT) is to foster learners' communicative competence rather than performance, the latter referring to accuracy in

language use (Swan, 2005). CLT advocates L2/FL classroom practices emulate conditions of L1 acquisition, embedding grammar learning in meaning-focused activities and maximising target language use, so that the structures are acquired through exposure (Krashen, 1987; Ellis & Shintani, 2014b).

However, the immersion approach promoted by CLT may not adequately equip learners with structural competence, i.e. the ability to understand and use the grammatical structures accurately (Nagai et al, 2015). Empirical evidence supports this concern: a meta-analysis of 77 studies by Norris and Ortega (2000) demonstrated that instructed learners outperformed their counterparts exposed only to implicit instruction on grammatical measures. Additionally, CLT overlooks a fundamental distinction between L1 and L2/FL acquisition: unlike L1 learners, these learners have limited exposure to the target language and rely heavily on classroom input (Jaekel et al, 2017). As a result, explicit grammar instruction remains an integral component of SLA curriculum, particularly at the secondary school level (Jaekel et al, 2017).

These limitations of CLT highlight the need for alternative means to support grammar acquisition, one of which being leveraging on students' mother tongue. The specific details of such approaches are discussed below.

1.3 Benefits of the use of L1

As aforementioned in Section 1.2, CLT views learners' use of own language a problem and encourages an exclusive use of target language (Ellis & Shintani, 2014b). This monolingual orientation is often grounded in the assumption that immersion maximises exposure to the target language. Krashen (1976) drew on Carroll (1963)'s Time-on-Task Hypothesis and argued that the amount of exposure to the L2 predicts proficiency development. Under this view, the L1 was seen as detrimental to L2 fluency, as it reduced the amount of comprehensible input learners received, and limited opportunities for negotiation of meaning, where learners interact to overcome communication breakdowns, developing fluency in the process (Long, 1983).

Proponents of language separation also associated native language use with off-task behaviour, suggesting that without using L2, students engage less in productive learning (Swain & Lapkin, 2000; Macaro, 2005). Furthermore, alternating between L1 and L2, i.e. codeswitching, in the classroom was argued to cause interference and confuse the learners (Goodman & Tastenbak, 2021, Campa & Nassaji, 2009). Such 'confusion' prevents native-like fluency, which is

regarded as the primary goal of language learning (Ellis & Shantani, 2014b). Despite the lack of empirical evidence substantiating the correlation between L2 exposure and proficiency development (Macaro, 2005), this belief underpins language policies at both classroom and institution level, which mandate the use of the target language only (Blackledge & Creese, 2010).

While immersion-based instruction remains popular, L1 use has almost always been an integral element of L2 language classrooms (Inciman & Akay, 2022). The stance of monolingual bias is increasingly being challenged by scholars, highlighting the cognitive, pedagogical and sociocultural benefits of incorporating learners' L1 into the process of L2 learning. Research suggests that the L1 and L2 draw from a shared pool of cognitive and linguistic resources: Cummins (1979)'s Common Underlying Proficiency suggested that skills, knowledge and concepts learnt in one language can transfer to another, and by inference, the development and maintenance of L1 serves as a cognitive tool that supports that of L2 (Linguistic Interdependence Hypothesis). Various studies (e.g. Bhatnagar, 1980; Chuang et. al, 2012; Yeung et al, 2000) presented evidence of bilingual children outperforming monolinguals in proficiency measures, emphasising the importance of acknowledging the native language of bilinguals.

Incorporating the L1 into the classroom can also promote language learning by supporting cognitive processing. During communication breakdowns, L1 translation and/or codeswitching can reduce the learners' need to allocate selective attention to resolve meaning (Macaro, 2005), therefore lowering the cognitive load on working memory (WM) (Baddeley, 2006). The reduced cognitive demand allows WM to focus on establishing a link between the new L2 input and existing L1 knowledge, facilitating deeper learning and retention (Macaro, 2005). In other words, dual-language use makes input more comprehensible, contributing to the prerequisite of language acquisition proposed by Krashen (1987).

Another advantage of integrating L1 in L2 instruction lies in its sociocultural value. Strict language separation often results in silencing effects, rote memorisation and rigid language routines, conditions which can hinder the language development and academic achievements of bilinguals, especially minority-language speakers (García & Li, 2014; Günther-van der Meij et al, 2020). The L2-only condition is also considered unnatural, as it fails to reflect and authenticate the linguistic reality of bi/multilingual learners (Üstünel, 2016). Strategic use of the students' home language in the classroom affirms their bilingual and cultural identities and validates the multilingual reality outside classrooms (Campa & Nassaji, 2009; Macaro, 2005).

Although the importance of L1 is increasingly recognised, teachers still report reservations towards the bi/multilingual approach (Inciman & Akay, 2022), and feel unprepared and guilty for integrating learners' native language into the classroom (Campa & Nassaji, 2009; Macaro, 2005).

1.4 Explicit grammar in primary and secondary school contexts

L1-mediated interventions targeting vocabulary development have been widely examined (e.g. Chalmers, 2019; Busse et al, 2020). However, Cenoz and Gorter (2020) suggested that the impact of translanguaging, as one of the means of L1 mediation, may not be on vocabulary but on other language domains. Hence, focusing on grammar as an outcome of translanguaging can provide a more nuanced understanding of its effects beyond lexical development. Nevertheless, despite its centrality to language processing and learning, relevant research has been limited due to the dominance of CLT – its emphasis on communicative function of grammar over form and accuracy may have sidelined research on FonFS (Ellis & Shintani, 2014a; Pawlak, 2022). Furthermore, existing research mainly contrasts the effectiveness of implicit and explicit instructions (Pawlak, 2022), leaving the specific role of L1 in mediating explicit grammar teaching largely underexplored.

As the presumably the most up-to-date large-scale methodological synthesis available, Sok et al (2019) replicated the meta-analysis by Norris and Ortega (2000) and coalesced 91 FFI studies published between 1980 and 2015. Sok et al (2019) found that adolescents and children were examined far less frequently (22%) than adult learners (64%), and only a small fraction of studies were situated in KS1-4 contexts. In terms of intervention, the reviewers reported a sharp decline in studies on FonFS instruction, which dropped almost by half from 67% in Norris and Ortega (2000) to 33%. Meanwhile, studies on implicit instructions almost doubled, from 39% to 70%, and FonF studies dominated (68%). These patterns reflect a pedagogical and research shift towards meaning-focused grammar instruction in the field.

However, these trends in research do not always mirror classroom realities, particularly in primary and secondary school contexts. In English language teaching (ELT), many EFL and ESL classrooms and assessments are still preoccupied with the accurate use of grammar structures, i.e. FonFS approaches (Ellis & Shintani, 2014a). This is largely due to time constraints and the demands of high-stakes, form-focused examinations (e.g. Sun & Zhang, 2021; Schurz & Coumel, 2023).

1.5 Rationale and outline of the dissertation

The gap between pedagogical ideals and classroom practice, along with the rise of multilingual education recognising learners' own language provide a clear rationale for investigating the role of L1 in explicit grammar instruction, the focus of the present study (Swain & Lapkin, 2000; Ellis & Shintani, 2014b).

This dissertation comprises six chapters. Following this introductory chapter, Chapter 2 reviews literature on L1-mediated strategies, highlighting research gaps across approaches. Chapter 3 presents the methodological framework and the analytical procedures for the systematic review. Chapter 4 reports review findings synthesised according to key themes emerged from the data. Chapter 5 discusses these findings and their implications for multilingual approaches to explicit grammar teaching. Chapter 6 summarises key contributions of this study and suggests directions for further research. This study, however, does not aim to compare the effectiveness of different L1-mediated interventions.

Chapter 2 Literature Review

Building on the background presented in the introductory chapter, this literature review critically examines current research on mother-tongue-based strategies in FL teaching and learning, highlighting key findings and identifying relevant research gaps. The chapter concludes by presenting two review questions that guide the present inquiry.

2.1 Operationalising different plurilingual pedagogical approaches

A wide range of pedagogical approaches embrace plurilingualism, including the 'use of L1', 'codeswitching', '(pedagogical) translation', and '(pedagogical) translanguaging'. Among these, translanguaging, a relatively new approach (Seals et al, 2020), has gained currency since García (2009) asserted its significance in the classroom. Given its growing prominence and the ongoing debates surrounding it, translanguaging will first be discussed. This section is structured to first provide a brief conceptualisation of each strategy, followed by an examination of how they are operationalised in language classrooms.

2.1.1 Translanguaging and pedagogical translanguaging

The crux of translanguaging pedagogy lies in utilising the entire linguistic repertoire of learners, including home, school, and additional languages, in the process of teaching and learning (Lewis et al, 2012a; Günther-van der Meij et al, 2020). As opposed to relying unidirectionally on the stronger language, presumably the L1, to scaffold L2 development, translanguaging emphasises the fluid mobilisation of linguistic resources to support learning across languages (García & Li, 2014; Cenoz & Gorter, 2020). This flexible use of languages challenges the notion of language compartmentalisation in a bi/multilingual cognitive system: García (2009) proposed dissolving rigid boundaries between named languages and treating them as a unitary repertoire. Cummins (2025), however, warned against assuming a completed integrated system, arguing instead for the coexistence of a shared and language-specific structures in a bi/multilingual brain.

Teacher-led translanguaging involves planned activities with purposeful language switching, such as cross-linguistic comparisons (Nap et al, 2024); student-led translanguaging allows learners to select the most appropriate language for a task (Lewis et al, 2012b). In both cases, the activation of multiple languages is integral. It legitimises the bi/multilingual mode of processing and enables deeper processing, contributing to language development, cognitive growth, and content learning (Lewis et al, 2012a; 2012b; Nap et al, 2024). Translanguaging

also positions multilingual learners as active co-constructor of knowledge (García & Li, 2014; Nap et al, 2024), drawing on their linguistic knowledge, skills, and the associated cultural values to support learning (Duarte, 2020; Seals et al, 2020). It thus acknowledges their sociolinguistic backgrounds and cultural identities (Lewis et al, 2012b).

2.1.2 Pedagogical translanguaging (PTL)

While translanguaging is primarily learner-oriented (Lewis et al, 2012b), pedagogical translanguaging (PTL) gives weight to teacher-led translanguaging in the classroom (Cenoz & Gorter, 2020). PTL is characterised by planned instruction that utilises learners' entire linguistic repertoire for input and output through strategies such as translation and identifying cognates (Cenoz, 2017). By leveraging the instinct of bilinguals in linking their prior knowledge of a language to any new one(s) they are learning, PTL aims to enhance both language development and metalinguistic awareness (Cenoz & Gorter, 2020).

A systematic review by Kim and Weng (2022) considered the implementation of PTL across 20 TESOL studies conducted before February 2022. Rather than adhering to Cenoz's (2017) definition of PTL, which emphasises the intentional and strategic use of the L1, translanguaging was found to be most frequently enacted through spontaneous translation. PTL in EFL contexts facilitated meaning communication, while contributed partially to social inclusion in ESL classrooms. Its adoption in grammar instruction was characterised by translation and/or by providing L1 examples that made the target grammar structures culturally familiar to pupils.

The reviewers highlighted a few discrepancies between the ideals of PTL and its classroom implementation. Across studies, learners' contributions were largely guided by the teacher, and there was a lack of linguistic diversity in the TESOL classrooms surveyed. Regarding linguistic homogeneity, Seals et al (2020) found that classroom interactions dominated by majority-language speakers could lead speakers of other languages to adopt the majority language and refrain from using their own, not to mention sharing it with the class for co-construction of knowledge (Seals et al, 2020). In other words, PTL in practice may not provide learners with full access to their linguistic repertoire (an exception involving Arabic-learning is discussed in Section 2.3). Without drawing on the entirety of learners' linguistic resources, Kim and Weng (2022) argued that PTL was not fully implemented, noting that teachers' perceptions towards translanguaging and bilingualism could inform classroom enactment of PTL, a finding similar to Prilutskaya's (2021).

2.1.3 Code-switching

Cook (2001) considered the act of codeswitching, i.e. the intra-sentential and inter-sentential use of L1 and L2, a highly skilled bilingual behaviour (Üstünel, 2016). The pedagogical application of codeswitching serves the communicative function of ensuring comprehension and compensating for learners' lack of L2 competence (Macaro, 2005). This explains the more frequent, teacher-led use of codeswitching with young and/or emergent bilinguals (Macaro, 2005); Inciman and Akay (2022) also found the strategy to be common for learners whose L1 was underdeveloped. Beyond this, codeswitching can scaffold the acquisition of complex language items, especially difficult grammar constructions (Macaro, 2005; Ellis & Shintani, 2014b). During explicit grammar instruction, teachers were often reported to alternate between languages to 1) aid learners in distinguishing the grammar properties of two different languages, 2) introduce a new item without a parallel form in L2, and 3) explain difficult structures (Üstünel, 2016; Inciman & Akay, 2022), and 4) provide metalinguistic information (Macaro, 2005; Inciman & Akay, 2022). The dual-code input mediates intake, enabling learners to establish L1-L2 connections (Macaro, 2005; Üstünel, 2016).

While translanguaging incorporates the L1 at all stages of language learning (Kim & Weng, 2022), codeswitching views students' own language as progressively less important with increasing proficiency, and FL should ultimately aims for full immersion in the L2 (Macaro, 2005). Yet, current research has not established the optimal amount of L1 use in codeswitching. Two key implications emerge: 1) there is no clear guidance on the proportion of a lesson that can be allocated to codeswitching without impeding L2 development and intake (Üstünel, 2016); and 2) there is insufficient evidence regarding the appropriate degree of codeswitching at different levels of L2 proficiency. In other words, it is unclear when mixed-code input should be replaced by L2-only instruction (Üstünel, 2016). Given that grammatical structures increase in complexity with proficiency, it raises the question of whether the L1 continues to facilitate grammar development at advanced stages of learning, which warrants investigation.

2.1.4 Pedagogical translation

Translation in language teaching (TILT), or pedagogical translation, describes the classroom practice of reproducing words or utterances from one language to another (G. Cook, 2010). According to Guy Cook (2010), TILT augments comprehension through the introduction of L1-equivalents (Lewis et al, 2012b). In grammar instruction, TILT is often integrated into consciousness-raising or language awareness activities that juxtapose the L1 with the target

structure in the new language to highlight distinctive grammar properties (Samardali & Ismael, 2017). However, as denoted by G. Cook (2010), word-for-word translation can be problematic because 1) there is not always a parallel form in the other language, and 2) translation must conform to the target language's grammar.

2.1.5 Use of the L1

The most generic term 'use of the L1' is mentioned last because it is widely applied to capture the essence and benefits of the aforementioned strategies. The term has been introduced in the form of cross-linguistic comparison, codeswitching (also codemixing), and translation through the provision of L1 equivalents (Campa & Nassaji, 2010; Copland & Neokleous, 2011). In grammar instruction, teachers reported deliberately choosing students' mother tongue when explaining complex grammar structures to beginner-level (Campa & Nassaji, 2010) and intermediate learners (Copland & Neokleous, 2011). Adopting learners' own language was deemed particularly efficient and contributed substantially to reducing learner anxiety, as it was similar to their prior language learning experiences (Campa & Nassaji, 2010). It also accommodated to the low level of L2, where synonyms in the new language was not always the most helpful (Campa & Nassaji, 2010). The explicit reference to another language, as argued by Bell and colleagues (2020), enhances bilingual pupil's metalinguistic awareness.

In summary, the discussion on the conceptualisation and implementation of various pedagogies leveraging learners' L1 revealed two key points: 1) bilingual education is receiving growing attention in both research and practice, and 2) there is a range of approaches that share the same overarching goal but differ slightly in their theoretical foundations.

2.2 'L1-mediated strategy' as the umbrella term

Since its popularisation in 2009, translanguaging has increasingly been adopted in place of earlier constructs such as 'use of L1', 'codeswitching' and 'translation' (Treffers-Daller, 2025). There is a trend towards using translanguaging as an umbrella term (Gutierrez, 2021), contributing to its emergence as a 'buzzword' (Goodman & Tastanbek, 2021, p. 30). Nevertheless, treating 'translanguaging' as the overarching framework could be a problematic for the following reasons.

The prime reason lies in the ambiguity of the conceptualisation and operationalisation within translanguaging studies. Translanguaging in its ideal form should involve bi/multilinguals drawing on their entire linguistic repertoire (García & Li, 2014). However, the systematic review by Huang and Chalmers (2023) highlighted inconsistencies in implementation: while

some studies define translanguaging as the use of multiple languages within a lesson or pedagogical task, others extended the concept to curriculum level – they posit that multilingualism should be manifested across broader curriculum, even if only one language is adopted in a lesson (c.f. Cenoz & Gorter, 2022).

Another limitation of employing 'translanguaging' as the overarching label concerns its reporting and presentation in research. Treffers-Daller (2025) argued that much of the literature on L1-mediated interventions adopt the term 'translanguaging' without adapting their analytical frameworks or presenting new findings, nor is sufficient empirical evidence provided. Goodman and Tastanbek (2021), focusing on codeswitching in the era of translanguaging, highlighted the blurred boundaries between the two constructs in both research and practice. They maintained that translanguaging and codeswitching illustrate a pedagogical and ideological shift, and thus should not be treated as hierarchal. Furthermore, Cummins (2025) critiqued translanguaging for assuming bilinguals process and store their languages within one unitary language system. This assumption may lead to the marginalisation of empirical studies that recognise that bilingual cognitive system contains language-specific knowledge (Cummins, 2022).

Beyond critiques of translanguaging per se, categorising all mother-tongue-based strategies under the broad label of translanguaging risks obscuring the conceptual underpinnings behind. Two examples illustrate this. First, unlike translanguaging which considers multilinguals' language knowledge as a collective whole (García, 2009; Chan, 2022), codeswitching and translation treat named languages as compartmentalised constructs (Goodman & Tastanbek, 2021; Lewis et al, 2012b). Second, in codeswitching and translation, the dominant language supports the development of the weaker one[1], whereas translanguaging treats all languages as equally valuable resource (Lewis et al, 2012b). Codeswitching frameworks therefore target a reducing reliance on the L1 in teaching advanced learners (Macaro, 2005), while translanguaging insists on a continued inclusion of learners' own language at all stages of language development (García, 2009; García & Li, 2014).

In other words, adopting translanguaging as an umbrella term would imply that 1) bilingual language systems function as an integrated whole and that 2) translanguaging is measured against other approaches, potentially reinforcing hierarchal assumptions, which is contrary to the goal of this review. To avoid this conceptual flaw, the present study adopts the term 'L1-

[1] The dominant language is not necessarily the L1 of bilinguals, especially among those experiencing L1 attrition.

mediated strategies' to more accurately capture the diverse conceptualisation and pedagogical implementation of L1.

2.3 Existing research

Having outlined the theoretical underpinnings, this section now reviews existing empirical research featuring pedagogical strategies using the pupils' mother tongue.

With the role of one's mother tongue being increasingly acknowledged by researchers and practitioners, L1-mediated pedagogical approaches have been frequently studied. Numerous studies were interested in the teachers' use of first language in the classroom and reported data from classroom observations, conversations and/or discourse analysis (e.g. Campa & Nassaji, 2009; Nap et al, 2024; Izquierdo et al, 2016; Vaish, 2019).

Busse et al (2020) investigated the effect of vocabulary gains using a translanguaging approach vis-à-vis motivational factors, finding some benefits of plurilingual instruction for vocabulary retention. There were greater gains in productive vocabulary use than receptive vocabulary knowledge. This was interesting because the participants had spent less time using the textbook and thus receiving less written input as a result of engaging in the intervention activities. However, in the delayed post-test 4 weeks later, a decline was observed, with a significant drop in productive vocabulary knowledge. The study design targeted both language use (translanguaging vs English-only) and different affective-experimental activities, making it difficult to isolate the effects of language use and motivational factors.

Izquierdo et al (2016) presented classroom data on L1 use by Grade 3 EFL teachers across nine Mexican secondary schools. During FFI, teachers employed English (the target language) four times more frequently than Spanish (the students' L1) to direct attention to form. At first glance, this finding diverged from previous studies reporting teachers' reliance on the L1 for FFI (e.g. Campa & Nassaji, 2009; Copland & Neokleous, 2011). However, further analysis of teacher discourse revealed a functional differentiation between the two languages – English (L2) was systematically used to highlight form-related aspects of the language, for example, to elicit metalanguage terms, provide grammar explanations and introduce new vocabulary for students to repeat or translate into Spanish. Meanwhile, Spanish served more meaning-oriented purposes, often leveraged spontaneously to elaborate on textbook content, which supported content retention and task completion. Nevertheless, as Spanish occurred more frequently than English (n=7,522 vs n=6,427), the researchers characterised the pattern as an overreliance on the L1 and recommended a gradual increase of L2 to develop learners' communicative

competence. However, no actual effect of teachers' language choice on learning outcomes was measured, meaning the claim that L1 use limits communicative competence development was unsubstantiated. In fact, this seems to be an issue with research on L1 use during grammar instruction – regardless of whether their focus was on teachers, students, or both: most studies only collected attitudinal data alongside classroom observations, without triangulation of the effects of L1 use on L2 acquisition (e.g. Abourehab, & Azaz, 2020; Leeming, 2011; Yasar Yuzlu, & Dikilitas, 2021, Meng, 2014).

A study by Abourehab and Azaz (2020) examined PTL in a linguistically heterogeneous classroom at a community language school for Modern Standard Arabic (MSA); an MSA-only policy was upheld there to maximise exposure and support heritage learners' access to Islamic culture and religious texts. Participants included the teacher and novice teenage learners (aged 13-16); they spoke English (their L1 or dominant language) and an Arabic dialect (e.g. Syrian, Egyptian) or a non-Arabic language (e.g. Sudanese, Somali). Classroom observations documented the teacher using English to explain nunation in MSA adverbs, supplemented with English examples and translations while pupils responded with examples in their dialects to confirm understanding. As the only study conducted in a linguistically diverse context, Abourehab and Azaz's (2020) findings underscore the challenges of implementing PTL in a multilingual classroom: While PTL emphasises planned use of learners' linguistic resources (Cenoz, 2017), teacher's and learners' language choices were not always predictable. The observations led the researchers to conclude that PTL cannot be fully 'developed beforehand' (p. 407).

Although the study illustrated how L1 use facilitated learners' trans-/multidialectal competence, it did not measure grammar gains. Among the few that did, the outcomes are mixed. For example, Swain and Lapkin (2000) examined the role of learners' L1 (English) in the acquisition of French morphosyntax, reflexive verbs, in a French immersion classroom. Two classes of 8th graders were recruited and assigned to either a dictogloss (n=30) or a jigsaw group (n=35) and worked in pairs. Learners in the dictogloss group listened to a story in French twice before collaboratively reconstructing it using the notes they took. Dyads in the jigsaw group took turns narrating segments of the same story based on the visual stimulus provide, then jointly produced a written version. Both conditions required the use of French reflexive verbs in the written output. Classroom recordings of the pair-work showed that, in addition to task management, learners used the L1 to access vocabulary, provide metalinguistic explanations, retrieve L2 grammar information, and correct one another. More proficient dyads relied less

on the L1, whereas weaker learners in the jigsaw group depended more on English (their L1) than those in the dictogloss group. The study, however, failed to obtain a significant correlation between the amount of L1 used and the learners' writing task performance. The authors attributed the nil result to the high variability of the two conditions: the visual prompt for the jigsaw group and textual input for the dictogloss group yielded different degrees of L1-use for comprehension and vocabulary access, thereby affecting the comparability of the task performance across the groups.

