Language Teaching (2026), 1–22
doi:10.1017/S0261444826101335

CAMBRIDGE UNIVERSITY PRESS

STUDY

A multilevel Bayesian meta-analysis of the effects of translanguaging pedagogy on L2 achievement

Xiaoqi Wang and Lawrence Jun Zhang

Faculty of Arts and Education, University of Auckland, Auckland, New Zealand
Corresponding author: Lawrence Jun Zhang; Email: lj.zhang@auckland.ac.nz

(Received: 8 January 2026; accepted: 7 May 2026)

Abstract

Globalisation and the multilingual turn in applied linguistics have exposed a persistent tension between learners' linguistic repertoires and traditional monolingual pedagogies. It is out of this tension that translanguaging pedagogy (TP) emerged as an instructional framework grounded in translanguaging theory, witnessing a burgeoning empirical literature on its classroom value. However, the empirical evidence has failed to produce a coherent account of TP's effectiveness in L2 education, as findings have exhibited considerable heterogeneity across instructional contexts, learner populations, and outcome measures. The present study sought to reconcile these inconsistent findings through a multilevel Bayesian meta-analysis of 108 effect sizes drawn from 40 independent studies among 3,145 learners. The pooled effect size of Hedges' $g = 1.014$ (standard error = 0.138, 95% credible interval [0.748, 1.291]) establishes a positive and statistically credible effect of TP on L2 achievement. Moderator analyses identified educational level and target language skills as credible sources of heterogeneity, with tertiary-level learners and speaking outcomes yielding the largest effect sizes. Target language, linguistic distance, and treatment duration did not reach the posterior probability threshold for credible inference. These findings advance the theoretical understanding of TP and furnish an empirically grounded basis for instructional decision-making in L2 education.

Keywords: L2 achievement; multilevel Bayesian meta-analysis; second language acquisition; translanguaging; translanguaging pedagogy

1. Introduction

The 'multilingual turn' contests the monolingual ideology by repositioning linguistic practices as a semiotic repertoire that encourages researchers to examine how language users mobilise linguistic, cultural, and social resources to construct meaning in multilingual contexts (May, 2014; see also Li, 2018; Liu & Fang, 2022). This conceptual reorientation is rooted in contemporary heteroglossic environments, where named languages, dialects, and registers circulate across domestic, educational, and digital spaces (e.g., Blackledge & Creese, 2014; Nagashima & Lawrence, 2022; Song, 2016; Zhao & Flewitt, 2020). Such heterogeneity reflects the complexity of language use in today's globalised world. It is, therefore, irrational to expect multilingual individuals to behave as monolinguals in that their language use is deeply connected to personal and cultural spheres and the interactions among them (e.g., Abourehab & Azaz, 2023; Fisher et al., 2020; Liao et al., 2025; Liu & Fang, 2022; Richards & Wilson, 2019). This multilingual environment calls for pedagogical practices that recognise and respond to multilingual learners' language practice patterns. Against this backdrop, translanguaging pedagogy (TP), as a pedagogical framework that legitimates learners' use of their full multilingual

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

resources, has increasingly gained scholarly attention (Cenoz & Gorter, 2020; Cummins, 2019; see also interviews with Li and Canagarajah by Zhang, 2022a, 2022b).

An increasing body of research has documented that TP nurtures language development, cognitive flexibility, emotional security, as well as learner agency by satisfying learners' multilingual needs across different instructional contexts (Dryden et al., 2021; Rajendram, 2023; Wawire & Barnes-Story, 2023; Yasar Yuzlu & Dikilitas, 2022). Although the majority of studies have highlighted positive impacts of TP on language development (e.g., Busse et al., 2020; Luo & Sun, 2025; Yasar Yuzlu & Dikilitas, 2022), Hopp et al. (2021) and Qureshi and Aljanadbah (2022) held different opinions. They found that TP did not contribute significantly to language performance. The mixed results might be attributed to the differences in study designs, instructional contexts, assessment standards, or sample sizes. This variability may prevent teachers from determining when, how, and to what extent such practices can be applied in instructional contexts. Previous studies have provided valuable insights into the effects of TP and have clarified its theoretical grounding (e.g., Sun & Zhang, 2022). However, this evidence is largely fragmented, with findings drawn from individual studies that vary in design, context, and outcome measures. This lack of cumulative evidence limits a comprehensive understanding of TP effectiveness across contexts.