Scott and Fuente (2008) also explored the role of L1 embedded in a consciousness-raising, form-focused task. In the experiment, 6 pairs of English-L1 intermediate learners of French and another 6 pairs learning Spanish were compared. These participants read a short text to infer the function of two relative pronouns in learners' L2, which was French and Spanish respectively. Three dyads from each language group selected as experiment group interacted for 5 minutes using English and French/Spanish, while the comparison group adhered to L2 throughout. Conversations of the experimental groups were characterised by continued dialogue, with elements of interaction and collaboration, and both members of the dyads contributed in a balanced manner. They made collaborative deductions on the usage of the relative pronouns rather quickly and used metalanguage in their discussion. Conversely, their L2-only counterparts demonstrated a higher variability in contribution and participation, and their interaction was filled with long silences. In the subsequent stimulated recall, the control group expressed frustration due to not being able to verbalise their thoughts in the L2 despite knowing the ideas in their L1. The authors attributed the difficulty displayed by the control group to the high cognitive task demand – although it was evident that both groups relied on L1 in processing, the L2 group had the additional task of processing their L1-driven thoughts into the L2 for output. Interestingly, despite recognising the facilitative effect of the L1, the authors declared themselves 'proponents of nearly exclusive use of the L2 for communicative interactions', arguing against 'random use of the L1 in the FL classroom' (p. 110). Unfortunately, no further justifications on this view or on what classifies as 'random use' were available.

A similar study design was introduced by Bell et al (2020), who compared primary and secondary school students' use of French (L1) in a verbal error correction task in their L2, English. The accuracy of error correction was evaluated vis-à-vis the type of crosslinguistic connections that students demonstrated during the dialogue. Three types of connections were coded: 1) when students articulated a grammar rule and made explicit reference to French (e.g.

for theirs faces: ‘it’s like in French leur face [...] if you have the plural mark on the noun, their it’s like what it’s a determiner’; p. 103), 2) verbalising a rule without explicitly referring to a French parallel (e.g. ‘Non because [...] y ont pas plusieurs faces’; translation: no because [...] they don't have more than one face’; p. 104), and 3) no verbalised rule while French was used. Although the younger group in primary school demonstrated fewer use of crosslinguistic connections than those in secondary schools, learners of both age groups were most accurate when explicitly pinpointing the grammar rule while French was involved.

Another study featuring grammar (Hopp et al, 2022) was conducted in eight EFL classrooms across four German primary schools. Both German-monolinguals and German-dominant speakers of heritage languages were recruited. Two intact classes per school were assigned to either an intervention (n=134) or a comparison condition (n=124). For the treatment group, 20% of the class time was dedicated to PTL activities, such as cross-linguistic comparisons and generating equivalents and further examples of target grammar items. All-English instruction was maintained for the control group. Over 23 weeks, both groups received 35 English lessons (45 minutes each), covering various grammar structures (e.g. forms of address, capitalisation, wh-questions, adverb placements, and double-digit numbers). Post-intervention grammar proficiency test, which measured general receptive grammar knowledge, showed comparable gains across groups. Individual learning gains was significantly predicted by pretest scores, cognitive skills, and the interaction of language background and phonological awareness (a proxy for metalinguistic awareness), but not their language status as monolingual or multilingual speakers of German. The intended delayed measurement was cancelled due to COVID, but results of the immediate test after the 6-month intervention indicated that the short-term learning gains on specific grammar items did not translate to general proficiency growth. Despite that, as the between-group differences were not significant, the authors argued that the reduced exposure to English due to plurilingual activities did not compromise overall FL development. They also suggested that the intervention might have been too brief for effects of interlingual inferencing strategies to surface and become observable in general proficiency measurements.

In a lab setting, McManus and Marsden (2019a) examined the performance of 53 advanced French-L2 university students learning imparfait, the French past tense for habits or ongoing events in the past. Participants were divided into three groups: L2-only (control), L2+L1 practice, and L1+L2. All groups received explicit instruction on imparfait and practised the target in French. The L2+L1 practice group completed practices in both English (their L1) and

French, while the L1+L2 group both received input and practised in both languages. The total time spent on input and practice was balanced across groups. Performance was assessed through reaction time and accuracy during practice, where participants matched the French stimuli with descriptions. The L1+L2 group, in particular, became significantly faster and more accurate across the four sessions, especially in the last two, where earlier content was reintroduced. However, the study did not specify the proportion of L1 to L2 used during explicit input, which could have affected processing (Macaro, 2005). Nevertheless, the findings demonstrated L1-based instruction and practice can contribute to the acquisition of complex grammar structures among highly proficient learners.

The same researchers in another study (McManus & Marsden, 2019b) examined learners' use French imparfait through two oral production tasks. As in their other experiment (2019a), three groups (L2-only, L2+L1 practice, and L1+L2) received input and practised form-meaning mappings either entirely in French or with L1 support, while a fourth control group received no treatment. All treatment groups outperformed the control group in their oral production of imparfait, as measured by appropriateness and accuracy on the immediate posttest. However, only the L1+L2 group maintained the advantage in the delayed posttest at Week 12. The L2-only and L2+L1 practice groups performed similarly to the control group in the delayed measure. Within-group analyses further suggested that L2-only input improved the accuracy with imparfait for habitual actions but not with ongoing actions. The delayed tests results raise questions about the durability of learning gains from L1 mediation and why the benefits persisted for one function of imparfait but not the other. Juxtaposing the two studies (McManus & Marsden, 2019a, 2019b), an important issue is whether the immediate receptive gain, as measured by reaction time and accuracy in (2019a) can be generalised to production tasks (2019b) and be sustained over time. Given that both experiments featured exclusively advanced adult learners, a further question is whether similar effects of L1 use during input and practice would extend to learners of different ages and proficiency levels, or to grammar structures of varying complexity.

In summary, the pedagogical use learners' own language has been widely reported. However, its effects on grammar learning remains unclear. This is partially because studies do not always provide direct, quantifiable measures of learning outcomes of explicit instruction (e.g. Izquierdo et al, 2016; Hopp et al, 2022; Scott & Fuente, 2008). Confounders, such as task design and the proportion of L1 to L2, make it difficult to determine the effects of language choice on grammar development (e.g. Swain & Lapkin, 2000; McManus & Marsden, 2019a).

2.4 Review questions

The notion of target-language-only is challenged by the role of the L1 in L2 learning. Empirical evidence provides grounds for its inclusion in the classroom to facilitate grammar development. Amidst the rising popularity of translanguaging in research and classroom practice, two issues emerge concerning 1) the unequivocal terminologies used and 2) the misalignment between theory, practice and assessment. Using 'L1-mediated strategies' as an umbrella term, the present study aims to provide an overview of approaches that leverage learners' mother tongue in classroom-based grammar instruction, and to examine their impact on English grammar development. It does not, however, attempt to evaluate which approach is more effective.

The two review questions are as follows:

1. What L1-mediated strategies have been used for explicit grammar teaching in primary and secondary English language classrooms?
2. What are the effects of these approaches on English learners' grammatical development?

Given the scope of the study, a systematic review is adopted. The research design will be detailed in the next chapter.

Chapter 3 Methodology

This study adopts a systematic review design, a research method which synthesises evidence available in the field to provide a well-informed, evidence-based answer (Dickson et al, 2024) to the two review questions. The following section provides a detailed explanation of the research design.

3.1. Systematic Review

Systematic review refers to the method of identifying, pooling, appraising and synthesising large bodies of literature (Dickson et al, 2024). This approach has the advantage of producing a more reliable conclusion and eliminating the problem of inconsistency and potential bias in primary findings (Petticrew & Roberts, 2006; Macaro, 2020). Because the research design follows a rigorous, systematic and transparent procedure (Macaro, 2020), it also allows for further reviews and replications (Zawacki-Richter et al, 2020).

The burgeoning volume of literature in the field of Applied Linguistics in recent years warrants the adoption of a systematic review as the research method (Macaro, 2020). First, it addresses the methodological limitations of individual primary studies, which tend to be small in scale (Dickson et al, 2024), and may not capture the multifarious approaches involving the L1 in English grammar instruction. Given their narrow focus, the findings of empirical literature may not be generalisable to those targeting other grammar structures or other L2 contexts (Macaro, 2020). Second, as discussed in the Literature Review, research on L1-use in English grammar instruction remains contentious, with studies often yielding conflicting results. A systematic review is particularly suitable for the topic as it synthesises and objectively evaluates diverse relevant classroom practices (Dickson et al, 2024) under the pedagogical shift to CLT and the rising popularity of (pedagogical) translanguaging in the field. The coalescence of findings can contribute to a richer conceptual understanding of the phenomenon (Macaro, 2020).

By surveying the existing literature (Petticrew & Roberts, 2006; Zawacki-Richter et al, 2020), this research method allows for the identification of possible gaps in the field that need filling, and/or on of possible reasons behind the conflicting results regarding the use of the L1 (Macaro, 2020). A close examination of existing evidence would challenge the assumptions derived from prior practices (Petticrew & Roberts, 2006). It also offers insights to the actual implementation

of L1 use in different language classrooms, thus potentially bridging the research-practice gap (Macaro, 2020).

3.2 Procedures

Procedures for this review were guided by the PRISMA statement for systematic reviews (Page et al, 2021).

3.2.1. Initial search

A search for other systematic reviews on the topic was conducted in May 2025 to identify any existing systematic reviews and avoid duplication. It was performed on several databases related to the Education, Linguistics and Social Sciences disciplines, namely the International Database of Education Systematic Reviews (IDESR), Prospero, EPPI Centre, and Campbell Collaboration Online Library. The search terms 'use of L1', 'English' and 'grammar learning and teaching' were applied. The scoping search yielded one published and one ongoing review. However, both of these reviews focused on impact of grammar in written compositions and neither matched the aims of the current study. To the best of the author's knowledge, the present review explores a relatively under-researched aspect of English grammar teaching that incorporates learners' first language.

3.2.2. Protocol registration

To enhance the transparency of the study, the protocol was registered in IDESR on 8 May prior to the search. It was published on 6 June (Appendix 1).

3.2.3. Eligibility criteria

To identify the most relevant body of work that addresses the review questions, a list of eligibility criteria was developed prior to the search. The relevance of each study was assessed according to the inclusion and exclusion criteria in Table 1. Those studies that failed to meet the inclusion criteria were excluded, with the reasons documented.

Table 1. Inclusion and exclusion criteria.

	Inclusion criteria		**Rationale/Remarks**
Reference	Includes:	Studies with complete reference	
	Excludes:	Studies with incomplete reference	
Time	Includes:	Studies published between 2009 and April 30, 2025	Pedagogical translanguaging became popular at around 2009, so no relevant research was available before then.
	Excludes:	Studies published before 2009	
Publication status	Includes:	All studies regardless of their publication status	To avoid metabiases (Macaro, 2020; Dundar et al, 2024)
Language	Includes:	All studies regardless of the language of publication. Advanced language learners/translators would be consulted for studies written in languages other than English and Chinese	
Study types	Includes:	Research articles and book chapters that report on empirical data	The focus of this study is on effects of L1-mediated grammar instructions in actual primary and secondary classrooms
	Excludes:	Articles and book chapters that do not report on empirical data (e.g. commentary, conceptual or theoretical papers)	
Education level	Includes:	Studies conducted in primary to secondary school English language learning classrooms	

	Excludes:	Studies conducted outside of primary and secondary contexts (e.g. academic language courses offered by tertiary institutions, EAP, ESP courses)	
Research contexts	Includes:	Studies with classroom settings where the English is the second and/or foreign/additional language of the learners	
	Excludes:	Studies conducted in a laboratory	
Participants	Includes:	Studies where the population is largely or exclusively typically developing language learners	For studies that include both typically developing and SEND students in mixed-group settings, only data from the former will be extracted, provided it is identifiable and available.
	Excludes:	Studies featuring exclusively learners with special education needs (SEND) or developmental (language) disorders	
Outcomes	Includes:	Studies that report either a) data on teachers' implementation of multilingual pedagogies based on observed practice, classroom recording, or other forms of empirical evidence; or b) learner outcomes of the L1-mediated pedagogies	As the use of L1 could be embedded in the pedagogical approach of form-focused instructions and/or consciousness-raising activities, studies with these two labels would not be eliminated at the start

	Excludes:	Studies that report data on neither. Examples include impact of corrective feedback, or those that only report stakeholders' perception of the L1-mediated strategies without evidence of implementation or learner outcomes	
Type(s) of intervention	Includes:	Approaches that describe the intervention that focuses on explicit grammar instructions using learners' L1 in an English language learning classroom	
	Excludes:	Studies where learners' L1 is used for other purposes, for example, vocabulary teaching, classroom management, etc	

A research librarian at the Education Library, University of Oxford, was consulted to identify suitable bibliographic databases and develop effective search strategies. Based on the consultation, searches were conducted across databases relevant to Linguistics, Education and multidisciplinary studies:

Education Resources Information Centre (ERIC)
British Education Index
Teacher Reference Centre
Education Abstracts
MLA International Bibliography
Linguistics Collection
Language and Linguistics Behaviour Abstracts (LLBA)

SCOPUS

ProQuest Dissertations and Theses

Chinese National Knowledge Infrastructure (CNKI)

Oxford Research Archive (ORA)

In addition to published work, Oxford Research Archive (ORA) and ProQuest Dissertations were scanned for grey literature. Unpublished work, such as PhD theses, was included as it may contain more recent work (KCL, n.d). Searches were limited to publications in English and Chinese published between 2009 to April 30, 2025.

3.2.4. Search strategy

Boolean operators 'AND' and 'OR' were used to combine search terms. When searching free-text words, wildcard characters were used. The asterisk (*) enabled the searching of a word with different spelling. The search terms used can be found in Table 2 below; specific syntax for individual databases (Dundar et al, 2024) was adapted and the results returned are available in Appendix 2.

Table 2. Boolean search terms used.

L1-mediated pedagogies	(Pedagogical translanguaging OR code-switch* OR code-mix* OR code-mesh* OR cross-linguistic comparison OR us* of L1 OR L1-mediat* OR leverag* L1 OR pedagogical translat* OR comparative grammar OR plurilingual pedagog* OR multilingual pedagog*)
AND	
Grammar teaching and learning	(explicit grammar teach* OR grammar/grammatical learn* OR grammar teaching strateg* OR grammar/grammatical develop* OR syntactic develop* OR morphological develop* OR morphosyntactic develop* OR grammar acqui*)
AND	
Context and participants	(ELT OR "English language teach*" OR ESL "English as second language" OR EFL OR "English as a foreign language" OR EAL "English as an additional language" OR TEFL OR Teaching English as a foreign language* OR TESOL OR "Teaching English to speakers of

	other language*" OR ESOL OR "English for speakers of other language*")
AND/OR	
Learning outcomes	(effect* OR outcome* OR score* OR benefit* OR result* OR perception* OR attitude*)

3.3. Data collection

The search was conducted between 16 May and 20 June 2025. The process involved selecting suitable records through title and abstract screening, then full-text screening, and citation chaining.

3.3.1 Title and abstract screening

Search results from the databases were first uploaded to Rayyan, a systematic review management tool (Ouzzani et al, 2016), for deduplication. The remaining records were then screened independently by the author by applying the eligibility criteria stipulated in Section 3.2.3. Titles and abstracts that did not meet the inclusion criteria were excluded. Records that could not be confidently assessed based on this information were retained for the next stage.

3.3.2 Full-text screening

Prior to the start of this stage, full reports of all potentially eligible studies were retrieved and uploaded to Rayyan. The selection criteria were then applied to identify records that clearly met all the criteria for the review.

3.3.3 Chain citation

To ensure a comprehensive search, citation chaining was conducted. Backward searching, which involved examining the bibliography of eligible studies to identify additional relevant publications, was performed. Forward searching, which entailed locating studies that had cited the included publications, was also introduced using Google Scholar (Dundar et al, 2024).

3.3.4 Quality assurance

To ensure the reliability of the screening process, a random sample of 20% of the titles and abstracts (n=227) and full-text articles (n=21) was independently reviewed by a peer, who is a graduate student in Applied Linguistics. Any discrepancies were discussed and resolved collaboratively until an interrater agreement of 90% was reached – a common threshold for high reliability in systematic reviews (Higgins et al, 2022). Studies meeting all inclusion criteria were then advanced to the next stage for data extraction and synthesis.

3.4. Data extraction

A data extraction table was developed with reference to Cochrane Effective Practice and Organisation of Care (EPOC) (See Appendix 3). The table for data extraction was piloted with two of the included studies to ensure all information related to the review questions was captured accurately and comprehensively (Fleeman & Dundar, 2024). A column for the L1-mediated strategy label, and for how it was reportedly implemented were added to address the two review questions.

Both descriptive and analytical data relevant to the two review questions were collected and recorded from the included studies, if available. Descriptive data, which included information related to the study characteristics (e.g. author(s)' name(s), year of publication), participants (e.g. L1, length of ELL, sample size), and study design (e.g. setting, duration, intervention(s) and comparator(s), if appropriate), were extracted. The specific L1-mediated strategy stated, and its actual implementation reported, as well as the grammar focus for each study were also recorded.

The reported outcomes of L1-mediated grammar instruction, including primary and secondary findings, and relevant statistical analyses were also documented. The quantitative results (e.g. learning outcomes of the intervention) and/or qualitative results (e.g. interview data) were clearly labelled. Since RQ2 concerns outcomes of explicit grammar instruction, information related to participants' language preference was deemed irrelevant and was not coded.

3.5. Critical quality appraisal

A quality assessment was then conducted to critically evaluate the design, implementation and reporting of each included study (Greenhalgh & Brown, 2024). The validity of a study depends

on the extent to which it minimises bias and error in its design, conduct and analysis. Hence, studies with high quality ratings are in general more accurate in representing the impact of the intervention and are more credible (Greenhalgh & Brown, 2024). Quality appraisal helped ensure that the conclusions drawn from the evidence were well-founded and contributed to more confident answers to the two review questions (Greenhalgh & Brown, 2024).

Quality appraisal in this systematic review was intentionally introduced after data extraction. The reasons behind this decision were twofold. 1) It aimed to minimise bias in selection and reporting in the review (Greenhalgh & Brown, 2024), which was particularly crucial as the current study sought to develop an understanding of all grammar instruction involving L1 use (RQ1). 2) It allowed for a more informed examination of the quality of work/reports given the reviewer's familiarity with individual research reports following data extraction, again enhancing the robustness of the conclusion this systematic review is attempting to draw (Petticrew & Roberts, 2006).

3.5.1 Risk of bias (RoB)

An important element of the quality appraisal process is assessing bias (Greenhalgh & Brown, 2024). As the first review questions focuses on existing approaches to English grammar teaching using learners' L1, the risk of bias assessment was not relevant. For the second review question, which concerned the effect(s) of L1-mediated strategies, bias assessment was indispensable to appraise the overall credibility, transparency and methodological soundness of individual studies (Greenhalgh & Brown, 2024).

Instead of using a generic tool to appraise study quality, design-specific assessment tools were employed according to the study design. There are two benefits to this approach. First, it recognises elements unique to each study type, ensuring that key strengths and limitations of individual study designs are not masked (Greenhalgh & Brown, 2024). It also facilitates a more accurate evaluation of the design, scale, fidelity and validity of each study (Greenhalgh & Brown, 2024).

Appraisal tools developed by the Joanna Briggs Institute (JBI), a research organisation that develops evidence-based information to improve healthcare practice, were selected for the RoB assessment (Refer to Appendix 4 for specific appraisal tools chosen). JBI appraisal tools are commonly used for assessing RoB in education-related studies (Greenhalgh & Brown, 2024).

The Education Endowment Foundation (EEF) Padlock Rating, initially chosen to assess qualitative and mixed methods studies research, was replaced by the relevant JBI tool. Although the EEF Padlock is designed for educational research, it places greater emphasis on quantitative data and may inadequately capture the nuances of qualitative findings.

3.5.2 Data synthesis

Narrative synthesis was adopted to synthesise data to address both review questions. This textual method is particularly appropriate for exploring the heterogeneity across studies, and enables the integration of diverse types of evidence, instead of intervention effectiveness alone (Petticrew & Roberts, 2006). It also facilitates the interpretation of findings across studies. The four elements of a narrative synthesis include: 1) developing a theory of how the intervention works, why and for whom across multiple studies, 2) producing a preliminary synthesis of findings, 3) exploring relationships within and across studies, and 4) assessing the robustness of the synthesis (Popay et al, 2006).

A meta-analysis was not performed: as the selected studies employed different research designs, combining RCTs and non-RCTs in the same meta-analysis could potentially inflate effect sizes (Petticrew & Roberts, 2006). Furthermore, the heterogeneity of the statistical data precluded a meaningful statistical synthesis (Brown & Chaplin, 2024). The studies featured measured a wide variety of learning outcomes, including receptive knowledge, learner awareness, reaction times, and the frequency and accuracy of target structure use. Differences in assessment tools and participant characteristics (e.g. length of ELL instructions, English proficiency) further limited the comparability across studies. As a result, a narrative synthesis was deemed more appropriate for summarising and interpreting the findings, which were characterised by substantial variability.

3.5.3 Confidence of Evidence (CoE)

The GRADE Approach was employed to assess confidence in cumulative evidence derived from quantitative data (Neumann et al, 2025). The robustness of findings across studies would be downgraded for concerns related to inconsistency, imprecision, risk of bias, indirectness or publication bias. Likewise, qualitative evidence was evaluated using the GRADE-CERQual framework. CoE would be lowered because of methodological limitation, lack of coherence, adequacy and relevance using GRADE-CERQual (Lewin et al, 2018).

3.5.4 Quality assessment of this systematic review

The final component of the systematic review was a critical evaluation of the quality of this review per se. The PRISMA checklist was employed to ensure the robustness of the review process (Page et al, 2021), that it was transparent, rigourous and replicable such that the findings could be generalised/transferred to other studies/contexts (Macaro, 2020).

3.6. Summary

Adopting the method of systematic review allowed for a rigorous approach in addressing the two review questions, which were to map existing L1-mediated strategies used for explicit grammar instruction in ELL classrooms on the primary and secondary school level, and to explore their respective effects on learners' grammatical development. By adhering to robust, transparent procedures, it was possible to draw relevant conclusions from the findings of primary studies.

Chapter 4 Results

Building on the methodology outlined in the previous chapter, this section presents findings in relation to the two review questions, namely 1) what L1-mediated strategies have been used for explicit grammar teaching in primary and secondary ELL classrooms, and 2) what the effects of these approaches on ELL learners' grammatical development are.

This chapter first documents the search results in Section 4.1 and data extraction in Section 4.2, then key findings in Sections 4.3 and 4.4. The data drawn from the included studies are analysed by themes, according to the type of L1-mediated strategies adopted (Popay et al, 2006).

4.1 Search results

The initial literature search resulted in a total of 1,223 records. After deduplication on Rayyan, there were 1,145 for screening.

4.1.1 Initial screening

At the title and screening stage, 1,027 records were excluded for failing to meet the eligibility criteria in the following areas: outcome (n=730), population (n= 174), publication type (n=103) or study design (n=20). Studies that could not be confidently excluded based on the information in the title and abstract (n=118) were kept for the subsequent full-text screening stage. Most of them featured consciousness-raising activities, form-focused instructions, and collaborative learning activities without specifying whether the L1 was involved. As a fidelity check, 20% of the titles and abstracts (n=227) were randomly selected and screened by a peer. The interrater agreement was 99%.

4.1.2 Full-text screening

The full version of the reports that were potentially eligible (n=118) were obtained before screening. 110 records were deemed unsuitable for this review for the following reasons: no L1 use (n=55), unrelated to ELL (n=6), unrelated to grammar (n=18), no implementation (n=16), and indirect measures (e.g. only using general proficiency tests to measure learning gains from interventions) (n=15). Among those labelled as 'indirect', many included a high level of variability, for example other teaching activities were involved; they were eliminated

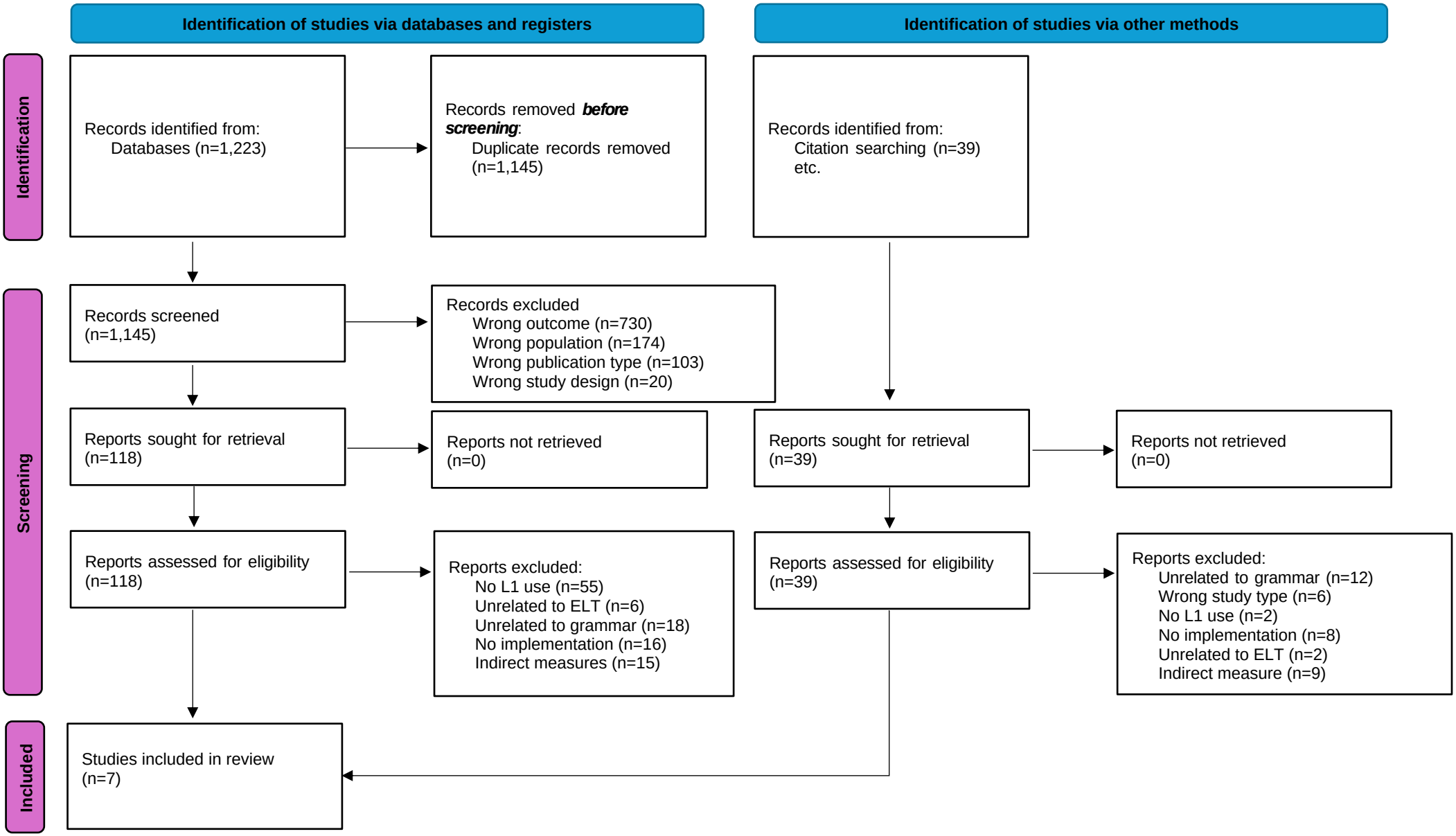
as it was difficult isolate results of language choice from the activity. On the other hand, eight were considered completely relevant to the two review questions. One study (Hosseininik & Sangani, 2014), however, was excluded as no clear demographic information about the participants' age was available despite efforts in contacting the contributing authors. Hence, a total of seven studies were selected for data extraction (Full reference in Appendix 5). 21 full-text articles (20%) were reviewed by the same peer rater. An interrater agreement of 95% indicated that the eligibility criteria had been applied rigorously.

4.1.3 Snowballing

To ensure the comprehensiveness of the search, citation chaining, i.e. backward and forward searching, was then conducted based on the seven included studies. Among the seven records, Hopp and Thoma (2021) was part of a research project called MEG-SKoRe. Therefore, the publications on the project website were also scrutinised. None of the 39 results returned from snowballing reported the use of the L1 in English grammar teaching or learning in classroom contexts at primary and/or secondary levels (See Appendix 6 for specific reasons for exclusion). The total number of studies for synthesis remained seven.

The entire selection process was recorded and reported in the PRISMA flow diagram in Figure 1 (Moher et al, 2009); specific reasons of exclusion is also documented. Several highly relevant titles were excluded due to methodological limitations (e.g. mixed-age groups without subgroup analysis, self-reported gains, or general grammar measures). Details are provided in Appendix 7.

Figure 1. PRISMA 2020 Flow Diagram (From Page et al, 2021).



4.2 Data extraction

4.2.1 Summary of studies characteristics

The seven studies suitable for this review were published between 2015 and 2023; five were peer-reviewed journal articles and two were master's theses. Research designs included (clustered) randomised control trial (RCT) (n=3), quasi-experimental (n=2), and mixed-methods approaches, presenting both quantitative and qualitative data (n=3, including one that was counted as RCT). All but one (Khonamri et al, 2022) used between-subjects comparison; the exception adopted a within-subjects design, comparing participants' performance in different intervention conditions (Dörnyei, 2007). A cross-over design (He, 2017) was also documented. Intervention duration ranged from 24 minutes to 5 weeks; the two-lesson experiment by Hopp and Thoma (2021) was conducted within the 6-month MEG-SKoRe II project on plurilingual FL teaching among beginner learners. Consistent with trends in Applied Linguistics research, cross-sectional research dominated (Dörnyei, 2007). Only immediate post-intervention outcomes were reported; no long-term data were presented. The data extraction table is provided in Appendix 3, with a sample entry in Appendix 8.

4.2.2 Study and demographic characteristics

Five studies were conducted in secondary school classrooms and two in primary contexts, with participants aged between 8 and 17. Sample sizes of ranged from 25 to 258 pupils. Although the review aimed to cover both ESL and EFL contexts, all results returned were set in EFL contexts, featuring classrooms in Germany, Iran, China, Spain and Italy. Most samples consisted of monolingual learners of English, except for the language minority group in Hopp and Thoma (2021), which featured participants from diverse L1 backgrounds, and the Turkman-Persian bilinguals in Khonamri et al (2022). Gallagher and Colohan (2017) was carried out in a Geography classroom under a Content-Language Integrated Learning (CLIL) curriculum, while Fan (2022) introduced special arrangements due to COVID-19. The participants' length of EFL instruction and proficiency level were not always reported, but most were classified as elementary to intermediate learners.

A wide range of grammatical structures was investigated, including verb-level (modal perfect, past perfect tense, auxiliary 'be' in 'there be'), clause-level (relative clause, conditional Type 2, compound sentences), and syntax-level (word order in wh-questions, adverbs of frequency and place, and question tags) items (Swan, 2016). Most studies focused on a singular

grammatical structure except for Khonamri et al (2022), which tested six structures, and for Hopp and Thoma's (2021), which tested two.

4.2.3 Outcome characteristics

Both linguistic and non-linguistic outcomes were examined. Language outcomes can be further divided into receptive and productive grammar knowledge. The former was mainly measured through multiple-choice questions, gap-filling tasks, and reaction time and accuracy in Grammaticality Judgement Tests (GJTs); productive knowledge was assessed through translation tasks, sentence construction, and short writing tasks using the target structures. Fan (2022) also analysed the frequency and accuracy of target-structure use during peer interactions. Potential transfer effects were explored in Hopp and Thoma (2021), who incorporated relative clauses and wh-questions as non-target items in the posttests. Alongside learning gains, learners' perception towards the interventions and the effects of grammar instruction were also frequently investigated through questionnaires and interviews. A summary of study characteristics can be found in Table 3.

Table 3. Summary of studies included.

Study ID	Design	L1-mediated strategy labelled by authors	Classroom activity	Grammar focus	Duration	Location	Primary/ Secondary	Sample size; Age by years	Learners' L1	Length of ELL/Eng level	Results of treatment	Risk of bias
Mehrseresht et al (2015)	RCT (pre-posttest)	Use of L1	Unclear	Items taken from Lessons 6-8; exact item unspecified	5 weeks	Iran	Secondary	50, all female/14-15	Persian	Unclear	Significant*	High
Gallagher & Colohan (2017)	Quasi-experimental qualitative (between-subject comparison)	Codeswitching/ translanguaging	Translation	Positions of adverbs of frequency and place	2 lessons	Italy	Secondary	59; 14-15	Italian	CEFR A2 or higher	Reported higher cross-linguistic awareness via interviews	High
He (2017)	Quasi-experimental mixed methods (pre-posttest)	Code-switching	Code-switching	Attributive clause with relative pronoun and adverbs	3 45-minute periods for two weeks	China	Secondary	112; 16	Mandarin Chinese	Elementary (Senior middle school entrance exam)	Mixed results	Moderate
Hopp & Thoma (2021)	Clustered RCT (pre-posttest)	Pedagogical translanguaging	Cross-linguistic comparison	Wh-questions, passivisation	2 45-minute lessons, 20% PTL	Germany	Primary	258; 8-9	German L1: 141 (66%), non-German L1: 117 (33%)	1 year 3 months	Significant	Low
Khonamri et al (2022)	Quasi-experimental mixed methods (within-subjects design)	Translanguaging	Unclear	Passive, tag questions, compound sentences, relative clauses, Type 2 conditionals, past perfect tense	4 hours/week, 6 consecutive sessions	Iran	Secondary	25, all male; age not specified but all were in Year 12	L1: Turkman L2: Persian (Farsi) – National language of Iran/MOI	Intermediate (Oxford Placement Test)	Significant	High

Fan (2022)	RCT mixed methods (pre-posttest)	Pedagogical translanguaging	Cross-linguistic comparison, code-switching	There ‘be’	24 minutes	China	Primary	145; 12-13	Mandarin Chinese	3 years	Null effect	Low
Shabaka-Fernández (2023)	Quasi-experimental (pre-posttest)	Use of L1	Code-switching, translation,	Modal perfect	55-minute session, 3 times a week for two weeks	Spain	Secondary	60; 15-17	Spanish, except two (3.3%) whose L1 is unknown	Unclear	Null effect	Moderate

* A potential interpretation error is found. In Table 5 (p.224) of the original study, the variable ‘q1’ was not clearly defined but likely represented pretest scores. The reported statistics show F=64.287, p<.001 for q1 (significant), and F=3.768, p=.058 for instruction (not significant). Despite this, the authors claimed instruction was effective in promoting grammar learning, suggesting a possible conflation of results with instructional effect.

4.2.4 Risk of bias (RoB) in studies

The quality of the eligible studies was assessed using the respective JBI Critical Appraisal Tools: the results were mixed. Of the seven empirical studies, only two, both clustered RCTs, were considered robust due to rigourous research design, methodology and data analysis. Two studies demonstrated a moderate RoB because of potential methodological flaws in baseline control, test conditions and reporting, which undermined the power of their claims. Three were rated low in trustworthiness because of possible bias in reporting and study design. Results of RoB appraisal are summarised in Table 3 above; the detailed RoB ratings for each study is available in Appendix 9.

4.3 RQ1: Types of L1-mediated strategies

The first review question concerned what explicit grammar instructions involving learners' L1 have been used in primary and secondary classrooms. Based on the labels provided in individual studies, the reported strategies include:

use of L1 (Shabaka-Fernández, 2023, Mehrseresht et al, 2015),
code-switching (He, 2017),
translanguaging (Khonamri et al, 2022),
pedagogical translanguaging (PTL) (Hopp & Thoma, 2021; Fan, 2022)
code-switching/translanguaging (Gallagher & Colohan, 2017)

PTL was implemented only in primary classrooms, whereas a wider variety of L1-facilitated grammar instruction was observed at the secondary level.

4.3.1 Prescribed labels

Translanguaging, be it the plain term or with the preceding label 'pedagogical', was defined and interpreted differently across studies. Drawing reference to Baker (2011, c.f. Wright & Baker, 2021), Gallagher and Colohan (2017) defined translanguaging as one of several 'bilingual instructional techniques' that involved the 'strategic and judicious use of the L1' (p. 487). According to the authors, translanguaging could be enacted through codeswitching, translation, and other means. Nevertheless, they did not make a distinction between codeswitching and translanguaging and used the two labels in tandem. In contrast, Khonamri et al (2022) operationalised three experimental conditions as 'L1-use', 'L2-use' and

'translanguaging' (involving L1, L2 and English); they treated translanguaging as a distinct construct, albeit not entirely independent of L1 or L2 use.

Similarly, Fan (2022) applied a stricter conceptualisation of PTL, emphasising the use of learners' entire language repertoire for input and output (Cenoz, 2017). She noted a slight deviation from PTL in Hopp and Thoma (2021): their use of language comparisons to raise students' metalinguistic awareness goes beyond the original goal of translanguaging (Fan, 2022, p. 21); Fan, however, did not clearly distinguish translanguaging and PTL in her report. These suggest the interpretations of PTL vary, with some applying the concept to the development of metalinguistic awareness beyond the specific language item, while others stress PTL's function in structuring planned input and output.

4.3.2 Classroom implementation

While multiple labels were used to describe plurilingual instruction, the variety of actual classroom practices was relatively less diverse. The most common way the mother tongue was introduced was codeswitching, followed by cross-linguistic comparison. In interventions involving codeswitching (e.g. He, 2017; Fan, 2022; Shabaka-Fernández, 2023), the MOI shifted from English to the learners' L1, and learners could also code-mix/switch in their responses. In Gallagher and Colohan (2017), code-switching/translanguaging was operationalised in the form of translation – participants translated aural stimuli from English into Italian, their L1. In Shabaka-Fernández (2023), participants translated examples containing modal perfect into their L1. Cross-linguistic comparison was sometimes facilitated by the teacher (Hopp & Thoma, 2021; Fan, 2022) or performed independently by students (Gallagher & Colohan, 2017). Some studies (Mehrseresht et al, 2015; Khonamri et al, 2022) did not describe specific practice employed. In addition to verbal instruction, the L1 was also presented in written form in teaching materials or task sheets (He, 2017; Fan, 2022).

Juxtaposing the labels and their implementation, it can be observed that one L1-facilitated strategy may be associated with multiple classroom practices, and vice versa. This many-to-many mapping highlights overlap among the labels themselves and among the classroom practices (Günther-van der Meij et al, 2020).

Table 4. Labels and implementations reported.

Labels	Implementation (Corresponding studies)
Translanguaging	Combined use of L1, L2 and English (Khonamri et al, 2022) Translation (Gallagher & Colohan, 2017)
Pedagogical translanguaging	Cross-linguistic comparison (Hopp & Thoma, 2021; Fan, 2022) Eliciting L1 equivalents/examples (Hopp & Thoma, 2021) Code-switching (Fan, 2022)
Use of L1	Exclusive use of L1 as MOI (Khonamri et al, 2022) Code-switching (Shabaka-Fernández, 2023) Certain degree of L1 use (Mehrseresht et al, 2015)
Code-switching	Code-switching (He, 2017)

4.3.3 Purpose of L1-use

The L1 served different purposes across studies. Among those that clearly described the circumstances of L1 use, learners' native language, and sometimes, other languages known, was elicited to contrast the target English grammar. The rationale was to help students notice cross-language differences in form (Hopp & Thoma, 2021; Gallagher & Colohan, 2017; Shabaka-Fernández, 2023) and meaning (Fan, 2022). For example, in Gallagher and Colohan (2017), word-by-word translation highlighted differences in word order for adverbs of frequency and place between English and Italian. Likewise, Fan (2022) compared the target form 'there ***is/are***' and the L1-driven error 'there ***has*** (an apple)' in a video shown prior to an interactive activity.

Another function of the L1 was to provide explanations of grammar rules and usage (He, 2017; Shabaka-Fernández, 2023). English, on the other hand, was used for giving examples of the target structure(s) and other purposes such as classroom and task management. In He (2017) and Fan (2022), metalinguistic terms such as 'subject' and 'participles' were presented in the L1 when explaining the forms of attributive clauses, and the 'there "be"' structure. However, grammar rules, be it in English or L1, were not always explicitly taught (Hopp & Thoma, 2021).

4.3.4 Amount of L1

An important finding was the proportion of English to L1 switching. While some studies offered quantifiable data, others included only qualitative descriptions of when language switch

occurred; several omitted such details entirely. For example, Hopp and Thoma (2021) stated that approximately 20% of the two 45-minute intervention sessions involved plurilingual elements, while Gallagher and Colohan (2017) provided indirect evidence by including at least one adverbial per line in their 'twisted dictation' task. In contrast, studies like He (2017), Fan (2022), and Shabaka-Fernández (2023) described L1 use only functionally – e.g. when the teacher explained grammar rules or when students translated the examples (and rules) – without stating the frequency or duration. Poorer quality studies (e.g. Mehrseresht et al, 2015; Khonamri et al, 2022 for the translanguaging condition) failed to report when or how frequently the L1 was used relative to English, making it difficult to assess whether the L1-based instruction was structured or incidental. Also notable is Fan (2022)'s study – while all three groups (PTL, English-only, and control) received L1-mediated input, the Translanguaging and English-only groups engaged in an additional 12-minute practice. This design introduced a difference in exposure between the groups.

In summary, the collation of studies highlighted two important aspects of L1 use in grammar instruction. One, a variety of terms have been employed to describe similar or distinct classroom practices. Second, both the extent and purpose of L1 use differ across studies, which could have potential influence on the effects of the interventions and could complicate direct comparisons of results.

4.4 RQ2: Effects of grammar instructions using L1

The second review question examined the effects of L1-mediated grammar instruction. Changes in grammar knowledge were measured through tests, reaction times, or the number of correct applications of target structures during peer interactions. The impact of the intervention also extended beyond grammar knowledge, with participants' perceived effect on their grammar development studied. However, as noted in Section 4.2.5, the trustworthiness of the studies warrants caution when interpreting the actual effects. The findings will be organised according to L1-mediated strategy, adhering to the prescribed labels in the original studies.

4.4.1 Linguistic outcomes

Different L1-mediated strategies yielded differing outcomes in grammar knowledge development.

Translanguaging

The two studies discussed the effects of translanguaging. The evidence of learning and increased form awareness presented, however, was of limited quality because of the high risk of bias.

Khonamri et al (2022) distinguished between the multilingual strategy of 'translanguaging' and the exclusive use of students' L1 (Turkman) and L2 (Persian, the national language of Iran). They found that grammar instruction was most effective when all three languages were incorporated. Passive voice and question tags were taught solely in Persian (L2), compound sentences and relative clauses in the L1. If-conditional Type 2 and past perfect, identified by the schoolteacher as 'the most problematic and confusing structures' (p.4), were taught through translanguaging. Two separate grammar tests were administered after each language strategy, so a total of six tests took place. A within-subjects repeated measures ANOVA was used to compare results. With only the mean scores of the two tests from each phase computed for comparison, it was observed that learners performed best in translanguaging phase, then L1-only, and finally L2. It was concluded that 1) L1 facilitated grammar development, and 2) translanguaging was more effective than relying only on one language.

Several limitations were identified – It was unclear whether the test items assessed corresponded to the content of each intervention phase, nor was there a control of possible confounding variables (e.g. practice effect). Combined with the lack of transparency in outcome measurement, the study indicated a serious risk of bias and indirectness. The certainty of evidence for the effects of translanguaging was hence rated as low.

Pedagogical translanguaging (PTL)

Two robust studies reported effects of PTL in primary schools, though their findings were mixed.

Hopp and Thoma (2021) examined the potential impact of multilingual instruction on learners' receptive knowledge of English grammar structures. Of particular interest was whether the multilingual instruction was modulated by the differences and similarities between English and German (learners' L1), and whether PTL would benefit bi/multilingual learners of English more than their monolingual peers. In Study 1, wh-questions were introduced, as question formation in English involves changes in word order in marking noncanonical orders, whereas German does not. Object questions (e.g. Which animal does the horse bite?) were particularly

challenging to German speakers. In Study 2, passivisation was introduced to explore whether cross-linguistic similarities facilitate learning, as both languages share the feature of overt word-order changes. Learners' comprehension of syntactically similar items – relative clauses in Study 1 and wh-questions in Study 2 – was also assessed to examine if learners could transfer gains from instruction to these non-target items.

258 fourth-graders from four public primary schools in Germany were recruited; 33% were German-dominant bilinguals learning English as their third language, while the majority were monolingual German speakers. About 20% of instruction involved PTL, during which teachers elicited equivalents in German and other home languages. Participants were then guided to notice structural similarities and differences across languages to infer form and meaning, without explicit presentation of grammar rules (Ellis & Shintani, 2014a).

Although both groups struggled with object-questions in Study 1, the PTL group showed greater gains (p=.001). The PTL group achieved a medium effect size, p=.029, d=0.466, 95% CI [0.110, 2.012], while the control group showed a small, non-significant effect, with CI crossing zero, p=.626, d=0.134, 95% CI [–1.279, 0.769]. Similarly, both groups improved in passives in Study 2, with a medium effect size for each.

Learning gains from instruction, however, did not transfer to non-target items. Additionally, no significant differences emerged between monolingual and their bi/multilingual counterparts in both experiments. The authors suggested this could be attributable to bi/multilinguals being German-dominant. Hopp and Thoma (2021) also concluded that PTL may compensate for the reduced amount of English input and support grammar learning without requiring extra time, and without explicitly presenting grammar rules.

Fan (2022) also provided quality evidence of PTL influencing primary school ELL learners' grammatical development, specifically on the copula 'be' in 'there+be'. Contrary to Hopp and Thoma (2021), a nil result was reported for between-group comparisons despite overall gains. 145 beginner-learners of English in a Chinese primary school were randomly assigned to a Translanguaging group (n=50), an English-only group (n=50), or a control group (n=45). All groups first watched a 12-minute bilingual video on 'there+be' in the existential clause (Swan, 2016) (e.g. 'there **is** a dog'), which is commonly mistaken as the erroneous form 'there **has** a dog' among Chinese learners. The video also compared the target with the stative verb 'have' when expressing possession in 'someone **has** a dog'. Accompanied by bilingual slides, rules

were explained in Chinese while examples in English. Afterwards, the Translanguaging and English-only group paired up within their respective groups and completed a 12-minute information gap activity called ‘spot the difference’ to practise the structure; the control group waited in another classroom. Learning outcomes were assessed through an oral translation test (TT), a Grammaticality Judgement Test (GJT, measured in accuracy and reaction time), and an analysis of the frequency and accuracy ‘there be’ in the interaction task.