Due to the heterogeneity in research designs, sample characteristics, and effect sizes reported across the current literature, a multilevel Bayesian meta-analytic approach might be appropriate. Unlike traditional frequentist meta-analysis, Bayesian methods can provide probabilistic statements about effect sizes and accommodate various sources of uncertainty (Bürkner, 2017; Reis et al., 2023; Vehtari et al., 2023 2017; Williams et al., 2018). This approach also allows researchers to quantify the probability that the true effect falls within a range of practical significance. Moreover, a three-level meta-analytic model is particularly suitable when multiple effect sizes are reported within individual studies, as it accounts for the dependency among effect sizes nested within the same study (Assink & Wibbelink, 2016; Cheung, 2014). Therefore, the present study aims to achieve two research objectives: (1) to estimate the overall effectiveness of TP on L2 achievement by synthesising effect sizes from existing empirical studies; and (2) to identify and examine potential moderators that may account for variability in the effectiveness of TP. This study will contribute to research syntheses in applied linguistics by providing a methodological demonstration of Bayesian multilevel meta-analysis and advancing empirical understanding of TP effectiveness.

2. Literature review

2.1. Translanguaging and TP

The theory of translanguaging rose from classroom practice, then returned there to be tested and refined. Each cycle of implementation reshaped what the theory claimed. It is recognised that translanguaging originates from bilingual classrooms in Wales, where Williams (1994) first coined the term TRAWSIEITHU to describe the practice of switching between Welsh and English in language teaching practices. It was Baker (2001) who provided a voice in English for the Welsh term trawsieithu by translating it into translanguaging. Since its introduction, translanguaging has increasingly attracted attention as a pedagogy and a theoretical framework for analysing diverse language usage. García (2009) extended earlier models to conceive translanguaging as a dynamic discourse practice through which bilinguals mobilise their full linguistic repertoire to make sense of their multilingual worlds across language boundaries. Canagarajah (2011) foregrounded attention towards the rhetorical and strategic nature of multilingual communication, defining translanguaging as 'the ability of multilingual speakers to shuttle between languages, treating the diverse languages that form their repertoire as an integrated system' (p. 401). He emphasised multilinguals' strategic use of their integrated repertoires to negotiate meaning in situated communicative contexts. García and Li (2015) further propelled the evolution of translanguaging theory. In their collaborative work, they argue that translanguaging challenges monolingual ideologies and valorises marginalised linguistic resources.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Translanguaging, in their view, can create spaces to afford bilingual and multilingual learners the opportunity to exercise epistemic agency, negotiate identities, and engage with academic content through their full spectrum of linguistic and semiotic repertoires. More recently, Li (2018) described translanguaging as a practical theory of language, where meaning develops through the dynamic and integrated use of multilingual, multisemiotic, multisensory, and multimodal resources in interaction. From his perspective, communication is an active process. It is multilingual, multisemiotic, multisensory, and multimodal, shaped by the coordination of cognitive, semiotic, sensory, and multimodal resources in real-time social interaction. Based on these insights, we can argue that translanguaging can be understood as a dynamic and intentional process through which multilingual individuals draw on integrated linguistic and semiotic resources to make meaning and participate in social and academic practices.

The theoretical development of translanguaging has led to related teaching approaches, most notably TP. This approach draws on students' multilingual resources to raise awareness of language, support both language and subject learning, and promote active participation in linguistically diverse classrooms (Cenoz & Gorter, 2021, 2020). García et al. (2017) outlined TP in terms of three connected components: stance, design, and shift. Stance, a philosophical, ideological, or beliefsystem that teachers draw from to design pedagogical frameworks, treats learners' language repertoire as a resource, never as a deficit. Design involves the systematic planning of instruction and assessment around the integration of the language and cultural practices that students bring from their home, school, and community environments. Shifts capture the teachers' moment-by-moment decisions through which the course of the lesson and assessment together with language use are adjusted to release and support students' voices. Cenoz and Gorter (2017) differentiated between pedagogical and spontaneous translanguaging. Pedagogical translanguaging involves the planned and systematic use of students' full language resources across teaching activities, with the aim of supporting learning in an organised way. Spontaneous translanguaging, in contrast, occurs naturally and without prior planning in response to learners' communicative needs. Although it is unplanned, spontaneous translanguaging still holds pedagogical value because it reflects the reality of how language resources are naturally used and teachers can use it as an instructional strategy in the classroom (Cenoz & Gorter, 2021). Rather than treating pedagogical and spontaneous translanguaging as separate or competing practices, Cenoz and Gorter (2020, 2021) view them as points along a single continuum. This continuum captures the close relationship between teacher planning and learner agency in multilingual classrooms. By encouraging learners to draw on their full linguistic and semiotic resources, TP broadens opportunities for meaning-making beyond the boundaries of named languages (Cenoz & Gorter, 2020, 2021). This flexibility supports learners in engaging with complex academic content and in taking part in the shared construction of subject knowledge through multilingual and multimodal interaction (Li, 2018). Based on these theoretical and pedagogical insights, we define TP in the present study as instructional practices that deliberately integrate students' full linguistic resources to support content learning and meaning-making. This definition includes both planned classroom practices and teachers' responsive adjustments to learners' spontaneous cross-linguistic use during interaction.

2.2. Effects of TP on L2 achievement

TP has turned out to not merely contribute to general language development but also tailor its contributions to specific language skills, such as vocabulary, speaking, listening, grammar, and writing. In the following section, we review TP studies that focus on these specific language skills to reveal its nuanced role in language learning.

For listening skills, Zhou et al. (2024) carried out a mixed-methods study to explore the influence of multilingual note-taking on listening comprehension among 90 Chinese English-as-a-Foreign-Language (EFL) university students. Based on the principles of translanguaging, Zhou et al. discovered that students who adopted multilingual and multimodal strategies attained better listening comprehension as compared to the students who were restricted to using only English or Chinese

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

for taking notes. By leveraging their entire linguistic and semiotic resources, these students benefited from organising and interpreting the listening content and showed a stronger inclination to insert multilingual and semiotic assets in their listening tasks. Robillos (2023) also implemented translanguaging into metacognitive listening and writing instruction in his 11-session research targeting 16 college students. The author found a significant increase in the listening comprehension scores of students after the intervention using TP. The mean of the pre-test was 9.19, while the mean of the post-test was 15.56. Turning to speaking skills, Luo and Sun (2025) ran a quasi-experiment testing translanguaging against L2-only planning in EFL speaking tasks. Drawing on task-based teaching and Levelt's (1989) speech production model, they reported that the translanguaging group demonstrated significantly greater syntactic complexity and idea development (e.g., higher clause ratios, longer AS-units, and more idea units) than the comparison group. Translanguaging also supports reading comprehension. Mgijima (2021) tested this in a quasi-experiment with Grade 4 isiXhosa-English bilinguals in South Africa. Students who received translanguaging-based instruction showed some improvement in predicting narrative events, with stronger and more consistent gains observed in their home language (isiXhosa) than in their target language (English). But the evidence was mixed across different analyses. Zhang (2023) reported similar gains in translation-based writing. When Chinese EFL students were encouraged to draw on their L1 during English composition, it worked as a productive resource, supporting both writing performance and learner confidence. Moreover, TP has also lent support to grammar and vocabulary acquisition (e.g., Hopp & Thoma, 2021; Sun, 2024). For example, Hopp and Thoma (2021) demonstrated that pedagogical translanguaging, through cross-linguistic comparisons among learners' majority, minority, and target languages, facilitated the acquisition of English object wh-questions. However, for grammatical features that were similar across languages, such as passives, both groups showed comparable gains.

While TP has shown the above pedagogical benefits, other studies have yielded inconsistent findings. Doubts about its effectiveness have been raised by teachers, who worry that the use of home languages may hinder target language development (Ticheloven et al., 2021). Students have also reported confusion during language alternation and expressed concern that such practices might lead to language mixing and affect their speaking and writing. Moreover, Hopp et al. (2021) did not find significant advantages of TP over monolingual instruction in vocabulary and grammar development for early foreign language learners. Qureshi and Aljanadbah (2022) also observed similar null effects for translanguaging instruction on L2 reading comprehension among Arabic–English bilingual learners in the United Arab Emirates. Sun and Zhang (2022) observed that although there were writing gains in the first few rounds of translanguaging-supported peer feedback, these advantages disappeared in later stages. These findings suggest that the pedagogical value of TP may not be consistent but conditional on specific instructional and learner-related factors. Wen et al. (2022) introduced a translanguaging-informed, genre-based pedagogical framework for designing tasks and classroom practices in business communication courses. They highlighted both the benefits and the challenges of applying this approach to curriculum development and professional training in superdiverse megacities such as the Great Bay Area.

2.3. Potential moderators

Among the factors that may influence the effectiveness of TP, educational level has been a hot topic in meta-analytic scholarship. Based on a comparative analysis of Korean heritage language learners in first and third grades, Lee (2024) found that, although students at both levels engaged in translanguaging to facilitate meaning- and sense-making processes, the specific mechanism differed. Younger learners tended to rely on translanguaging to deal with lexical uncertainty, and their use could not be separated from teachers' language use. By contrast, older students held more strategic regulation of language choices. Their translanguaging practices operated as a strategic resource for articulating evaluative stances. It also supported flexible alignment with interlocutors' language preferences and

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

interactional roles. For inter-educational-level differences, Zainuddin and Zaki (2023) found variance in translanguaging practices among Malaysian secondary and tertiary English as a second language (ESL) classrooms. Ulum (2024) also found educational-level differences, with students at different educational stages differing in how they evaluated and experienced translanguaging in terms of confidenceand in how they perceived its academic and classroom-related benefits. These perceptual and psychological differences across educational levels may condition how TP operates.

Despite growing empirical attention to the instructional benefits of TP, little research has examined whether its effectiveness varies with target language characteristics or with the linguistic distance between learners' first language and the target language. The similarities between first and second languages could influence cross-linguistic transfer and language development (Chiswick & Miller, 