Pre-posttests results indicated all three groups improved. For the PTL group, mixed-ANOVA analyses of performance in TT and GJT showed effect sizes ranging from marginally significant to medium. However, there was no significant interaction between Group and Test, suggesting comparable gains across conditions. Fan therefore concluded that PTL did not hinder the acquisition of ‘there be’, even though the intervention was short (24 minutes), but the approach could not be shown to be more effective than monolingual instruction in teaching ‘there be’. Notably, the researcher acknowledged that the bilingual video shown to all groups may have influenced the results.

This empirical study was also the only among all seven included studies to have examined student-led translanguaging, when the Translanguaging group engaged in the ‘spot the difference’ activity. Although it was the teacher who set the restriction of allowing Chinese only for unknown vocabulary, student-led translanguaging was evident because learners had the autonomy to decide when to introduce L1 and the amount used. That said, instances of translanguaging were reported among the English-only group, when pupils used Chinese pinyin in an English accent (p. 49, 54), literal translation for vocabulary items (p. 55) and body language. The translanguaging group produced more target structures and fewer incorrect uses than the English-only group ($t=-1.885$, $p=.033$).

Although findings of Fan (2022) and Hopp and Thoma (2021) differed, both studies were judged to provide high-certainty evidence.

Codeswitching

This strategy refers to the alternating use of English and the L1. In a moderate-quality study, He (2017) found that code-switching contributed to the learners’ receptive but not productive use of attributive clauses (also called relative clauses). 112 first-year senior high school students (average age: 16) were divided into two groups of 56 members. A cross-over design was employed:

Table 5. Cross-over design adopted in He (2017).

Phase	Relative pronouns modifying	Group A	Group B	Test
1	Person or thing (e.g. *who, whom, whose, which, that)*	English-only	Codeswitching	1
2	Place, time or reason (e.g. when, where, why)	Codeswitching	English-only	2

In Phase 1, Group A (n=55, with one absentee) received instruction entirely in English, whereas Group B (n=56) was taught attributive clauses through codeswitching. After completing Test 1, the groups proceeded to attributive adverbs, with the language conditions reversed (see Table 2). Receptive competence was assessed through sentence-level multiple-choice and gap-filling questions, while their productive ability was measured via paragraph-level gap-filling, translation, and sentence rewriting tasks. Dictionaries and reference books were permitted in Test 1 but not in Test 2.

An independent samples t-test of Test 1 showed that Group A (English-only) scored significantly lower than Group B (codeswitching) on receptive knowledge, $p<.001$. The opposite pattern emerged for the productive use of relative pronouns: Group A significantly outperformed their codeswitching counterpart, $p=.009$. A similar pattern was observed in Test 2. Group A, now receiving bilingual instruction on relative adverbs, scored higher than Group B (English-only) in the receptive component, $p<.001$. In the production section, Group B, taught under monolingual instruction, scored significantly higher than Group A, $p<.001$. It was concluded that codeswitching facilitated the comprehension of relative clauses, but it was not as effective as the English monolingual strategy for production.

Despite the variation in test conditions (i.e. dictionary use in Test 1 but not Test 2), the cross-over design mitigated some of the risk of bias. The overall certainty of evidence, therefore, remained moderate.

Use of L1

Mixed results were reported in the three studies adopting the strategy of L1 use. First, Shabaka-Fernández (2023) found evidence of improvements, but the between-group difference failed to reach statistical significance. In the pre-posttest experiment, two Year 4 secondary school classes were compared. The L1 group (n=30) received Spanish (their L1) alongside English

input when modal perfect was introduced. Spanish was used when students translated examples of the given structure; English was used for the remainder of the lesson. The L2 group (n=30) received English-only instruction throughout. No information about the pupils' English proficiency level was available. After six 55-minute sessions, an unannounced posttest took place. The L1 group showed improvements, with their mean in the two tests rising by 35.5% (*SD*=25.9). However, no significant difference in scores was obtained between groups (*p*=.09). The complexity of modal perfect and the potential interference from modal verb as a recently introduced item were suggested as explanations. It must, however, be remarked that the L1 group was considered more proficient and disciplined by their schoolteacher, but a visual inspection of the raw data provided in the pre-experiment grammar test suggested otherwise (L1 group M=33.6%, SD=27.8; L2 group M=39.9%, SD=31.2). The subsequent questionnaire revealed that the L1 group were twice as likely to claim that they had developed a higher degree of '(complete) understanding' of modal perfect. Since the definition of 'understanding' was not explained, this measure may be subject to measurement bias.

Mehrseresht et al (2015) also examined L1 use, although the frequency or amount of the L1 was unspecified. 50 third-grade girls in a junior high school were randomly assigned to either a Persian (L1) or English (L2) group, each with 25 participants. Three target structures were taught, but these were not explicitly identified. After five weeks, learning was assessed via a 30-item multiple-choice test, with 10 questions per structure; the test served as both the pre- and posttest. The authors suggested that using Persian as MOI improved grammar achievements. However, statistical results of between-group comparison yielded a p-value over p<.05 (*F*=3.768, *p*=.058), which does not the threshold of statistical significance in Applied Linguistics (Dörnyei, 2007). Therefore, the claim that L1 instruction being more effective than English-only instruction was not supported by their own data.

The final study examining L1 use is Khonamri et al (2022). As noted in the 'Translanguaging' section, the exclusive use of Turkman, learners' mother tongue, was not as effective as translanguaging, which combined Turkman, Persian and English (p=.005). On the L1- and L2-only conditions, the follow-up questionnaire suggested that Turkman use promoted engagement, while students' lack of proficiency in Persian (L2) may have accounted for the lower performance in the L2-only condition. Although the mean scores of Turkman-only instruction was higher, the difference between L1- and L2-only conditions did not reach statistical significance (p=.06). This suggested factors beyond L2 proficiency may have

influenced the outcomes. Meanwhile, the within-subjects design and the absence of controls for potential confounders such as practice effects and variation in item difficulty (see Translanguaging section) undermined the power of the researchers' conclusions.

Taken together, the confidence of evidence for these studies labelled as 'L1 use' was very low due to serious risk of bias in two out of the three publications (Mehrseresht et al, 2015; Khonamri et al, 2022). Mehrseresht et al (2015) provided insufficient intervention details and may have misinterpreted statistical results; Khonamri et al (2022) showed unclear alignment between test content and intervention phases. Although Shabaka-Fernández (2023) presented fewer limitations, it was underpowered and did not offset the overall risk of bias.

4.4.2 Metalinguistic awareness

In addition to mapping form and meaning, qualitative data shed light on the metalinguistic awareness fostered by different L1 strategies. Three eligible studies addressed this aspect of grammatical development.

As the only study relying exclusively on qualitative evidence, Gallagher and Colohan (2017) provided insights into learners' perceived awareness of the forms of adverbials of frequency and place. In the intervention activity called 'Twisted Dictation', students first wrote down descriptors about rainforests and deserts presented to them aurally (dictation) and then sorted them under the corresponding label. For instance, the statement 'It rarely rains' would be placed under 'deserts'. The experimental group undertook an additional task of translating the English stimuli into Italian, their L1 – thus the 'twisted' element. Learners' reaction to the task was explored in a questionnaire the next day. Students in the experiment group (n=30) were reported to be more aware of the English-Italian word order differences and were likely to be able to articulate the contrasts. Meanwhile, the control group focused primarily on the pronunciation of content-related vocabulary; those who noticed the structural difference did not elucidate the cross-linguistic contrast. The authors concluded that translanguaging, operationalised in the form of codeswitching and translation, facilitated the noticing of the target structure's form.

Fan (2022) reported a secondary outcome concerning learners' metalinguistic awareness, specifically their recognition of the absence of verb agreement in Chinese when producing the target 'there be' structure. In the translanguaging group, participants were permitted to codemix

by using L1 equivalents for unknown words. This often led to the omission of the plural suffix -s (e.g. 'ladybug**s**'), which in turn caused errors in subject-verb agreement (e.g. 'there **is** some **七星**' [ladybug**[s]**, with the plural marker omitted despite a plural meaning indicated by 'some'). One interviewee noted that this reduced opportunities to practise the target structure during the activity. Such remark reflected a heighted awareness of non-target structures (i.e. subject-verb agreement) in English. Fan's observation was corroborated by interview data in Khonamri et al (2022), where learners reported noticing a greater extent of grammatical similarities between English and Turkman (L1) than between English and Persian (L2).

Although three studies reported findings on metalinguistic awareness, Gallagher and Colohan (2017) was the only one to focus on it directly. Nevertheless, its lack of transparency in data analysis resulted in a high RoB. Meanwhile, the increase in metalinguistic awareness was a secondary outcome of Fan (2022) and Khonamir et al (2022) and was not richly described. With reference to GRADE-CERQual (Lewin et al, 2018), the tool for qualitative studies, the overall confidence in these findings was therefore rated low due to serious methodological limitations and inadequacy of the data.

4.4.3 Non-linguistic outcomes

Overall, participants across the reviewed studies expressed a generally positive attitude towards integrating their L1 in grammar instruction. Reported benefits included enhanced participation, a more comfortable and less anxiety-inducing learning environment (Khonamri et al, 2022), and enjoyment (Fan, 2022). The translanguaging approach was found to promote the development of a bi/multilingual learner identity (Gallagher & Colohan, 2017). As most of the attitudinal data did not explicitly address grammar learning, they were not analysed in detail here. However, students' perceived rate of grammar learning is considered to be of relevance and are discussed below. As the discussion will make clear, findings in this area remain inconclusive.

Perceived faster rate of learning

Fan (2022), He (2017), and Khonamri et al (2022) reported that participants believed bi/multilingual grammar instruction accelerated their acquisition of target structures. Worth discussing is the anecdotal evidence from the Translanguaging group in Fan (2022) on the spot-the-difference task: learners revealed that the more flexible language choice for unknown

vocabulary enabled them to focus on processing and producing the 'there be' structure within the limited time available. By contrast, the English-only group reportedly prioritised lexical retrieval to avoid communication breakdowns, which appeared to reduce fluency. Translanguaging students also described instances of noticing, self-correction, and correcting their peers. This perceived improvement in rate of grammar learning was further supported by the frequency analysis, which revealed that the group produced the target structure more frequently and with greater accuracy compared to their English-only counterparts.

Similar observations were made in other studies. Learners in Khonamri et al (2022) reported drawing on Turkman (their L1) to support comprehension due to its higher level of syntactic similarity with English. They also perceived the combined use of Turkman and Persian (L2) in the translanguaging condition as beneficial, as the distinct grammatical properties of each language were leveraged to help address difficulties with learning English grammar. Likewise, Shabaka-Fernández (2023) and He (2017) both found that students preferred integrating the L1 when addressing challenging features of the target grammar. In addition, participants in He (2017) considered teacher codeswitching enhanced explanations, which allowed them to grasp a large volume of information within limited time, thus enhancing the efficiency of form-meaning mapping.

Perceived slower rate of learning

Conversely, the incorporation of the L1 led to concerns over the perceived rate of learning. According to interviewees in He (2017), teachers' reliance on codeswitching could be distracting: they found themselves relying on Chinese to process English grammar, which to them was inefficient and even counterproductive. They also found the excessive use of Chinese to learn attributive clause potentially tedious.

These contradictory findings were based on three studies that examined the perceived impact of grammar learning under L1-facilitated instruction. Given the limited number of studies and adequacy of data available, the overall confidence of evidence for this outcome was rated as moderate.

4.5 Chapter Summary

The present chapter presented findings from seven studies featuring learners' own language in grammar instruction. The languaging strategies included 'translanguaging', 'pedagogical

translanguaging', 'codeswitching' and 'use of L1', each applied to varying degrees and for different purposes. Most studies reported teacher-led approaches except for Fan (2022), which incorporated student-directed approaches. Gains in learning were generally observed, but the achievement gap between experimental and English-only groups did not always reach significance. That said, none of the studies reported negative impacts of bi/multilingual grammar instruction, despite the reduced target language input.

Chapter 5 Discussion

This review set out to address two questions: 1) What L1-mediated strategies have been used for explicit grammar teaching in primary and secondary English language classrooms? and 2) What are the effects of these approaches on English learners' grammatical development? A systematic search identified seven studies of varying quality. The multilingual pedagogies examined included translanguaging, pedagogical translanguaging, code-switching and use of the L1. The findings revealed mixed outcomes across these approaches.

Chapter 5 is organised around the two review questions, with Section 5.1 highlighting the inconsistencies surrounding the strategies and their implementations, and Section 5.2 presenting the synthesised effects of the measures.

5.1 RQ 1: L1-mediated strategies

A range of explicit grammar instruction mediated by learners' mother tongue was reported in studies published between 2015 and 2023, identified through a systematic search spanning 2009-2025. The strategies examined included 'translanguaging', 'pedagogical translanguaging', 'codeswitching' and 'use of L1'. However, there is considerable inconsistency in the conceptualisation and operationalisation of these pedagogies. This resembles the 'jingle-jangle fallacy' in L2 Motivation Self System, a different strand of research in Applied Linguistics. The fallacy is described as follows:

> *Jingle fallacies at the construct level involve "assumptions of similarity. Because they bear the same or a nearly identical name, constructs are conceptually distinct can be treated as equivalent. Jangle fallacies, on the other hand, involve assumptions of dissimilarity. Identical constructs are differently named (Marsh et al, 2019; Ponnock et al, 2020)"* (Henry & Liu, 2024, p. 738)

Studies in this review sometimes apply the same term (e.g. translanguaging) to describe different pedagogical approaches (*jingle*), or different terms (e.g. use of L1, codeswitching and translanguaging) for similar practices (*jangle*). The specific terminological ambiguity will be elucidated in the following sections, beginning with the jangle fallacy as it concerns primarily the conceptual framework of the labels.

5.1.1 Jangle fallacy

An assumption of dissimilarity is identified in the labelling of L1-mediated classroom practices. Four out of the seven included studies explicitly described their pedagogical approach as 'translanguaging', 'pedagogical translanguaging', or 'code-switching/translanguaging'; two referred to their approach as 'use of L1' and the remaining as 'codeswitching'. Based on these labels, the data reflect the growing popularity of translanguaging in language education (Goodman & Tastanbek, 2021; Treffers-Daller, 2025). Meanwhile, other mother-tongue-based pedagogies, such as codeswitching and L1-use, continue to be treated as separate approaches. This variation in labelling illustrates a jangle fallacy, where conceptually similar practices – leveraging learners' mother tongue to support language learning – are categorised under different terms. There seems to be no consensus in current literature on whether 'L1 use' and codeswitching are parallel constructs to translanguaging, or whether translanguaging functions as an overarching umbrella term that encompasses these practices (Goodman & Tastanbek, 2021; Gutierrez, 2021).

The jangle effect, where different terms are adopted for similar concepts and practices, is evident among the translanguaging-based studies reviewed. Cross-linguistic comparisons were featured in studies adopting both 'translanguaging' and the PTL label. The two PTL cases (Hopp & Thoma, 2021; Fan, 2022) involved teacher-planned language comparisons, drawing on students' L1 and L2 (and heritage languages in Hopp & Thoma, 2021), and thus their full linguistic repertoire (Cenoz, 2017; Cenoz & Gorter, 2020). In contrast, Gallagher and Colohan (2017), which also documented planned language comparisons, did not subscribe to the PTL label. Instead, they referred to the structured L1 use during 'Twisted Dictation' simply as 'translanguaging' achieved through codeswitching. Specifically, they classified codeswitching as a practice informed by 'translanguaging' (p. 489), implying a hierarchical relationship, while simultaneously calling their approach 'codeswitching/translanguaging'. These variations in labelling for conceptually similar classroom practices reflect inconsistencies in the conceptual framing of translanguaging-based studies rather than their implementation. Such misalignment reduces comparability across translanguaging-based studies, as the jangle effect implies possible conceptual differences even though the practices themselves are similar.

Translation illustrates both jangle and jingle effects (the latter discussed in the next section). The jangle effect arises as the same practice of translation, i.e. converting English input into the L1, is applied across several studies but labelled differently. For example, learners translating a target item into their L1 was described as 'use of the L1' in Shabaka-Fernández

(2023; learners also translated grammar rules into Spanish), as PTL in Hopp and Thoma (2021), and 'codeswitching/translanguaging' in Gallagher and Colohan (2017). Adopting different labels for the same activity creates the illusion of conceptual differences despite similarities in classroom enactment.

A closer inspection highlights commonalities that extend to their conceptual underpinnings. First, translation across studies often preceded language comparison activities, aligning with Guy Cook's (2010) definition of pedagogical translation, or TILT, which treats translation as a learning tool to promote language awareness through L1-L2 contrasts. Second, while all three studies acknowledged the pedagogical value of translation, none explicitly labelled their approach as '(pedagogical) translation'. This avoidance likely reflects the term's association with the Grammar Translation Method, which is widely considered outdated in the era of CLT due to its emphasis on memorisation and treated knowledge of L1 equivalents as evidence of acquisition (G. Cook, 2010; Ellis & Shintani, 2014a; 2014b).

5.1.2 Jingle fallacy

Grouping distinct concepts and practices under the same label gives an illusion of similarity, while the nuances in the theoretical underpinnings or implementation can be overlooked. The jingle fallacy pertains to the grammar teaching strategies identified in this review.

A comparison of the coalesced studies reveals that one label is applied to distinct practices, creating a false sense of similarity. The most prominent example is the term 'use of the L1'. It encompassed teacher codeswitching to explain grammar rules and present examples alongside student translation in Shabaka-Fernández (2023), a certain degree of L1-based instruction in Mehrseresht et al (2015), and the exclusive adoption of L1 as the MOI in Khonamri et al (2022), who treated the condition as a construct separate from translanguaging. The varied degree and purpose of L1 use by students and teachers captured under one single label exemplifies a jingle fallacy. The assumption of similarity may obscure the distinct effects of approaches labelled as 'L1 use'.

Using the term 'translation' across studies risks assuming similarity, while closer examination of classroom implementation reveals nuanced differences in interpretation. Rather than viewing traditional one-to-one equivalence as the end goal (G. Cook, 2010; Gutierrez, 2021), translation in the reviewed studies was of a different nature – foregrounding learner-led, L1-mediated practices as a means to grammar learning (Gutierrez, 2021). A notable example is

Gallagher and Colohan (2017). They recorded students writing down aural English (L2) prompts in Italian (their L1). This multimodal form of translation challenges the assumption that translation is 1) the target of learning, and 2) is restricted to a single modality (written or oral). It taps into Gutierrez (2021)ṡ proposed framework of translanguaging-based pedagogical translation, which can be transsemiotic and multimodal (e.g. subtitling). While this instance is underpowered by measurement and reporting bias, it illustrates a subtle shift in conceptualising translation within translanguaging in the classroom. Using the same term across studies may risk obscuring the differences between practices that treat translation for equivalence as the target, and those that leverage it as a scaffolding tool, exemplifying the jingle effect.

Narrowing the scope now to studies adopting a translanguaging framework, a similar tension can be observed between its conceptualisation and operationalisation, illustrating the jingle fallacy. Huang and Chalmers (2023) reported difficulties in establishing an unequivocal definition of translanguaging, with interpretations ranging from an ideological stance of softening language boundaries (García, 2009; Blackledge & Creese, 2010) to its pedagogical application of multilingual classrooms. The present review offers further insight into the ambiguities. Translanguaging in the four studies reviewed was primarily implemented at the lesson or discourse level, through structured activities such as teacher-led L1-L2 grammar comparisons, eliciting examples in students' heritage languages, or code-switching for metalinguistic explanations (Gallagher & Colohan, 2017; Hopp & Thoma, 2021; Fan, 2022). None of these translanguaging-focused studies explicitly addressed the ideological aim of softening language boundaries or explored curriculum-level language integration. Similarly, lessons delivered exclusively in the L1 or L2, despite drawing on learners' knowledge in other languages, were not classified as translanguaging by Khonamri et al (2022); only lessons combining L1 (Turkman), L2 (Persian), and English were considered translanguaging. These patterns emerged from this synthesis substantiated the observation that the distinction in the conceptualisation of translanguaging is not always perceived as empirically important (Huang & Chalmers, 2023, p.16; Günther-van der Meij et al, 2020; Goodman & Tastanbek, 2021). The continued ambiguity contributes to a jingle effect – applying the same label across conceptually distinct practices creates an impression of equivalence, which may mask important distinctions, such as what constitutes as translanguaging.

5.1.3 Function and extent of L1 use

Putting side the jingle-jangle fallacy surrounding prescribed labels and their implementations, learners' own language consistently functioned as a scaffolding tool across studies. This shows that researchers and practitioners alike recognise the pedagogical value of learners' L1 as a cognitive tool in supporting form-meaning mapping in grammar (Kim & Weng, 2022; Ellis & Shintani, 2014b). The reviewed studies align with existing literature based on teacher self-reports and classroom discourse analysis, which found that the L1 is favoured when explaining complex grammatical structures, providing metalinguistic information, and facilitating connections between learners' L1 and L2 (e.g. Copland & Neokleous, 2011; Izquierdo et al, 2016; Inciman & Akay, 2022; Campa & Nassaji, 2009).

This shared recognition extended into practice, as manifested in a unidirectional pattern of language use. The stronger language, usually the L1, was consistently leveraged to support the development of the weaker (English). Such a pattern aligns with the function of the L1 in promoting L2 development across non-translanguaging approaches, including codeswitching (Macaro, 2005, Üstünel, 2016; Goodman & Tastanbek, 2021), pedagogical translation (G. Cook, 2010), and general 'use of L1' strategies. However, the same unidirectionality was also evident in translanguaging-focused studies. Understandably, with English grammar development being the goal, only learning gains on L2 were measured. Nevertheless, the translanguaging studies in this review did not report an intention to strengthen both languages or to retain bilingual proficiency, nor was there any reciprocative effect of L2 supporting the L1 reported, which are hallmark features of translanguaging (Duarte, 2020; Lewis et al, 2012b). This contrasts with the ideological goals of translanguaging, which extend beyond scaffolding and acknowledging learners' language identities to enhancing content and language knowledge by engaging the entirety of learners' bi/multilingual repertoire (William, 2002; Lewis et al, 2012b; Duarte, 2020). With translanguaging in this study demonstrating the same overlapping implementation and rationale as its non-translanguaging counterparts, its status as a distinct construct or even an overarching term, at least in the context of grammar instruction, appears questionable (Goodman & Tastanbek, 2021).

Despite ambiguities in their conceptualisation, the role of L1 in scaffolding English grammar learning appears consistent. On the other hand, there is considerable variability in how its use is reported. The frequency or proportion of L1 during interventions was not always clearly documented: Hopp and Thoma (2021) controlled L1 use through a teacher script, whereas in

other studies, language switching appeared arbitrary, often depending on how often grammar rules were explained and how many L2 examples were translated (e.g. Shabaka-Fernández, 2023; He, 2017). Some (e.g. Mehrseresht et al, 2015) omitted such information entirely. Furthermore, few studies employed thematic or discourse analysis to examine actual classroom language use. A notable exception was He (2017), who presented excerpts of teacher-student interaction under monolingual and codeswitching conditions, although no formal thematic analysis was conducted. Such incongruities in empirical reporting were found across translanguaging-focused and broader L1-facilitated approaches. On the one hand, the lack of transparency hinders understanding of the relationship between L1 use and L2 outcomes (Üstünel, 2016). On the other hand, this observation aligns with Treffer-Daller's (2025) critique that much research (re)branded as 'translanguaging' has neither sufficiently adapted its analytical frameworks, nor provides robust empirical evidence regarding classroom implementation. This critique seems applicable to studies not explicitly framed around translanguaging (e.g. Mehrseresht et al, 2015).

The first review question asked what L1-mediated strategies have been introduced for explicit grammar instruction. While multiple attempts have been made to distinguish between translanguaging and other related L1-mediated strategies (e.g. Lewis et al, 2012b; Goodman & Tastanbek, 2021; García & Li, 2014), the synthesis of empirical studies in this review highlights areas of overlap between translanguaging, PTL, codeswitching and the use of L1. It also underscores a lack of clear distinction in how frontline teachers employ these strategies (Günther-van der Meij et al, 2020).

5.2 RQ2: Effects on grammar development

The second question, which concerned the effects of L1-delivered explicit grammar instruction, yielded mixed findings across the included studies, with the power of conclusions constrained by variability in study quality.

5.2.1 Linguistic outcomes

Among studies measuring linguistic outcomes using quantitative methods, only Hopp and Thoma (2019) found significant effects of PTL compared to English-only instruction. Nonetheless, the tentative quality of the studies and limited evidence of between-group differences do not negate the potential benefits of L1 mediation for grammatical development.

Indeed, none of the studies reported the L1 being a hindrance. Rather, the findings confirmed its facilitative role: the L1 appears to aid, rather than interfere with, the acquisition of complex domains like grammar (Cenoz & Gorter, 2020). The learning gains found in this review is also in line with Cummin's (2005) Linguistic Interdependence Hypothesis, which posits that proficiency in the L1 can transfer to L2 learning. Moreover, even when English input was reduced to accommodate L1-facilitated instruction, linguistic growth was evident in most studies (e.g. He, 2017; Shabaka-Fernández, 2023; Mehrseresht et al, 2015). This aligns with the findings of most studies on L1 mediating grammar instruction (e.g. Swain & Lapkin, 2000; McManus & Marsden, 2019a; 2019b), supporting the claim that bi/multilingual grammar instruction does not compromise English proficiency (Macaro, 2005; Swain & Lapkin, 2000). These findings address concerns from advocates of language separation that L1 usage might reduce exposure to English and negatively affect L2 development (Macaro, 2005; Üstünel, 2016).

This review found evidence that L1-mediated strategies, like cross-linguistic comparisons, were effective for drawing students' attention to form even though no explanations on grammar rules were provided in Hopp and Thoma (2021) and Gallagher and Colohan (2017). FonFS instruction, operationalised as language comparison activities, optimises noticing (Schmidt, 1990) – Even under an inductive approach (Ellis, 2006), the process of highlighting the formal and structural differences between languages allows learners to develop explicit knowledge for form-meaning mapping, which can then contribute to long-term acquisition (Ellis & Shintani, 2014a). The results of this review, which is based on beginning-to-intermediate level learners, complemented the findings of Scott and Fuente (2008), who observed that university-level intermediate learners of Spanish and French were able to better infer the functions of relative pronouns with the support of English, their L1.

A potential factor contributing to the reported grammatical development could be the higher level of engagement noted in the studies reviewed – many participants contended that the process of grammar instruction was more enjoyable and easier to understand because their L1 was involved (e.g. Shabaka-Fernández, 2023; Fan, 2022; He, 2017). Macaro (2005) suggested that bilingual instruction, particularly during early stages of L2 acquisition, helps reduce cognitive load during input processing. This reduction frees up working memory, allowing beginner learners to allocate more resources to form-meaning mapping (Baddeley, 2006; Macaro, 2005). However, because the results across studies were tentative, they only provided

partial support for the claim that L1-mediated instruction may be especially beneficial for younger learners, whose L1 proficiency is still developing, or for beginner learners (Inciman & Akay, 2022).

Nevertheless, the lack of transferability across language outcome measures warrants discussion. He (2017) reported that codeswitching improved learners' comprehension of attributive clauses, but this effect did not extend to production. Similarly, in Hopp & Thoma (2021), the enhanced awareness of word order and passive voice after instruction was not transferred to the non-target structures. In Fan (2022), the more accurate 'there be' output during peer interaction was observed, but this improvement did not generalise across modes of assessment – the pre-posttest gains did not reach statistical significance in the GJT test, which assessed automaticity of grammar processing.

These results contrast with McManus and Marsden (2019a; 2019b). Their advanced French-learning university students receiving bilingual input and practice outperformed their peers in both comprehension and production of habitual and continuous aspects of French imparfait (2019a); they also demonstrated emergent automaticity, as reflected in their reaction times (2019b). The restricted generalisability across items and skills, on the other hand, is in line with Hopp et al (2022), where gains from instruction on specific items were be reflected in general receptive grammar tests. These patterns suggest a possible interaction between learners' proficiency, age, and the effects of explicit grammar teaching using the L1, which necessitates further research.

5.2.2 Metalinguistic awareness

Although metalinguistic awareness was not directly measured, the data suggest that using the L1 may contribute to its development. Beyond noticing the structural differences across languages (e.g. Gallagher & Colohan, 2017; Khonamri et al, 2022), learners demonstrated such awareness through their ability to make corrections, as reported in Fan (2022). Allowing the participants (aged 12-13) to code-mix for unknown vocabulary reduced the need for lexical retrieval. The lighter cognitive load enabled students to monitor their own output, which explains the instances of self- and peer corrections (Baddeley, 2006). Similar outcomes were observed in Bell et al (2020), who analysed error correction in a cohort 11-year-olds in an ESL context. While learners in the ESL study were explicitly instructed to make corrections orally and justify themselves, the EFL group in Fan (2022) were engaged in an information gap

activity and error correction on the target took place without being prompted. While metalinguistic awareness is generally associated with linguistic gains (Ellis & Shintani, 2014a), such enhanced awareness in Fan (2022) did not translate into stronger results on other quantitative measures in the study.

5.2.3 Non-linguistic outcomes - Perceived rate of learning

Students' self-reported learning rate varied considerably, according to the qualitative data synthesised. Campa and Nassaji (2010) suggested that a perceived acceleration could be due to participants' prior experiences with L1-facilitated instruction in other previous FL experiences, particularly at the beginning stage where proficiency level was too low for an immersion approach. Familiarity with L1 support during FL instruction allowed learners to focus on input rather than adapting to an all-English environment, enabling them to process more information (Campa & Nassaji, 2010; Macaro, 2005). Similarly, classroom interactions involving learners' home language, as found in Fan (2022), were more continuous and collaborative, which potentially promoted engagement and reduced frustration when learners knew a concept but could not express it in English (e.g. Scott & Fuenta, 2008; Swain & Lapkin, 2000; Busse et al, 2020). However, as none of the studies reported the usual MOI or participants' prior FL experience, Campa and Nassaji's (2010) explanation cannot be directly verified.

On the other hand, some learners in Fan (2022) felt that reliance on the L1 slowed their L2 development, viewing it as an interference in developing 'native-like habits' (e.g. processing information in the L2; Ellis & Shintani, 2014b, p. 226). It, albeit being a secondary outcome of Fan (2022), reflects a deep-seated preference for a L2-only environment, as also shown in Scott and Fuente (2008), who upheld such a belief despite acknowledging the benefits of L1 use. Many still regard immersion as the sine qua non of achieving native-like competence, though the L2-only approach and native-like competence are being increasingly contested in the SLA field (Üstünel, 2016; Ellis & Shintani, 2014b).

5.3 Contributions of the included studies

5.3.1 Addressing knowledge gaps

Despite the small sample size (n=7), the empirical data presented in this review sheds light on grammar instruction, a relatively less-researched yet crucial area of language development (Ellis & Shintani, 2014a; Pawlak, 2022). It also explores Cenoz and Gorter's (2020) claim that

the effects of translanguaging are more apparent on other language domains than vocabulary. By featuring speakers of languages that are typologically different from English (e.g. Chinese, Turkman, and Persian), this review also addresses under-studied demographic groups in form-focused instruction research (Sok et al, 2019). This helps broaden understanding of the facilitative role of the mother tongue in English teaching, an area dominated by linguistically similar languages, such as French, German, and Spanish (Prilutskaya, 2021; Kim & Weng, 2022).

Featuring empirical studies conducted in authentic primary and secondary schools, this review adds ecological validity to the interpretation and implication of the results. It offers insights for language teachers seeking to integrate additional language resources into English grammar lessons and the wider curriculum; the findings are also relevant for education policymakers fine-tuning MOI-policies at the institutional level (Blackledge & Creese, 2010).

5.3.2 Limited generalisability

Despite attempts to explore effects of the L1 in harnessing FFI, the findings of this synthesis may have limited generalisability. One key reason is the inconsistent reporting of learners' English proficiency levels. Competence was measured and documented in multiple ways, from using recognised frameworks such as CEFR criteria and the Oxford Placement Test, to national-level assessments, to the length of EFL instruction; some studies provided no such information at all. The variability, combined with the interaction between grammar-item complexity and pupils' language competence, made it difficult to interpret the effects of language choice on explicit grammar teaching, let alone to compare results across studies.

Having data only from basic to intermediate level EFL learners limits the scope of understanding. Translanguaging is believed to be more suitable for learners with some established L2 proficiency (Blackledge & Creese, 2010). Without any advanced learners, it is difficult to verify the effect of translanguaging as a developmental tool on grammar development across proficiency levels (Baker, 2011; Kim & Weng, 2022). Beyond translanguaging, it is also hard to examine Cummins' (1979) Threshold Hypothesis, which posits that certain cognitive benefits of bilingualism, such as metalinguistic awareness, emerge only after achieving minimum proficiency in both languages. Similarly, Macaro's (2005) suggestion that codeswitching is temporary and unnecessary for advanced learners cannot be

tested. Lab evidence from proficient learners in McManus and Marsden (2019a; 2019b) may not generalise to less proficient learners.

The combined results lack generalisability, as most interventions took place in linguistically homogeneous EFL classrooms. Most participants shared the same L1 (and L2) – even minority-language speakers had the same dominant language as their monolingual peers (Hopp & Thoma, 2021), or their numbers were too small (n=2 out of 60) to be considered (Shabaka-Fernández, 2023). While Ellis (2006) suggested that cross-linguistic comparison is difficult to implement in classrooms with varied L1s, Abourehab and Azaz (2020) demonstrated English (learners' L1) and dialects could be combined in grammar instruction in a linguistically diverse classroom. The near-uniform language backgrounds in most studies limits both the examination of Ellis' (2006) claim and understanding of how varied linguistic resources can be utilised. Additionally, except for those in Khonamri et al (2022) and the minority group in Hopp and Thoma (2021), almost all learners were acquiring English as their L2. Their learning trajectories may not be applicable to L3 or L4 English classrooms.

The generalisability of the findings is further undermined by variability in study measures. Most experiments only assessed receptive grammar knowledge and learners' communicative use of the structures taught was not always evaluated, while some (e.g. Khonamri et al, 2022) combined receptive and productive outcomes in their analysis. Additionally, assessment methods ranged from textbook exercises to researcher-designed multiple choice, cloze passages, sentence rewriting, grammaticality judgement tests). These variabilities suggest that claims that L1-mediated strategies enhance grammar development warrants a more cautious interpretation.

Finally, no studies included longitudinal data: interventions (ranging from 24 minutes to five weeks) were followed only by immediate posttests, with no delayed measures to evaluate retention. Focusing solely on short-term gains precludes a more thorough understanding of whether and how immediate outcomes translate into implicit, durable grammar knowledge (Ellis & Shintani, 2014a; McManus & Marsden, 2019b). Given the iterative, non-linear nature of grammar learning (Long, 1991), it fails to address Hopp et al (2022)'s question on whether specific gains generalise to overall grammar proficiency and the time needed.

5.4 Chapter summary

This chapter explored the types of L1-assisted strategies harnessed to support grammar instruction. The synthesis of the seven included studies highlights the jingle-jangle fallacy in the conceptualisation and implementation of the relevant pedagogical approaches. The observed inconsistencies, along with the mixed outcome on both linguistic performance and perceived rate of learning, underscores a limited generalisability of the findings to other contexts, setting the stage or further research.

Chapter 6 Conclusion

This concluding chapter summarises key findings of the review, highlighting patterns and inconsistencies in L1-mediated strategies for grammar instruction. It then discusses the limitations of the review and the pedagogical and research implications of the findings.

6.1 Summary of findings

Against widespread monolingual bias and the marginalisation of form-focused instruction (FFI) within Communicative Language Teaching, both the use of learners' mother tongue and explicit grammar teaching remain a staple in foreign language classrooms. This study sought to address this belief-practice gap in language education. Adopting a systematic review design, it synthesised findings from seven studies to examine the various L1-mediated strategies during FFI in primary and secondary ELT classrooms and their effects.

Four major types of L1-facilitated grammar teaching strategies were identified: translanguaging, pedagogical translanguaging, codeswitching, and general L1 use, all introduced in EFL contexts. Despite being grounded in different theoretical orientations – most notably translanguaging versus simple L1 use – these approaches all recognise learners' linguistic background as a valuable cognitive tool for scaffolding English grammar development. However, the jingle-jangle effect in the labels and practices across strategies indicates that inconsistencies in the conceptualisation and implementation of translanguaging persist and extend to wider multilingual teaching approaches. Such ambiguity risks obscuring key distinctions between these approaches. Moreover, translanguaging-based and other mother-tongue-based studies alike demonstrate a reliance of the stronger language in supporting form-meaning mapping in the L2. Given its defining feature – the fluid use of learners' full linguistic repertoire to support content and language learning – the study raises the question of whether translanguaging represents a genuinely distinctive construct or merely a rebranding of established L1-facilitated practices.

Although linguistic gains varied across studies and educational levels, a consistent finding is that incorporating learners' home language into grammar instruction is not detrimental. Multilingual instruction appears to reduce cognitive load, freeing working memory capacity for processing more complex aspects of language, such as form-meaning mapping in grammar.

Studies also reported higher metalinguistic awareness, while perceived learning rate varied. Nevertheless, conclusion regarding the effects of explicit grammar teaching facilitated by the L1 remain tentative due to the short-term nature and variable quality of the studies.

6.2 Limitations of this review

A meta-analysis was not conducted despite the availability of quantitative data. One reason is the inconsistencies in the data available and their quality across studies (Petticrew & Roberts, 2006). More critically, despite some overlap in the theoretical underpinnings between translanguaging-informed and other non-translanguaging studies (García & Li, 2014; Lewis et al, 2012b; Goodman & Tastanbek, 2020; Gutierrez, 2022), synthesising them into a single meta-analysis would risk conflating these distinct pedagogical approaches, essentially comparing apples to oranges. While this limits the ability to draw a quantitative conclusion, the narrative approach aligns with the review's rationale of providing a broad overview of existing relevant approaches, not to compare them directly.

Another limitation concerns the search strategy. The reviewer recognises the plethora of terms associated with translanguaging, such as 'heteroglossia,' 'flexible/dynamic bilingualism,' 'code-meshing' and 'plurilingualism'. Likewise, the multifarious terminologies for other L1-mediated practices include 'addressing language diversity in a resource-oriented way' (e.g. Busse et al, 2020), 'language-switching' (e.g. McManus, 2021), 'L1 processing improving L2 processing' (e.g. McManus & Marsden, 2019b), and 'language comparison' (e.g. Nap et al, 2024). This list is by no means exhaustive. Furthermore, as the L1 is often embedded within consciousness-raising activities (Ellis & Shintani, 2014a) or input-enhancement/processing components (Ellis, 2006), identifying relevant records was challenging.

Such difficulty extended to grammar-focused studies: some did not specify language switching or reference grammar in the title or abstract. Studies involving Grammar Translation Method (GTM) were often excluded due to the unclear descriptions of language use; GTM studies were often compared to other instructional approaches, such as CLT (e.g. Çiftci & Ozcan, 2021) or discourse-based grammar teaching (e.g. Zhu, 2022). It was thus difficult to isolate the effects of L1-medication. Despite efforts to apply the eligibility criteria rigourously, the variability in terminology and scope made it practically impossible to capture all relevant studies. As a result, some potentially eligible records may have been inadvertently overlooked based on limitations in how the approaches were described in titles and abstracts.

6.3 Pedagogical and research implications

Although findings were mixed and the conclusions tentative due to the overall study quality, none of the studies in this synthesis indicated that L1 use hindered L2 grammar development. Contrary to the view that L2 proficiency requires maximum exposure to the target language (e.g. Krashen, 1976), this review provides evidence that reduced time on task is not counterproductive. This review addresses the belief-practice gap and lends support for English teachers who feel conflicted about their occasional use of codeswitching in response to the learners' needs (Blackledge & Creese, 2010). It also challenges the notion that L1 use is an illegitimate practice (Blackledge & Creese, 2010; Swain & Lapkin, 2000). Teacher-training programmes should therefore equip pre-service teachers with the knowledge and strategies needed to discern when and how multilingual instruction can be productively integrated into FL instruction.

However, for teacher-educators to provide such guidance, research must offer clearer answers about the conditions under which L1 use is most effective. Owing to inconsistencies in the report of quantity, frequency, function and purpose of L1 use, coupled with the small number of studies, it remains difficult to determine how much L1 should be used, for which level of L2 proficiency, and at which stage(s) of grammar instruction. Apart from highlighting grammar being an under-researched area (Ellis & Shintani, 2014a; Pawlak, 2022), this research gap also aligns with Treffers-Daller's (2025) critique of the lack of coherence in the analytical frameworks and reporting conventions within translanguaging research. Accordingly, this review calls for 1) more research on grammar and 2) a more standardised approach to measuring L2 proficiency outcomes and documenting classroom-based L1 use, be it by teachers, students or both parties. Greater empirical evidence, together with an alignment in analysis and reporting, would facilitate broader and more meaningful comparisons across studies. It also opens avenues for exploring key questions, such as whether the amount of L1 use correlates with learning outcomes (Macaro, 2005; Campa & Nassaji, 2010) and whether the L1 is more beneficial when used to explain grammar rules, present examples or make language comparisons. Addressing these questions is a prerequisite of defining what constitutes the 'judicious' use of mother tongue (Swain & Lapkin, 2000) as a pedagogical tool and for tailoring the strategies to individual classrooms.

Overall, this review highlights the potential of L1-mediated strategies in explicit grammar instruction in English. Several inconsistences in conceptualisation and practices were identified. Despite the limited study quality and scale of the studies, the findings underscore the value of the mother tongue as a cognitive tool in FL grammar development. Alignment in measuring and reporting classroom language use, alongside further research, would enhance understanding of the pedagogical value of multilingual instruction in language education.

References

Baddeley, A. (2006). Working memory: An Overview. In *Working memory and education* (pp. 1–31). Elsevier Inc. https://doi.org/10.1016/B978-012554465-8/50003-X

Bell, P., Fortier, V., & Gauvin, I. (2020). Using L1 knowledge about language during L2 error correction: Do students make cross-linguistic connections? *Language Awareness, 29*(2), 95–113. https://doi.org/10.1080/09658416.2020.1729783

Bhatnagar, J. K. (1980). Linguistic behaviour and adjustment of immigrant children in French and English schools in Montreal. *Applied Psychology, 29*(1–2), 141–158. https://doi.org/10.1111/j.1464-0597.1980.tb00887.x

Blackledge, A., & Creese, A. (2010). Translanguaging as pedagogy in the bilingual classroom. In *Multilingualism: A critical perspective* (pp. 201–214). Bloomsbury Academic. https://doi.org/10.5040/9781474212281.ch-010

Busse, V., Cenoz, J., Dalmann, N., & Rogge, F. (2020). Addressing linguistic diversity in the language classroom in a resource-oriented way: An intervention study with primary school children. *Language Learning, 70*(2), 382–419. https://doi.org/10.1111/lang.12382

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*, 723–733.

Cenoz, J. (2017). Translanguaging in school contexts: International perspectives. *Journal of Language, Identity & Education, 16*(4), 193–198. https://doi.org/10.1080/15348458.2017.1327816

Cenoz, J., & Gorter, D. (2020). Teaching English through pedagogical translanguaging. *World Englishes, 39*(2), 300–311. https://doi.org/10.1111/weng.12462

Chalmers, H. W. (2019). *Leveraging the L1: The role of EAL learners' first language in their acquisition of English vocabulary* (Doctoral dissertation, Oxford Brookes University). Oxford Brookes University.

Chan, B. H.-S. (2022). Translanguaging or code-switching? Reassessing mixing of English in Hong Kong Cantonese. *Chinese Language and Discourse, 13*(2), 167–196. https://doi.org/10.1075/cld.20003.cha

Chuang, H.-K., Joshi, R. M., & Dixon, L. Q. (2012). Cross-language transfer of reading ability: Evidence from Taiwanese ninth-grade adolescents. *Journal of Literacy Research, 44*(1), 97–119. https://doi.org/10.1177/1086296X11431157

Çiftci, H., & Ozcan, M. (2021). A contrastive analysis of traditional grammar translation method and communicative language teaching in teaching English grammar and vocabulary. *International Online Journal of Education and Teaching, 8*(2), 709–729.

Cochrane Effective Practice and Organisation of Care (EPOC). (2017). *Data collection form. EPOC resources for review authors*. Cochrane. https://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Resources-for-

authors2017/good_practice_data_extraction_form.doc (Retrieved on September 14, 2025).

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford University Press.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review, 57* , 402–423.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (2022). Pedagogical translanguaging: Examining the credibility of unitary versus crosslinguistic translanguaging theory. *OLBI Journal / Cahiers de l'ILOB, 12*, 33–55.

Cummins, J. (2025). Pedagogical translanguaging: Rescuing a legitimate and useful concept from the echo chamber of evidence-free assertions. *Linguistic Approaches to Bilingualism, 15*(1), 61–65. https://doi.org/10.1075/lab.24064.cum

Dickson, R., Pilkington, G., Cherry, M. G., & Boland, A. (2024). What is a systematic review? In M. G. Cherry, A. Boland, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (3rd ed., pp. 1–22). Sage Publications.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Dundar, Y., & Fleeman, N. (2024). Applying inclusion and exclusion criteria. In M. G. Cherry, A. Boland, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (3rd ed., pp. 93–108). SAGE Publications.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(1), 83–107. https://doi.org/10.2307/40264512

Ellis, R. (2016). *Focus on form: A critical review*. Language Teaching Research, 20(3), 405–428. https://doi.org/10.1177/1362168816628627

Ellis, R., & Shintani, N. (2014a). Explicit instruction and SLA. In *Exploring language pedagogy through second language acquisition research* (1st ed., pp. 83-114). Routledge. https://doi.org/10.4324/9780203796580-6

Ellis, R., & Shintani, N. (2014b). Using the L1 in the L2 classroom. In In *Exploring language pedagogy through second language acquisition research* (1st ed., pp. 225-248). Routledge. https://doi.org/10.4324/9780203796580-12

Fan, L. (2022). *Effects of pedagogical translanguaging on grammar learning for Chinese Primary School students in early foreign language learning*. (Masters thesis, University of Oxford). Oxford Research Archive.

Fleeman, N., & Dundar, Y. (2024). Data extraction: Where do I begin? In M. G. Cherry, A. Boland, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (3rd ed., pp. 109–122). SAGE Publications.

Gallagher, F., & Colohan, G. (2017). T(w)o and fro: using the L1 as a language teaching tool in the CLIL classroom. *Language Learning Journal*, *45*(4), 485–498. https://doi.org/10.1080/09571736.2014.947382

García, O. (2009). *Bilingual education in the 21st century: a global perspective.* Wiley-Blackwell.

García, O., & Li, W. (2014). *Translanguaging language, bilingualism and education*. Palgrave Macmillan.

Goodman, B., & Tastanbek, S. (2021). Making the Shift From a Codeswitching to a Translanguaging Lens in English Language Teacher Education. *TESOL Quarterly*, *55*(1), pp. 29-53. Wiley-Blackwell. https://doi.org/10.1002/tesq.571

Greenhalgh, J., & Brown, T. (2024). Quality assessment: Where do I begin? In M. G. Cherry, A. Boland, & R. Dickson (Eds.), *Doing a systematic review: A student's guide* (3rd ed., pp. 123–148). Sage Publications.

Gutiérrez, L. P. (2021). Translation in language teaching, pedagogical translation, and code-Switching: restructuring the boundaries. *Language Learning Journal*, *49*(2), 219-239. https://doi.org/10.1080/09571736.2018.1534260

He, Z. (2017). *A comparative study of teacher's code-switching and English-only explanations on the acquisition of attributive clause in senior high English grammar teaching [高中英语语法教学中教师语码转换与纯英解释对定语从句习得的对比研究]* (Master's thesis). Northwest Normal University.

Henry, A., & Liu, M. (2024). Jingle–Jangle Fallacies in L2 Motivational Self System Research: A Response to Al-Hoorie et al. (2024). *Applied Linguistics*, *45*(4), 738–746. https://doi.org/10.1093/applin/amae041

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2022). *Cochrane Handbook for Systematic Reviews of Interventions* (version 6.3). Cochrane.

Hopp, H., & Thoma, D. (2021). Effects of Plurilingual Teaching on Grammatical Development in Early Foreign-Language Learning. *The Modern Language Journal (Boulder, Colo.)*, *105*(2), 464–483. https://doi.org/10.1111/modl.12709

Hopp, H., Thoma, D., Kieseier, T., Jakisch, J., & Sturm, S. (2022). Early plurilingual foreign language teaching in mixed classrooms: Linguistic and metalinguistic predictors of foreign language development among majority-language and minority-language students in EFL classes. System (Linköping), 108, Article 102828. https://doi.org/10.1016/j.system.2022.102828

Huang, X., & Chalmers, H. (2023). Implementation and Effects of Pedagogical Translanguaging in EFL Classrooms: A Systematic Review. Languages, 8(3), 194. https://doi.org/10.3390/languages803019

Inciman, C. T., & Akay, C. (2022). Using code switching and L1 by teachers: A qualitative meta synthesis by entreq statement (1999-2021).*I-Manager's Journal on English Language Teaching*, *12*(3), 44. https://doi.org/10.26634/jelt.12.3.18617

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From Early Starters to Late Finishers? A Longitudinal Study of Early Foreign Language Learning in School. *Language Learning*, *67*(3), 631–664. https://doi.org/10.1111/lang.12242

Khonamri, F., Yaqubi, B., & Sokhandan, E. (2022). The Potential of Translanguaging Pedagogy for Promoting Iranian Bilingual Students' English Development. *MEXTESOL Journal*, *46*(4).

Kim, G. J. Y., & Weng, Z. (2022). A Systematic Review on Pedagogical Translanguaging in TESOL. *TESL-EJ (Berkeley, Calif.)*, *26*(3), 1. https://doi.org/10.55593/ej.26103a4

King's College London. (n.d.). *Grey literature - Systematic reviews - LibGuides at King's College London*. https://libguides.kcl.ac.uk/systematicreview/greylit (Retrieved on September 14, 2025)

Krashen, S. D. (1976). Formal and Informal Linguistic Environments in Language Acquisition and Language Learning. *TESOL Quarterly*, *10*(2), 157–168. https://doi.org/10.2307/3585637

Krashen, S. D. (1987). *Principles and practice in second language acquisition*. Prentice-Hall International English Language Teaching.

Leeming, P. (2011). Japanese high school students' use of L1 during pair-work. *International Journal of Applied Linguistics*, *21*(3), 360–382. https://doi.org/10.1111/j.1473-4192.2011.00284.x

Lewin, S., Bohren, M., Rashidian, A., Munthe-Kaas, H., Glenton, C., Colvin, C. J., Garside, R., Noyes, J., Booth, A., Tunçalp, Ö., Wainwright, M., Flottorp, S., Tucker, J. D., & Carlsen, B. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings—Paper 2: How to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. *Implementation Science*, *13*(Suppl 1), Article 10. https://doi.org/10.1186/s13012-017-0689-2

Lewis, G., Jones, B., & Baker, C. (2012a). Translanguaging: Origins and development from school to street and beyond. *Educational Research and Evaluation*, *18*(7), 641–654. https://doi.org/10.1080/13803611.2012.718488

Lewis, G., Jones, B., & Baker, C. (2012b). Translanguaging: Developing its conceptualisation and contextualisation. *Educational Research and Evaluation*, *18*(7), 655–670. https://doi.org/10.1080/13803611.2012.718490

Long, M. (1991). Focus on Form: A Design Feature in Language Teaching Methodology. In C. Kramsch, R. B. Ginsberg, & K. de Bot (Eds.), Foreign Language Research in Cross-Cultural Perspective (Vol. 2, pp. 39–52). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.2.07lon Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*(2), 126–141. https://doi.org/10.1093/applin/4.2.126

Macaro, E. (2005). Codeswitching in the L2 Classroom: A Communication and Learning Strategy. In E. Llurda (Ed.), Non-Native Language Teachers (pp. 63–84). Springer US. https://doi.org/10.1007/0-387-24565-0_5

Macaro, E. (2020). Systematic reviews in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 230–239). Routledge. https://doi.org/10.4324/9780367824471-20

Mehrseresht, K., Gowhary, H., & Azizifar, A. (2015). A comparative study of the use of Persian vs. English in teaching English grammar to Iranian students in junior high school. *Advances in Language and Literary Studies, 6*(4), 221–225. https://doi.org/10.7575/aiac.alls.v.6n.4p.221

Nagai, N., Ayano, S., Okada, K., & Nakanishi, T. (2015). What should be explicit in explicit grammar instruction? Language Learning in Higher Education (Berlin, Germany), 5(2), 375–396. https://doi.org/10.1515/cercles-2015-0018

Neumann, I., Brennan, S., Meerpohl, J., Davoli, M., Alonso Coello, P., Akl, E., Dahm, P., Skoetz, N., Xia, J., Brozek, J., & Schünemann, H. J. (2025, July 13). *Overview of the GRADE approach. GRADE Handbook.* https://book.gradepro.org/guideline/overview-of-the-grade-approach#conclusions

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning, 50*(3), 417–528. https://doi.org/10.1111/0023-8333.00136

Nunan, S. L. (2005). Forgiving Ourselves and Forging Ahead: Teaching Grammar in a New Millennium. *English Journal, 94*(4), 70–75. https://doi.org/info:doi/

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5, 210. 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., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Pawlak, M. (2022). Implicit versus explicit grammar learning and teaching. In E. Macaro & R. Woore (Eds.), *Debates in Second Language Education* (1st ed., Vol. 1, pp. 165–182). Routledge. https://doi.org/10.4324/9781003008361-12

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing. https://doi.org/10.1002/9780470754887

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme.* ESRC Methods Programme. https://doi.org/10.13140/2.1.1018.4643

Richards, J.C., & Schmidt, R. (2011). *S. Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Routledge. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.4324/9781315833835-294699

Samardali, M. F. S., & Ismael, A. M. H. (2017). Translation as a tool for teaching English as a second language. *Journal of Literature, Languages and Linguistics, 40*, 64-69.

Schurz, A., & Coumel, M. (2023). Grammar teaching in ELT: A cross-national comparison of teacher-reported practices. *Language Teaching Research: LTR, 27*(5), 1167–1192. https://doi.org/10.1177/1362168820964137

Shabaka-Fernández, S. (2023). The effect of teacher language use in Spanish EFL classrooms. *Language Learning Journal, 51*(2), 161–177. https://doi.org/10.1080/09571736.2021.1970794

Sok, S., Kang, E. Y., & Han, Z. (2019). Thirty-five years of ISLA on form-focused instruction: A methodological synthesis. *Language Teaching Research, 23*(4), 403–427. https://doi.org/10.1177/1362168818776673

Spada, N. (1997). Form-Focussed Instruction and Second Language Acquisition: A Review of Classroom and Laboratory Research. Language Teaching, 30(2), 73–87. https://doi.org/10.1017/S0261444800012799

Sturm, S., Jakisch, J., Kieseier, T., & Hopp, H. (2024). Plurilingual foreign-language teaching in early EFL: Exploring student perspectives and profiles. *International Journal of Multilingualism, 21*(2), 689–710. https://doi.org/10.1080/14790718.2022.2105849

Sun, Q., & Zhang, L. J. (2021). A Sociocultural Perspective on English-as-a-Foreign-Language (EFL) Teachers' Cognitions About Form-Focused Instruction. *Frontiers in Psychology, 12*, 593172. https://doi.org/10.3389/fpsyg.2021.593172

Swan, M. (2005). *Grammar*. Oxford University Press.

Swan, M. (2016). *Practical English usage (Fourth edition.*). Oxford University Press.

Treffers-Daller, J. (2025). Translanguaging: What is it besides smoke and mirrors? *Linguistic Approaches to Bilingualism, 15*(1), 1-26. https://doi.org/https://doi.org/10.1075/lab.24015.tre

Üstünel, E. (2016). Current debates in classroom codeswitching. In Üstünel, E. *EFL classroom code-switching* (pp. 83–176). Palgrave Macmillan. DOI 10.1057/978-1-137-55844-2

Vaish, V. (2019). Challenges and directions in implementing translanguaging pedagogy for low achieving students. *Classroom Discourse, 10*(3–4), 274–289. https://doi.org/10.1080/19463014.2019.1628790

Williams, C. (2002). *Ennill iaith: Astudiaeth o sefyllfa drochi yn 11–16 oed [A language gained: A study of language immersion at 11–16 years of age]*. Bangor University, School of Education.

Wright, W. E., & Baker, C. (2021). Effective Schools and Classrooms for Bilingual Students. In *Foundations of Bilingual Education and Bilingualism, 127*. Multilingual Matters.

Yeung, A. S., Marsh, H. W., & Suliman, R. (2000). Can Two Tongues Live in Harmony: Analysis of the National Education Longitudinal Study of 1988 (NELS88) Longitudinal

Data on the Maintenance of Home Language. *American Educational Research Journal, 37*(4), 1001-1026. https://doi.org/10.3102/00028312037004100l

Zawacki-Richter, O., Kerres, M., Bedenlier, S., Bond, M., & Buntins, K. (Eds.). (2020). *Systematic reviews in educational research: Methodology, perspectives and application.* Springer VS. https://doi.org/10.1007/978-3-658-27602-7

Zhu, Z. (2022). *Effects of discourse-based grammar teaching on English grammar learning of senior high school students.* (Masterś thesis, Northwest Normal University). Northwest Normal University.

Appendix 1 IDESR Registration Record

International Database of Education Systematic Reviews

Record details
IDESR ID: IDESR000216

Title
Effects of L1-mediated interventions in primary and secondary English grammar classrooms. A protocol for systematic review.

Main Contact/Corresponding Author
Chi LAM, University of Oxford. chi.lam@wolfson.ox.ac.uk
Wolfson College, Linton Road, Oxford, OX26UD

Additional Authors
Faidra Faitaki, University of Oxford. faidra.faitaki@education.ox.ac.uk

Review Question(s)
1. What L1-mediated interventions have been used for explicit grammar teaching in primary and secondary English language classrooms?
2. What are the effects of these approaches on FL learners' grammatical development?

Rationale
Cummin's Linguistic Interdependence Hypothesis (1979) posits that one's first language (L1) facilitates second language (L2) learning. Many pedagogical approaches acknowledge this hypothesis by incorporating learners' L1 in language instructions. Examples include code-switching, cross-linguistic comparison, and pedagogical translanguaging. Among these, pedagogical translanguaging has gained particular traction in both research and classroom practice since 2009.

Despite its popularity, the conceptualisation and operationalisation of pedagogical translanguaging remain ambiguous (Huang & Chalmers, 2023). Cummins (2025) critiques that pedagogical translanguaging is underpinned by the assumption that bilinguals process and store their languages within a unitary language system. In other words, instructional approaches that recognise and utilise language-specific knowledge in the bilingual cognitive system is overlooked (Cummins, 2022). The theoretical ambiguity extends to the empirical research: Treffers-Daller (2025) highlighted the lack of consistency in the operationalisation of pedagogical translanguaging research, which many L1-mediated interventional studies adopt the term 'translanguaging' without changing their analytical frameworks or presenting new findings, nor is there sufficient empirical evidence provided. The conceptual and methodological ambiguity of pedagogical translanguaging have led to the marginalisation of other L1-mediated practices in research discourse.

L1-mediated interventions have been explored at both lexical and syntactic levels. However, grammar, despite being important in language processing and learning, is relatively less studied. Existing research primarily contrasts the effectiveness of implicit and explicit instructions (Pawlak, 2022); work that examines the role of L1 is scarce although the L1 is commonly used in teaching English grammar as a second or foreign language. A preliminary search shows that no systematic synthesis has been done to synthesise the multifarious L1-mediated grammar

instructions, including but not limited to code-switching, pedagogical translation, cross-linguistic comparison, and pedagogical translanguaging.

This systematic review aims to provide a clearer overview of pedagogical approaches that leverage learners' L1 in classroom-based grammar instruction, and to evaluate their impact on English grammar learning.

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. Review of Educational Research, 49(2), 222-251. https://doi.org/10.3102/00346543049002222 (Original work published 1979)
Cummins, J. (2025). Pedagogical translanguaging: Rescuing a legitimate and useful concept from the echo chamber of evidence-free assertions. *Linguistic Approaches to Bilingualism*, *15*(1), 61–65. https://doi.org/10.1075/lab.24064.cum
Huang, X., & Chalmers, H. (2023). Implementation and Effects of Pedagogical Translanguaging in EFL Classrooms: A Systematic Review. *Languages*, *8*(3), 194. https://doi.org/10.3390/languages803019
Pawlak, M., Macaro, E., & Woore, R. (2022). Implicit versus explicit grammar learning and teaching. In Debates in Second Language Education (1st ed., Vol. 1, pp. 165–182). Routledge.https://doi.org/10.4324/9781003008361-12
Treffers-Daller, J. (2025). Translanguaging: What is it besides smoke and mirrors? Linguistic Approaches to Bilingualism, 15(1), 1-26. https://doi.org/https://doi.org/10.1075/lab.24015.tre

Inclusion Criteria

1. Reference

• Includes studies with complete reference
• Excludes studies whose reference is incomplete

2. Time

• Includes studies published between 2009 and April 30 2025
• Excludes studies published before 2009
• Rationale: Pedagogical translanguaging became popular at around 2009, so there is no relevant research before then.

3. Publication Status

• No studies will be excluded because of their publication status
• Rationale: All studies, including published and grey literature, are perused to avoid publication bias

4. Language

• No studies will be excluded because of the language in which they were written; advanced language learners/translators will be consulted when the papers are in languages other than English and Chinese
• Rationale: To ensure the studies included in the synthesis are comprehensive and representative of as many contexts as possible.

5. Study types

• Includes research articles and book chapters that report on empirical data
• Excludes articles, chapters, or books that do not report on empirical data (e.g. commentary, conceptual or theoretical papers)
• Rationale: We are interested in the effects of grammar instructions that uses L1 in actual classrooms.

6. Educational level

• Includes studies conducted in primary to secondary school foreign language classrooms.

• Excludes studies that are conducted outside of primary and secondary school contexts (e.g. academic language courses offered by tertiary institutions, EAP/ESP courses)
• Rationale: We are interested in how the L1 is used to teach grammar within primary and secondary school contexts

7. Research contexts

• Includes studies with classroom settings where the English is the second and/or foreign/additional language of the learners
• Excludes studies conducted in a laboratory
• Rationale: The focus of this study is on L1-mediated grammar instructions in classrooms

8. Participants

• Includes studies where the population is largely or exclusively typically developing language learners.
• Excludes studies featuring exclusively learners with special education needs (SEND) or developmental (language) disorders to allow a more homogeneous study comparison
• Note: For studies that include both typically developing and SEND students in mixed-group settings, only data from the former will be extracted, provided it is identifiable and available.

9. Outcomes

• Includes studies that report either a) data on teachers' implementation of multilingual pedagogies based on observed practice, classroom recording, or other forms of empirical evidence; or b) learner outcomes of the L1-mediated interventions. Studies meeting either criterion will be included.
• Excludes studies that report data on neither. Examples include impact of correctivefeedback, or those that only report stakeholders' perception of the interventions without evidence of implementation or learner outcomes

10. Intervention

• Includes approaches that describe the intervention that focuses on explicit grammar instructions using learners' L1 in a foreign language classroom (e.g. pedagogical translanguaging, translation equivalents, cross-linguistic comparison, code-switching, etc.)
• Excludes studies where learners' L1 is used for other purposes, for example, vocabulary teaching, classroom management, etc.

Information Sources

The selection of databases for this systematic review is informed by multiple sources. Research networks that publish systematic reviews, namely International Database of Education Systematic Reviews (IDESR) and Campbell Collaboration, are perused. A research librarian in the Education Library at the University of Oxford was consulted to help the researcher refine the list of databases and to develop relevant search strategies.

The following bibliographic databases will be assessed and searched electronically through the Search Oxford Libraries Online (SOLO) tool provided by the University of Oxford:

1. Education

• Education Resources Information Centre (ERIC)
• British Education Index
• Teacher Reference Centre
• Education Abstracts

2. Linguistics

• MLA International Bibliography (EBSCO)
• Linguistics Collection
• Language and Linguistics Behaviour Abstracts (LLBA)

3. Multidiscipline

• Oxford Research Archive
• SCOPUS
• ProQuest Dissertations and Theses
• Chinese National Knowledge Infrastructure (CNKI)

Search Strategy

Boolean operators 'AND' and 'OR' will be used to combine search terms. The year of publication will be limited to the period between 2009 and April 30 2025.

When searching free-text words, wildcard characters will be used. The asterisk (*) will enable the searching of a word with different spelling.

Four types of search terms are included. They are the ones related to 1) multilingual pedagogical practices, 2) the topic 'grammar teaching and learning', 3) the context and participants, and 4) learning outcomes.

The following search terms are identified to search the bibliographic databases:

1. **Multilingual pedagogies**: (Pedagogical translanguaging OR code-switch* OR code-mix* OR code-mesh* OR cross-linguistic comparison OR us* of L1 OR L1-mediat*OR leverag* L1 OR pedagogical translat* OR comparative grammar OR plurilingualpedagog* OR multilingual pedagog*)

AND

2. **Grammar teaching and learning:** (explicit grammar teach* OR grammar/grammatical learn* OR grammar teaching strateg* OR grammar/grammatical develop* OR syntactic develop* OR morphological develop* OR morphosyntactic develop* OR grammar acqui*)

AND

3. **Context and participants:** (ELT OR "English language teach*" OR ESL "English as second language" OR EFL OR "English as a foreign language" OR EAL "English as an additional language" OR TEFL OR Teaching English as a foreign language* OR TESOL OR "Teaching English to speakers of other language*" OR ESOL OR "English for speakers of other language*")

4. **Learning outcomes**: (effect* OR outcome* OR score* OR benefit* OR result* OR perception* OR attitude*)

Citation chaining, which includes both backward and forward searching, will be used as an additional search strategy. Backward searching involves reviewing the reference lists of reports that meet the inclusion criteria to identify any relevant studies that were not initially found. Forward searching entails examining the studies that have cited a particular paper to locate any related articles, if available, on the journal's website.

Data Management

The key search terms will be used to conduct database searches. The titles, abstracts, years of publication, authors, and keywords of all the studies identified will be exported as a RIS file to Rayyan, a tool for viewing, selecting and managing data in systematic reviews. The tool will help de-duplicate references and categorise abstracts based on inclusion and exclusion decisions. A Microsoft Excel file will be kept to track inclusion and exclusion decisions. The record-keeping table will include 1) reference details, 2) decisions made at the title and abstract screening stage, 3) whether full-paper was obtained, and 4) the decision at the full-text screening stage, and the reason for exclusion, if any.

After obtaining the full papers of all eligible studies, the reference management software Mendeley will be used for a full-text screening. The researcher will review the texts and record reasons for inclusion and exclusion. Mendeley will also be used in generating in-text citations and bibliography in the dissertation.

Following full-paper screening, all the papers fulfilling all inclusion criteria will be re-analysed in Mendeley for coding to address the research questions. All codes will be documented in a Microsoft Excel file. Then, the researcher will read through all the codes to develop analytical codes.

In addition, the researcher will keep a research activity log in another Microsoft Excel file for an up-to-date record of data management throughout the study. The record will be updated every week.

Selection Process

The researcher will first identify and remove any duplicate records using the results from Rayyan. Then the inclusion and exclusion criteria will be applied to screen the title and abstract of each study. The researcher will highlight the keywords and record the reason for inclusion or exclusion decisions for each study. Only the studies that cannot be excluded on the basis on the information contained in the abstract will be kept; studies that fail to meet any one of the inclusion criteria will be excluded. The full reports of all studies that cannot be excluded based on the information contained in the abstract will be sought.

After title and abstract screening, the researcher will obtain the full papers of all potentially eligible studies. These full papers will be screened according to the inclusion and exclusion criteria. The reasons for exclusion will be recorded.

To minimise researcher bias, a peer reviewer who majors in the field of education or applied linguistics will be invited to screen 20% of the titles and abstracts from the main search. A 90% agreement will be sought to ensure the inclusion and exclusion criteria have been successfully applied. If the agreement is less than 90%, the researcher and the peer reviewer discuss the reasons with each other. After that, another 20% of the sample will be screened again until a 90% agreement is reached.

If time is sufficient after the researcher assesses all the full papers obtained, the peer reviewer will be invited to repeat the above process to confirm the inclusion of the full-text citations is consistent with the inclusion and exclusion criteria.

A PRISMA flow diagram (Moher et al., 2009) will be used to present the process of inclusion.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine, 6(7), e1000097.

Data Collection Process

The process will begin with the researcher identifying the data that need to be extracted. The researcher will then develop a data extraction table in Microsoft Excel spreadsheets, and pilot it with 5 included studies, if available. The data extraction table will be revised and finalised after piloting.

Then, the researcher will read through all the included studies, extract data, and record the data in the data extraction table in Microsoft Excel spreadsheets.

Based on the data extraction table, the researcher will develop and complete data tables to report data in the systematic review.

Data Items
Two types of data will be extracted: descriptive and analytical data.

The former will comprise of the author(s)' name(s), publication year (for published studies), study location and duration, and study focus. Participant information will also be described, namely the number of participants (including any dropouts), and demographics of the participants (age, sex, ethnicity, languages). Interventions and comparators (if applicable), sources of funding, and setting, will also be extracted.

The analytical data features descriptions of the L1-mediated approaches to grammar teaching and learning. The outcomes can be quantitative (e.g. learning outcomes of the intervention) and/or qualitative (e.g. pupils' perception).

Risk of bias/trustworthiness of individual studies
The first review question concerns the existing approaches to English grammar teaching using learners' L1 in the specific timeframe. The data will mainly be an overview of the different strategies employed, so the risk of bias is not relevant.

The risk of bias will be assessed for the data addressing the second review question, which is related to the effect of the L1-mediated grammar teaching approaches, to evaluate the design, scale, attrition, fidelity, and validity of the relevant studies found. Different assessment tools will be used depending on the type of research design.
1. Randomised trials: JBI Critical appraisal tool for the assessment of risk of bias for randomised controlled trials
2. Quasi-experimental studies: JBI Critical appraisal tool for quasi-experimental studies
3. Qualitative and mixed-methods: EEF Padlock
After the assessment, the results and conclusions of the systematic review will be developed with consideration of the quality of these studies.

Data Synthesis
Narrative synthesis will be applied to synthesise data for both research questions. The textual approach uses words and text to summarise and interpret the results of a systematic review (Popay et. al., 2006). The four elements of a narrative synthesis include 1) developing a theory of how the intervention works, why and for whom, 2) developing a preliminary synthesis of findings of included studies, 3) exploring relationships in the data, and 4) assessing the robustness of the synthesis. Thematic analysis will be used as the technique for developing preliminary synthesis.

For the second review question, should there be studies with a similar research design and have student outcome measures, a meta-analysis will be conducted to synthesise numerical data. The effect sizes will be computed by calculating the Cohen's d.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S.(2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Meta-biases
To minimise meta-biases, data from multiple databases will be collected. Pre-registered protocols (e.g. PROSPERO, PRISMA) are used to ensure that relevant published studies and grey literature will be included, regardless of the outcome, language written in, on the premise that they meet all the inclusion criteria stated above. Non-English studies, mainly Chinese, will be consulted.

Confidence in cumulative evidence
The GRADE Approach will be adopted when assessing confidence in cumulative evidence. Link: https://www.gradeworkinggroup.org/

Sources of Funding
NONE

Role of Funders
NONE

Anticipated or actual start date
12 May 2025
Anticipated completion date
30 September 2025

Other language resources
N/A
Current Status
Ongoing

Details of Published Review
/
IDESR URL
https://idesr.org/article/IDESR000216

Appendix 2 Search terms used for specific databases and results returned

Database	Search syntax	Number of results
Education Resources Information Centre (ERIC), British Education Index, Teacher Reference Centre, Education Abstracts, and MLA International Bibliography (EBSCO) Accessed via EBSCOHost	((Pedagogical translanguaging OR code-switch* OR code-mix* OR code-mesh* OR cross-linguistic comparison OR us* of L1 OR L1-mediat* OR leverag* L1 OR pedagogical translat* OR comparative grammar OR plurilingual pedagog* OR multilingual pedagog*) AND) AND ((explicit grammar teach* OR grammar learn* OR grammatical learn* OR grammar teaching strateg* OR grammar develop* OR grammatical develop* OR syntactic develop* OR morphological develop* OR morphosyntactic develop* OR grammar acqui*)) AND ((ELT OR "English language teach*" OR ESL OR "English as second language" OR EFL OR "English as a foreign language" OR EAL OR "English as an additional language" OR TEFL OR "Teaching English as a foreign language*" OR TESOL OR "Teaching English to speakers of other language*" OR ESOL OR "English for speakers of other language*")) AND ((effect* OR outcome* OR score* OR benefit* OR result* OR perception* OR attitude*))	1,015*
Linguistics Collection, Language Linguistics Behaviour Abstracts (LLBA), and ProQuest Dissertations and Theses Accessed via ProQuest	(AB("pedagogical translanguaging" OR "code-switch*" OR "code-mix*" OR "code-mesh*" OR "cross-linguistic comparison" OR "use of L1" OR "L1-mediat*" OR "leverage L1" OR "pedagogical translat*" OR "comparative grammar" OR "plurilingual pedagog*" OR "multilingual pedagog*") OR TI("pedagogical translanguaging" OR "code-switch*" OR "code-mix*" OR "code-mesh*" OR "cross-linguistic comparison" OR "use of L1" OR "L1-mediat*" OR "leverage L1" OR "pedagogical translat*" OR "comparative grammar" OR "plurilingual pedagog*" OR "multilingual pedagog*")) AND (AB("explicit grammar teach*" OR "grammar learn*" OR "grammatical learn*" OR "grammar teaching strateg*" OR "grammar develop*" OR "grammatical develop*" OR "syntactic develop*" OR "morphological develop*" OR "morphosyntactic develop*" OR "grammar acqui*") OR TI("explicit grammar teach*" OR "grammar learn*" OR "grammatical learn*" OR "grammar teaching strateg*" OR "grammar develop*" OR "grammatical develop*" OR "syntactic develop*" OR "morphological develop*" OR "morphosyntactic develop*" OR "grammar acqui*")) AND (AB ("ELT" OR "English language teach*" OR "ESL" OR "English as a second language" OR "EFL" OR "English as a foreign language" OR "EAL" OR "English as an additional language" OR "TEFL" OR "Teaching English as a foreign language" OR "TESOL" OR "Teaching English to speakers of other languages" OR "ESOL" OR "English for speakers of other languages") OR TI("ELT" OR "English language	1

	teach*" OR "ESL" OR "English as a second language" OR "EFL" OR "English as a foreign language" OR "EAL" OR "English as an additional language" OR "TEFL" OR "Teaching English as a foreign language" OR "TESOL" OR "Teaching English to speakers of other languages" OR "ESOL" OR "English for speakers of other languages")) AND (AB("outcome*" OR "result*"OR "score*" OR "benefit*" OR "perception*" OR "attitude*") OR TI("outcome*" OR "result*"OR "score*" OR "benefit*" OR "perception*" OR "attitude*")) AND pd(20090101-20250430)	
SCOPUS	TITLE-ABS-KEY(("pedagogical translanguaging" OR "code-switch*" OR "code-mix*" OR "code-mesh*" OR "cross-linguistic comparison" OR "use of L1" OR "L1-mediat*" OR "leverag* L1" OR "pedagogical translat*" OR "comparative grammar" OR "plurilingual pedagog*" OR "multilingual pedagog*")) AND TITLE-ABS-KEY(("explicit grammar teach*" OR "grammar learn*" OR "grammatical learn*" OR "grammar teaching strateg*" OR "grammar develop*" OR "grammatical develop*" OR "syntactic develop*" OR "morphological develop*" OR "morphosyntactic develop*" OR "grammar acqui*")) AND TITLE-ABS-KEY(("ELT" OR "English language teach*" OR "ESL" OR "English as a second language" OR "EFL" OR "English as a foreign language" OR "EAL" OR "English as an additional language" OR "TEFL" OR "Teaching English as a foreign language" OR "TESOL" OR "Teaching English to speakers of other languages" OR "ESOL" OR "English for speakers of other languages")))	2
Oxford Research Archive	(code-mix* OR multilingual pedagog* OR plurilingual pedagog* OR comparative grammar OR pedagogical translation OR pedagogical translanguaging OR L1-mediated OR leverag* L1 OR cross-linguistic comparison OR use of L1 OR code switching OR codeswitching OR code-switching OR use of L1 grammar OR code-switching grammar OR pedagogical translanguaging) AND grammar AND English learning teaching	40
Chinese National Knowledge Infrastructure (CNKI)	(code-mix* OR multilingual pedagog* OR plurilingual pedagog* OR comparative grammar OR pedagogical translation OR pedagogical translanguaging OR L1-mediated OR leverag* L1 OR cross-linguistic comparison OR use of L1 OR code switching OR codeswitching OR code-switching OR use of L1 grammar OR code-switching grammar OR pedagogical translanguaging) AND grammar English learning teaching	165
	Total	1223

*A total of 1,102 results were retrieved from EBCOHost, and 87 duplicates were automatically removed by the platform during download of title records for initial screening.

Appendix 3 Data Extraction Table

Note: The table has been transposed for a more reader-friendly presentation.

	Item	Description	Data from study
Study identification	Title		
	Author(s)		
	DOI/URL		
	Publication year		
	Journal		
	Database	Source of information	
	Study Aims		
	Research questions		
Intervention Details	L1-mediated strategy label	Label(s) prescribed by authors	
	Implementation of strategy	How the actual strategy was actually carried out (e.g. cross-linguistic comparison, translation, teacher/student language switch)	
	Grammar focus	Specific grammar structure(s) targeted	
	Topic/Theme	Theme of lesson (e.g. crime)	
	Lesson activities	Tasks involved (e.g. dictation, information gap activity)	
	Instructional materials	e.g. textbook, handouts, whether they have been adapted	
	Duration	Length of lesson(s)/intervention, number of sessions	
	Teacher	Person conducting the intervention (e.g. researcher, regular teacher …)	
	Teachers' knowledge of learners' L1	Whether the teacher understands/can use students' L1	
Participants information	Context	Country	
	Education setting	Primary/Secondary	
	Year level		
	Participants' age	Range/average age	
	Students' L1	L1 of most, if not all, of the participants	
	Other languages spoken	Heritage language/Current dominant language/Other L1(s)/Additional languages (L2/L3)	
	Length of ELL	Years receiving formal English instructions	
	English proficiency level	Ss' level of English according to tests, CEFR, etc.	
	Sample size	Total number of participants	
	Intervention group (s)	number of Ss, group size	
	Control group(s)	number of Ss, group size	
	Grouping information	Intact class? Random assignment? Clustered randomisation?	
	Baseline control	Whether groups were equivalent at pre-test	
	Attrition and reasons	Drop-outs, missing data, etc.	
Methodology	Type of data collected	Qualitative/Quantitative/Mixed Methods	
	Study design	RCT, quasi-experimental, within-subject measures, etc.	
	Method of randomisation, if applicable	How participants were assigned to groups	

	Data collection tools/procedures	Tests, questionnaire, GJT, lesson observation, translation tests, recordings Assessment tool(s) used (e.g. tests/questionnaire/GJT/recording)	
	Validation of assessment	Information on reliability, fairness, or appropriacy of the assessment tool(s) used	
Outcomes	Outcome type	What was measured, e.g. learning gains, response rate, accuracy of response/production, frequency of target structure use, awareness	
	Outcome unit/data type	**Quantitative information**, e.g. test scores, number of correct responses, proportion of target structures **Qualitative information**, e.g. attitude towards task, reported understanding of target structures, observational notes, interview comments	
	Outcome		
	Statistical tests used	Inferential analyses applied, e.g. t-test, ANOVA, mixed ANOVA	
	Effect size, if available		
Others	Source of funding		
	Peer review status		
	Limitations	Reported study limitations	

Appendix 4 JBI Critical Appraisal Tools for Risk of Bias (RoB) Assessment

Three JBI appraisal tools were used to assess the RoB for individual studies. The table below indicates which tool was applied to each study type. For mixed-methods studies, a combination of tools was used: One for the quantitative component (RCT or quasi-experimental) and another for qualitative component.

Type of study	**JBI Critical Appraisal Tool for …**	**In Appendix**
RCT	**… Randomised Controlled Trials (2023 Version)** Barker, T. H., Stone, J. C., Sears, K., Klugar, M., Tufanaru, C., Leonardi-Bee, J., Aromataris, E., & Munn, Z. (2023). The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. *JBI Evidence Synthesis, 21*(3), 494–506. https://doi.org/10.11124/JBIES-22-00375	4a
Quasi-experimental	**… Quasi-experimental Studies (2024 Version)** Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., et al. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias in quasi-experimental studies. *JBI Evidence Synthesis, 22*(3), 378–388. https://doi.org/10.11124/JBIES-23-00217	4b
Qualitative	**… Qualitative studies** Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation. *International Journal of Evidence-Based Healthcare, 13*(3), 179–187. https://doi.org/10.1097/XEB.0000000000000062	4c
Mixed methods	1. … Randomised controlled trials **OR** … Quasi-experimental Studies **AND** 2. … Qualitative studies	4a/4b and 4c

Appendix 4a JBI critical appraisal tool for the assessment of risk of bias for randomised controlled trials

Source: Barker TH, Stone JC, Sears K, Klugar M, Tufanaru C, Leonardi-Bee J, Aromataris E, Munn Z. The revised JBI critical appraisal tool for the assessment of risk of bias for randomized controlled trials. JBI Evidence Synthesis. 2023;21(3):494-506

RoB Assessor:	**Date of Appraisal:**	**Record ID:**
Study Author:	**Study Title:**	**Study Year:**

Internal Validity		**Choice - Comments/Justification**	**Yes**	**No**	**Unclear**	**N/A**
Bias related to selection and allocation						
1	**Was true randomization used for assignment of participants to treatment groups?**		☐	☐	☐	☐
2	**Was allocation to treatment groups concealed?**		☐	☐	☐	☐
3	**Were treatment groups similar at the baseline?**		☐	☐	☐	☐
Bias related to administration of intervention/exposure						
4	**Were participants blind to treatment assignment?**		☐	☐	☐	☐
5	**Were those delivering the treatment blind to treatment assignment?**		☐	☐	☐	☐

6	**Were treatment groups treated identically other than the intervention of interest?**		☐	☐	☐	☐

Bias related to assessment, detection and measurement of the outcome

7	**Were outcome assessors blind to treatment assignment?**		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

8	**Were outcomes measured in the same way for treatment groups?**		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

9	**Were outcomes measured in a reliable way**		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

Bias related to participant retention

10	**Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed?**					
	Outcome		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

Statistical Conclusion Validity

11	**Were participants analysed in the groups to which they were randomized?**					
	Outcome 1		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐
12	**Was appropriate statistical analysis used?**					
	Outcome		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐
			Yes	**No**	**Unclear**	**N/A**
13	**Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?**		☐	☐	☐	☐

Overall appraisal: **High risk:** ☐
Moderate risk: ☐
Low risk: ☐

Comments:

Appendix 4b: JBI critical appraisal tool for the assessment of risk of bias quasi-experimental studies

Source: Barker TH, Habibi N, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias quasi-experimental studies. JBI Evid Synth. 2024;22(3):378-88.

RoB Assessor:	Date of Appraisal:	Record ID:
Study Author:	Study Title:	Study Year:

Internal Validity		Choice - Comments/Justification	Yes	No	Unclear	N/A
Bias related to temporal precedence						
1	It is clear in the study what is the "cause" and what is the "effect" (ie, there is no confusion about which variable comes first)?		☐	☐	☐	☐
Bias related to selection and allocation						
2	Was there a control group?		☐	☐	☐	☐
Bias related to confounding factors						
3	Were participants included in any comparisons similar?		☐	☐	☐	☐
Bias related to administration of intervention/exposure						
4	Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?		☐	☐	☐	☐

Bias related to assessment, detection, and measurement of the outcome

5	**Were there multiple measurements of the outcome, both pre and post the intervention/exposure?**		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐
6	**Were the outcomes of participants included in any comparisons measured in the same way?**		**Yes**	**No**	**Unclear**	**N/A**
			☐	☐	☐	☐
7	**Were outcomes measured in a reliable way**		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

Bias related to participant retention

8	**Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed?**					
	Outcome		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

Statistical Conclusion Validity

9	**Was appropriate statistical analysis used?**					
	Outcome		**Yes**	**No**	**Unclear**	**N/A**
	Click or tap here to enter text.		☐	☐	☐	☐

Overall appraisal:	**High risk:** ☐ **Moderate risk:** ☐ **Low risk:** ☐

Comments:

Appendix 4c: JBI Critical Appraisal Checklist for Qualitative Research

Source: Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc. 2015;13(3):179–187

Reviewer _____ Date_____

Author_____Year_____ Record ID _____

		Yes	No	Unclear	Not applicable
1.	Is there congruity between the stated philosophical perspective and the research methodology?	☐	☐	☐	☐
2.	Is there congruity between the research methodology and the research question or objectives?	☐	☐	☐	☐
3.	Is there congruity between the research methodology and the methods used to collect data?	☐	☐	☐	☐
4.	Is there congruity between the research methodology and the representation and analysis of data?	☐	☐	☐	☐
5.	Is there congruity between the research methodology and the interpretation of results?	☐	☐	☐	☐
6.	Is there a statement locating the researcher culturally or theoretically?	☐	☐	☐	☐
7.	Is the influence of the researcher on the research, and vice- versa, addressed?	☐	☐	☐	☐
8.	Are participants, and their voices, adequately represented?	☐	☐	☐	☐
9.	Is the research ethical according to current criteria or, for recent studies, and is there evidence of ethical approval by an appropriate body?	☐	☐	☐	☐
10.	Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data?	☐	☐	☐	☐

Overall appraisal: **High risk:** ☐ **Moderate risk:** ☐ **Low risk:**

Comments (Including reason for exclusion)

_____

_____

Appendix 5 Reference of all included studies

1. Mehrseresht, K., Gowhary, H., & Azizifar, A. (2015). A comparative study of the use of Persian vs. English in teaching English grammar to Iranian students in junior high school. Advances in Language and Literary Studies, 6(4), 221–227. https://doi.org/10.7575/aiac.alls.v.6n.4p.221
2. Gallagher, F., & Colohan, G. (2017). T(w)o and fro: using the L1 as a language teaching tool in the CLIL classroom. Language Learning Journal, 45(4), 485–498. https://doi.org/10.1080/09571736.2014.947382
3. He, Z. (2017). A comparative study of teacher's code-switching and English-only explanations on the acquisition of attributive clause in senior high English grammar teaching [**高中英语语法教学中教师语码转换与纯英解释对定语从句习得的对比研究**] (Master's thesis). Northwest Normal University.
4. Khonamri, F., Yaqubi, B., & Sokhandan, E. (2022). The Potential of Translanguaging Pedagogy for Promoting Iranian Bilingual Students' English Development. MEXTESOL Journal, 46(4).
5. Hopp, H., & Thoma, D. (2021). Effects of Plurilingual Teaching on Grammatical Development in Early Foreign-Language Learning. The Modern Language Journal (Boulder, Colo.), 105(2), 464–483. https://doi.org/10.1111/modl.12709
6. Fan, L. (2022). Effects of pedagogical translanguaging on grammar learning for Chinese Primary School students in early foreign language learning. (Master's thesis)
7. Shabaka-Fernández, S. (2023). The effect of teacher language use in Spanish EFL classrooms. Language Learning Journal, 51(2), 161–177. https://doi.org/10.1080/09571736.2021.1970794

Appendix 6 Results from chained-citation and reasons for exclusion

Title	Reason of exclusion
Leonet, O., & Saragueta, E. (2023). The case of a pedagogical translanguaging intervention in a trilingual primary school: the students' voice. International Journal of Multilingualism, 21(4), 2112–2130. https://doi.org/10.1080/14790718.2023.2224009	Unrelated to grammar
Mahan, K. R. (2020). The comprehending teacher: scaffolding in content and language integrated learning (CLIL). The Language Learning Journal, 50(1), 74–88. https://doi.org/10.1080/09571736.2019.1705879	
Karabassova, L., & San Isidro, X. (2020). Towards translanguaging in CLIL: a study on teachers' perceptions and practices in Kazakhstan. International Journal of Multilingualism, 20(2), 556–575. https://doi.org/10.1080/14790718.2020.1828426	
Pavón Vázquez, V., & Ramos Ordóñez, M. del C. (2018). Describing the use of the L1 in CLIL: an analysis of L1 communication strategies in classroom interaction. International Journal of Bilingual Education and Bilingualism, 22(1), 35–48. https://doi.org/10.1080/13670050.2018.1511681	
Fielding, R. (2015). Students' use of their plurilingual resources in Australian schools. Language and Education, 30(4), 361–377. https://doi.org/10.1080/09500782.2015.1114631	
Omidire, M. & Ayob, S. (2022). The utilisation of translanguaging for learning and teaching in multilingual primary classrooms. Multilingua, 41(1), 105-129. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1515/multi-2020-0072	
Costa, F. (2021). *Alternation between L1 (Italian) and L2 (English) in three CLIL and EMI contexts*. Cambridge Scholars Publishing.	
Gayete, G. (2025). Functions of translanguaging in primary school CLIL and EFL settings. *Journal of Language and Culture in Education, 2*(Special Issue), 76–94. https://doi.org/10.5281/zenodo.14725042	
Vaish, V. (2019). Challenges and directions in implementing translanguaging pedagogy for low achieving students. *Classroom Discourse, 10*(3–4), 274–289. https://doi.org/10.1080/19463014.2019.1628790	
Vaish, V. (2021). Linguistic distance and translanguaging: Teaching Malay-English bilingual learners. *Asia Pacific Journal of Education, 41*(4), 673–684. https://doi.org/10.1080/02188791.2021.1997708	
Aoyama, R. (2020). Exploring Japanese high school students' L1 use in translanguaging in the communicative EFL classroom. TESL-EJ, 23(4). http://www.tesl-ej.org/pdf/ej92/a12.pdf	
Yuzlu, M. Y., & Dikilitas, K. (2021). Translanguaging in the development of EFL learners' foreign language skills in the Turkish context. *Innovation in Language Learning and Teaching, 16*(2), 176–190. https://doi.org/10.1080/17501229.2021.1892698	
Lundmark, A. (2025). *L1 use in the Swedish EFL classroom and what language policy it implies* (Doctoral dissertation). Luleå University of Technology. https://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-112647	
Lorenz, E., Toprak-Yildiz, T. E., & Siemund, P. (2024). L3 English in the German secondary school context: longitudinal development of bilingual heritage speakers' multilingual repertoire. International Journal of Multilingualism, 22(1), 76–97. https://doi.org/10.1080/14790718.2024.2345215	Wrong study type
Maluch, Jessica T., Marko Neumann & Sebastian Kempert. 2016. Bilingualism as a resource for foreign language learning of language minority students? Empirical evidence from a longitudinal study during primary and secondary school in Germany. Learning and Individual Differences 51. 111–118.	
Ndhlovana, S. N. (2025). Unleashing the Power of Multilingualism Through Pedagogical Translanguaging in Foundation Phase. In E. Charamba & S. Ndhlovana (Eds.), Improving Academic Performance and Achievement With Inclusive Learning Practices (pp. 327-350). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-4501-7.ch014	
Pintado Gutiérrez, L. (2018). Translation in language teaching, pedagogical translation, and code-Switching: restructuring the boundaries. The Language Learning Journal, 49(2), 219–239. https://doi.org/10.1080/09571736.2018.1534260	

Moore, P. (2023). Translanguaging in CLIL. In D. L. Banegas & S. Zappa-Hollman (Eds.), The Routledge handbook of content and language integrated learning (1st ed., pp. 15–29). Routledge. https://doi.org/10.4324/9781003173151	
Goodman, B., and S Tastanbek. 2020. Making the shift from a codeswitching to a translanguaging lens in English language teacher education. TESOL Quarterly 55: 29–53. —> very important text	
Barberán-Recalde, M. T., & Basterrechea, M. (2025). Integration of grammar instruction and task-based interaction with young EFL learners: An investigation into the use of English possessive determiners (his/her). In M. d. P. García Mayo (Ed.), Investigating attention to form and individual differences: Research with EFL children (pp. 1-27. Springer. https://doi.org/10.1007/978-3-031-80924-8_1	No use of L1
Hopp, H. (2018). Cross-linguistic influence in the child L3 acquisition of grammar: Sentence comprehension and production among Turkish-German and German learners of English. International Journal of Bilingualism, 23(2), 567 - 583. DOI: 10.1177/1367006917752523	
Hopp, H., Vogelbacher, M., Kieseier, T., & Thoma, D. (2019). Bilingual advantages in early foreign language learning: Effects of the minority and the majority language. Learning and Instruction, 61, 99–110. https://doi.org/10.1016/j.learninstruc.2019.02.001	Indirect measure
Soto-corominas, A., Daskalaki, E., Paradis, J., Winters-difani, M., & Janaideh, R. A. (2022). Sources of variation at the onset of bilingualism: The differential effect of input factors, AOA, and cognitive skills on HL Arabic and L2 English syntax. Journal of Child Language, 49(4), 741–773. https://doi.org/10.1017/S0305000921000246	
Hopp, H., Thoma, D., Kieseier, T., Jakisch, J., & Sturm, S. (2022). Early plurilingual foreign language teaching in mixed classrooms: Linguistic and metalinguistic predictors of foreign language development among majority-language and minority-language students in EFL classes. System (Linköping), 108, Article 102828. https://doi.org/10.1016/j.system.2022.102828	
Ikeda, M. (2022). Pedagogical translanguaging in primary school math CLIL lessons in Japan. In Soft CLIL and English Language Teaching (1st ed., pp. 86–102). Routledge. https://doi.org/10.4324/9780429032332-5-6	
Neil, P., Salters, J., & McEwen, A. (1999). Teachers' use of target language in the German classroom. The Language Learning Journal, 19(1), 12–18. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1080/09571739985200041	
Gabillon, Z. (2020). Analysing content and language integrated learning (CLIL) and English as an additional language (EAL) interactions and exploring representations in a plurilingual context [Doctoral dissertation, Université de la Polynésie française]. TEL Archives ouvertes. https://hal.science/tel-03725623	Non-ELL context
Scott, V. M., & Fuente, M. J. D. L. (2008). What's the Problem? L2 Learners' Use of the L1 During Consciousness-Raising, Form-Focused Tasks. The Modern Language Journal (Boulder, Colo.), 92(1), 100–113. https://doi.org/10.1111/j.1540-4781.2008.00689.x	
Hopp, H., Steinlen, A., Schelletter, C., & Piske, T. (2019). Syntactic development in early foreign language learning: Effects of L1 transfer, input, and individual factors. Applied Psycholinguistics, 40(5), 1241–1267. https://doi.org/10.1017/S0142716419000249	No report of intervention
Sturm, S., Jakisch, J., Kieseier, T., & Hopp, H. (2024). Plurilingual foreign-language teaching in early EFL: exploring student perspectives and profiles. International Journal of Multilingualism, 21(2), 689–710. https://doi.org/10.1080/14790718.2022.2105849	
Coelho, D., Khalil, N., & Shankar, D. D. (2022). Teachers' perceptions and 'invisible' uses of plurilingual pedagogy in UAE K-12 schools. International Journal of Multilingualism, 21(2), 670–688. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1080/14790718.2022.2099400	
Tannenbaum, M., Haim, O., & Shohamy, E. (2024). Incorporating translanguaging in English instruction: teachers' beliefs, practices, and influencing factors. ***International Multilingual Research Journal***, 1–18. https://doi-org.ezproxy-prd.bodleian.ox.ac.uk/10.1080/19313152.2024.2418158	
Hopp, H., Kieseier, T., Jakisch, J., Sturm, S., & Thoma, D. (2021). Do minority-language and majority-language students benefit from pedagogical translanguaging in early foreign language development? Multilingua, 40(6), 815–837. https://doi.org/10.1515/multi-2020-0164	No measurement of effect on grammar development
Hopp, H., Schimke, S., Gastmann, F., Öwerdieck, D., & Poarch, G. J. (2024). Processing to learn noncanonical word orders: Exploring linguistic and cognitive predictors of reanalysis in early L2 sentence comprehension. Studies in Second Language Acquisition, 46(3), 686–709. doi:10.1017/S0272263123000451	
Busse, Vera, Jasone Cenoz, Nina Dalmann & Franziska Rogge. 2020. Addressing linguistic diversity in the language classroom in a resource-oriented way: An intervention study with primary school children. Language Learning 70(2). 382–419. https://doi.org/10.1111/lang.12382.	

Ticheloven, A., Schwenke-Lam, T., Fürstenau, S., Duarte, J., & Kirsch, C. (2020). Multilingual teaching practices in primary classrooms in Germany: Language comparisons. In Multilingual Approaches for Teaching and Learning (1st ed., pp. 34–51). Routledge. https://doi.org/10.4324/9780429059674-4	
Hopp, H., Schimke, S., Öwerdieck, D., Gastmann, F., & Poarch, G. J. (2025). Learning via Processing: Structural Priming Across Grammatical Structures and Languages in Early Second Language Development. Language Learning. https://doi.org/10.1111/lang.12711	Lab setting

Appendix 7 Excluded studies that are highly relevant but deemed unsuitable

Citation	Reason(s) for exclusion
Abourehab, Y., & Azaz, M. (2020). Pedagogical translanguaging in community/heritage Arabic language learning. *Journal of Multilingual and Multicultural Development*, *44*(5), 398–411. https://doi.org/10.1080/01434632.2020.1826496	Examines the use of PTL in **Arabic** grammar development
Abdurakhmonov, M. (2024). Own-language use in EFL classrooms in Uzbekistan: perceptions, practice, and purposes.	Participants include both teenagers and adults, but **no separate analyses** are reported for the two groups
Hosseininik, S. Y., & Hamid Rahmani Sangani. (2014). The Influence of Explicit Cross-Linguistic Consciousness- Raising on the EL Writing of the Iranian English Language Learners. Advances in Language and Literary Studies, 5(3), 147–157. https://doi.org/10.7575/aiac.alls.v.5n.3p.147	Grammar is assessed through error correction and translation tests, independent of writing competence. **Learners' age unspecified** and attempts to obtain this information from the authors were unsuccessful.
Meng, R. (2014). An analysis of senior high English teachers' code-switching in grammar teaching (Master's thesis). Sichuan Normal University.	Presents only **self-reported learning gains** from a questionnaire
Yasar Yuzlu, M., & Dikilitas, K. (2021). Translanguaging in the development of EFL learners' foreign language skills in Turkish context. *Innovation in Language Learning and Teaching*, *16*(2), 176–190. https://doi.org/10.1080/17501229.2021.1892698	Grammar is evaluated in a writing task; **no specific or isolated measures** are provided.
Hopp, H., Kieseier, T., Jakisch, J., Sturm, S. & Thoma, D. (2021). Do minority-language and majority-language students benefit from pedagogical translanguaging in early foreign language development?. Multilingua, 40(6), 815-837. https://doi.org/10.1515/multi-2020-0164	Learning is measured through a **general grammar proficiency test**. No clear evidence of growth on specific grammar items introduced.
Çiftci, H., & Ozcan, M. (2021). A contrastive analysis of Traditional Grammar Translation Method and Communicative Language Teaching in teaching English grammar and vocabulary. International Online Journal of Education and Teaching (IOJET), 8(2). 709-729.	L1 is incorporated, but **variability in teaching approaches** (e.g. discourse-based grammar teaching/CLT) makes it difficult to isolate the impact of language choice on grammar teaching.
Zhu, Z. (2022). Effects of discourse-based grammar teaching on English grammar learning of senior high school students (Master's thesis, Northwest Normal University). Northwest Normal University.	
Rodríguez-Valls, F. (2009) "Comparando simetrías lingüísticas: estudiantes y docentes s analizan el español y el inglés en el aula" *Archivos Analíticos de Políticas Educativas*,17 (19). Recuperado [30 June, 2025] de http://epaa.asu.edu/epaa/ [Translation: Rodríguez-Valls, F. (2009). "Comparing linguistic symmetries: Students and teachers analyze Spanish and English in the classroom." Analytical Archives of Educational Policies, 17(19). Retrieved [30 June, 2025] from http://epaa.asu.edu/epaa/]	L1 use emphasises cross-linguistic transfer both in and outside classroom. ELL focus is on **recognising learners' language identity**; fluency development is a secondary goal.

Appendix 8 Example entry from data extraction table

	Item	Data
Study identification	Title	Effects of Plurilingual Teaching on Grammatical Development in Early Foreign-Language Learning
	Author(s)	Hopp, H., Thoma, D.
	DOI/URL	https://doi.org/10.1111/modl.12709
	Publication year	2021
	Journal	Modern Language Journal (John Wiley & Sons, Inc.)
	Database	ERIC
	Study Aims	To investigate whether effects of plurilingual FL teaching are modulated according to the differences and similarities between grammars of the languages involved. To test whether majority language learners and minority language learners demonstrate differential learning gains to address whether plurilingual FL teaching has compensatory or additive effects in grammatical learning.
	Research questions	RQ1. Does PTL by virtue of the use of contrastive language activities promote the acquisition of foreign-language grammar? RQ2: Do effects of PTL extend across different grammatical phenomena? RQ3: Are there differences in the degrees to which PTL affects the grammatical development of majority language and minority language learners?
Intervention Details	L1-mediated strategy label	Pedagogical Translanguaging (PTL); approach by Cenoz & Gorter, 2017; 2020)
	Implementation of strategy	1. Cross-linguistic comparison involving German (L1) and minority languages 2. Eliciting examples of target structure in Ss' L1/heritage language 3. No presentation of explicit grammar rules (P. 479)
	Grammar focus	Experiment 1: Word order in wh-questions (Different from German) Experiment 2: Passivisation (Similar to German)
	Topic/Theme	Birthdays (Animal birthday party)
	Lesson activities	Lesson 1: Animacy contrasts in nouns (i.e. non-reversible actions) Presenting target structure using picture cards Practising target structures in comprehension and production tasks Lesson 2: Animacy contrasts in nouns (i.e. reversible actions) and **additional language comparisons (intervention)** Presenting target structure using picture cards Practising target structures in comprehension and production tasks **Language comparison: eliciting examples of target structure from other languages spoken by the learners, denoting cross-linguistic differences**

		PTL group wrote target structures in English, German, and other languages in a worksheet Control group: further practices in English
	Instructional materials	Playway 4 (Gerngross, Puchta & Becker, 2013)
	Duration	PTL group: Plurilingual instruction for 20% of class time Two 45-minute lessons (embedded in a 6-month project on effects of plurilingual FL teaching on the development of general (meta-)linguistic and cognitive skills
	Teacher	A project member, who is a qualified English teacher; regular English teachers were present during the interventions and supervised the teaching
	Teachers' knowledge of learners' L1	Not specified
Participants information	Context	German EFL; Experiment 1/Cohort 1: 2018-2019; Experiment 2/Cohort 2: 2019-2020
	Education setting	4 regular public primary schools in medium-sized cities in Lower Saxony
	Year level	4th Grade
	Participants' age	Cohort 1: 8.0-9.9 years Cohort 2: 7.0-9.7 years
	Students' L1	66% German-L1 33% German-dominant speakers of minority languages
	Other languages spoken	Albanian, Arabic, Chinese, Croatian, Farsi, Greek, Indonesian, Italian, Kurdish, Polish, Romanian, Russian, Spanish, Turkish, and Vietnamese
	Length of ELL	Two 45-minute lessons per week since Grade 3, i.e. about 1 year 3 months
	English proficiency level	Not specified, but results of English receptive and grammar tests are available P. 469
	Sample size	258 (Cohort 1: 125; Cohort 2: 131)
	Intervention group (s)	Cohort 1: 62 (33 were minority-language speakers) Cohort 2: 69 (30 minority-language speakers)
	Control group(s)	Cohort 1: 53 (26 were minority-language speakers) Cohort 2: 58 (26 minority-language speakers)
	Grouping information	2 intact classes in 4th grade at each school
	Baseline control	Cognitive processing 1. Nonverbal IQ test (CFT20-R; Weiβ, 2006) 2. Working memory (a forward digit span task adapted from HAWIK-IV; Petermann & Petermann, 2008). Language proficiency in German 1. A standardized 32-item picture naming task (Cross-Linguistic Lexical Task, CLT; Rinker & Gagarina, 2014) Language proficiency in English 1. A receptive vocabulary task (British Picture and Vocabulary Scale, BPVS3; Dunn, Dunn, & Styles, 2009)

		2. A receptive grammar task (Test for Reception of Grammar, TROG2; Bishop, 2003). Results: All groups were comparable in all measures (ps>.11). In Cohort 1, No significant differences in Group/Language status, except for lower German proficiency group among minority language students ($F(1,121) = 19.027$, $p < .001$)). Majority language speakers in Cohort 2 were younger ($F(1,136) = 7.020$, $p = .009$), with higher IQ ($F(1,136) = 7.510$, $p= .007$), and were more proficient in German ($F(1,136) = 10.589$, $p <.001$.).
	Attrition and reasons	Cohort 1: Pretest and posttest data from 115 students; one removed from control group (Reason: English being one of their L1s) Cohort 2: Data from 125 students: two from PTL and one from control were removed as they were English-L1 bilinguals. Three additional participants removed from analysis for scoring at or below chance on filler items at both pre- and posttests, suggesting they were not paying attention during the task.
Methodology	Type of data collected	Quantitative
	Study design	Clustered RCT, pre-posttest design
	Method of randomisation, if applicable	Random assignment to experimental or control condition
	Data collection tools/procedures	1. Duration of test: Around 25 minutes 2. 10 quadruplets of items with familiar animal names 3. Items in the tests: 10 wh-questions (5 subject and 5 object questions) 10 relative clauses (5 subject and 5 object questions) 10 filler items Students circle the target animal (i.e. the answer) 4. Procedures: Task explained in German and answered questions prior to the start of the test. Teacher introduced a practice and filler sentence to familiar students with the test. All animals were named, and the experimental sentences were read aloud twice in a slow pace before students circled the answers.
	Validation of assessment	1. Two lists of each test, counterbalancing the order of lists between pre and posttests across intervention and control group 2. Tests completed in class, under the supervision of the teacher 3. No feedback given.
Outcomes	Outcome type	Learning gains
	Outcome unit	Pre-posttests scores (Quantitative)
	Outcome	**Experiment 1 (Target: Wh-questions; non-target: Relative clause):** 1. Only PTL group showed learning gains on object wh-questions; both majority and minority language learners in PTL group experienced learning gains in comparable measure.

		2. High accuracy on subject questions and subject relative clauses; lower accuracy for object-first orders (significant main effect of orders) 3. Wh-questions: a marginally significant 3-way interaction between order, test and group; a significant 2-way interaction between test and group, and a main effect of test 4. Relative clause: No further significant effect beyond the main effect of order 5. Only PTL group showed an interaction between order and test (β = 1.21, SE= 0.38, z = 3.22, p = .001, illustrating that gains from pretest to posttest were specific to object questions in the PTL group (Mdiff = 14.8%; control group: Mdiff = 5.8%). In contrast, the control group evinced a main effect of test, which reflects a general moderate increase in performance from pretest to posttest) **Experiment 2 (Target: Passive, non-target: wh-questions):** 1. Both groups displayed high comprehension accuracy on active sentences. 2. Accuracy of passives increased for both groups from below to around chance in pre-posttest comparison. 3. Both groups obtained gains for passive sentences (PTL: Mdiff = 12.3%; control: Mdiff = 19.3%) 4. Differential learning gains for passive between majority and minority language learners: No interaction between test and language status (β = 0.60, SE = 0.40, z= 0.151, p = .880) 5. No transfer from passive to wh-questions, but students scored higher in object-questions in pre-posttest comparisons.
	Statistical tests used	Mixed logistic regression modelling, reporting fixed effects
	Effect size, if available	**Experiment 1: Object wh-questions** 1. PTL group: Medium effect size (β = 1.06, SE = 0.49, z = 2.187, p = .029, d = 0.466, 95% CI [0.110, 2.012]); 2. Control group: Small effect size that crossed zero (β = 0.26, *SE* = 0.52, *z* = 0.488, p = .626, d = 0.134, 95% CI [–1.279, 0.769].) **Experiment 2: Passivisation** 1. PTL group: Medium effect size (β = 0.79, SE = 0.20, z = 4.029, p < .001, d = 0.459, 95% CI [0.403, 1.168]; 2. Control group: Medium effect sizes (β = 1.15, SE = 0.21, z = 5.422, p < .001, d = 0.676, 95% CI [0.731, 1.558]) 3. Effect sizes of Control group>PTL, but CI for neither group crossed zero
Others	Source of funding	Part of MEG-SKoRe Project, details can be found in Hopp, H., Jakisch, J., Sturm, S., Becker, C., & Thoma, D. (2020). Integrating multilingualism into the early foreign language classroom: Empirical and teaching perspectives. International Multilingual

		Research Journal, 14(2), 146–162. https://doi.org/10.1080/19313152.2019.1669519
	Peer review status	Peer reviewed
	Limitations reported by authors	1. Possibility of learners in control group discovering the cross-linguistic parallels in passivisation among the majority group, but they may not have capitalised on the differences 2. Lack of cross-construction transfer across target and non-target structures: gains limited to pattern-specific learning, so longer instructional treatments are needed (Suggesting a potential Type 2 error) 3. Absence of productive knowledge test/delayed posttest in understanding generalisation of modalities or time 4. Limited generalisability because of age and linguistic backgrounds of learners

Appendix 9. Summary table of RoB appraisals for all seven included studies

Study Design			Hopp & Thoma (2021)	Mehrseresht et al (2015)	Fan (2022)	He (2017)	Shabaka-Fernardez (2023)	Khonamri et al (2022)	Gallagher & Colohan (2017)
RCT	Internal validity	Selection and allocation bias							
		Treatment bias							
		Assessment/outcome bias							
		Attrition bias							
	Statistical validity								
Quais-experimental	Internal validity	Temporal precedence bias							
		Selection and allocation bias							
		Confounding factors bias							
		Treatment bias							
		Assessment/outcome bias							
		Attrition bias							
	Statistical validity								
Qualitative	Congruity between research methodology and …	Stated philosophical perspective							
		Research questions/objectives							
		Data collection							
		Data representation and analysis							
		Interpretation of results							
		Researcher(s)' cultural/theoretical positionality							
		Researcher(s)' influence on the research and vice versa							
		Representation of participants and voice							
		Ethical approval							

		Alignment of conclusion with data							
Overall RoB									

High RoB	Moderate RoB	Low RoB

Appendix 10 Sample of a completed RoB Appraisal Form

10a. JBI critical appraisal tool for randomized controlled trials – Fan (2022)

RoB Assessor: Chi Lam	**Date of Appraisal: 9 July**	**Record Number: Fan (2022)**
Study Author: Liuyi Fan	**Study Title: Effects of Pedagogical Translanguaging on Grammar Learning for Chinese Primary School Students in Early Foreign Language Learning**	**Study Year: 2022**

Internal Validity		**Choice - Comments/Justification**	**Yes**	**No**	**Unclear**	**N/A**
Bias related to selection and allocation						
1	**Was true randomization used for assignment of participants to treatment groups?**	P. 26-27 145 participants were randomly divided into three groups, the Translanguaging (TL) group (n=50), the English-only (EO) group (n=50), and the Control group (n=45). The numbers of participants in the TL group and the EO group were set to be even numbers before the randomization because participants in these two groups will need to complete a task in pairs as part of their treatment.	☒	☐	☐	☐
2	**Was allocation to treatment groups concealed?**		☐	☐	☒	☐
3	**Were treatment groups similar at the baseline?**	Matching EFL length, age and proportion of gender	☒	☐	☐	☐
Bias related to administration of intervention/exposure						

4	**Were participants blind to treatment assignment?**	Learners seem to know the condition of the other group (P. 54-55) – 'trade everything to be in the TL group' Different days of assessment – possible communication among participants	☐	☒	☐	☐
5	**Were those delivering the treatment blind to treatment assignment?**	Researcher delivering the content, and classroom teacher administering the test. But both are not directly contributing to effect.	☐	☒	☐	☐
6	**Were treatment groups treated identically other than the intervention of interest?**	Same tutorial video	☒	☐	☐	☐
Bias related to assessment, detection and measurement of the outcome						
7	**Were outcome assessors blind to treatment assignment?**		**Yes**	**No**	**Unclear**	**N/A**
	Translation Test	Done by researcher, but objective measure	☐	☒	☐	☐
	GJT (Accuracy, RT)	Objective measure	☐	☐	☐	☒
8	**Were outcomes measured in the same way for treatment groups?**		**Yes**	**No**	**Unclear**	**N/A**
	TT, GJT (Accuracy and RT), Total number of translanguaging sentences, questionnaire		☒	☐	☐	☐
9	**Were outcomes measured in a reliable way**		**Yes**	**No**	**Unclear**	**N/A**
	Procedure	Participants were only informed of their partners immediately before the experiment activity to avoid possible communications	☒	☐	☐	☐

	Translation Test	P. 39 Researcher (assessor) blind to the condition (TL/EO/control) and test sessions (pre/post) Two versions of test	☒	☐	☐	☐
	Grammaticality judgement test	Two versions of test	☒	☐	☐	☐
	Production of translanguaging sentence during exchange	Recording	☒	☐	☐	☐

Bias related to participant retention

10	**Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analyzed?**					
	Click or tap here to enter text.		**Yes**	**No**	**Unclear**	**N/A**
	GJT	P. 40 One TL participant's data in GJT removed from analysis because of invalid response	☒	☐	☐	☐
	Total # of translanguaging sentences and incorrect output	P. 49 25>24 pairs in the EO group due to tech issues	☒	☐	☐	☐

Statistical Conclusion Validity

11	**Were participants analyzed in the groups to which they were randomized?**					
	Click or tap here to enter text.		**Yes**	**No**	**Unclear**	**N/A**

	Click or tap here to enter text.		☒	☐	☐	☐

12	**Was appropriate statistical analysis used?**					
	Translation test and GJT – pre-posttest		**Yes**	**No**	**Unclear**	**N/A**
	Correct rates	P. 43 Mixed ANOVA (Results on P. 45)	☒	☐	☐	☐
	Learning gains in correct rates	P. 43 One-way ANOVA and Helmert Contrasts	☒	☐	☐	☐
	Translanguaging sentences per pair		**Yes**	**No**	**Unclear**	**N/A**
	Total number of target structures	P. 49 Descriptive statistics, independent samples t-test	☒	☐	☐	☐
	% of Incorrect sentences	P. 51-52 independent samples t-test	☒	☐	☐	☐

			Yes	**No**	**Unclear**	**N/A**
13	**Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?**	P. 44, P. 46 Assumptions of ANOVA tested: Correct rates, gains of correct rates in TT & GJT (DVs) met the assumptions of normal distribution and homogeneity of variance (Appendix 14 and 15). P. 49, P. 51 Normal distribution for total target sentences (for t-test)	☒	☐	☐	☐

Overall appraisal:

High risk: ☐

Moderate risk: ☐

Low risk: ☒

Comments:

Master's thesis
Setting different from a usual classroom (pulled-out)
Test conducted on different dates – possible exchange of experience among participants

10b. JBI Critical Appraisal Checklist for Qualitative Research

Reviewer_____Chi Lam___Date__10 July 2025_____

Author__Liuyi Fan _____Year__2022_____ Record ID __Fan (2022)_____

	Yes	No	Unclear	Not applicable
11. Is there congruity between the stated philosophical perspective and the research methodology?	☒	☐	☐	☐
12. Is there congruity between the research methodology and the research question or objectives?	☒	☐	☐	☐
13. Is there congruity between the research methodology and the methods used to collect data? Group interview: more relaxing with presence of peers, can interact with one another; conducted in L1, but allowed to use either lang	☒	☐	☐	☐
14. Is there congruity between the research methodology and the representation and analysis of data? Thematic coding scheme, inter-coder agreement (96%) P. 41	☒	☐	☐	☐
15. Is there congruity between the research methodology and the interpretation of results?	☒	☐	☐	☐
16. Is there a statement locating the researcher culturally or theoretically? Does not affect implementation fidelity	☐	☒	☐	☐
17. Is the influence of the researcher on the research, and vice- versa, addressed? Researcher delivered the tutorial video on target structure	☒	☐	☐	☐
18. Are participants, and their voices, adequately represented?	☒	☐	☐	☐
19. Is the research ethical according to current criteria or, for recent studies, and is there evidence of ethical approval by an appropriate body? CUREC approved	☒	☐	☐	☐
20. Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data?	☒	☐	☐	☐

Overall appraisal: High risk ☐ Moderate risk ☐ Low risk ☒

Comments: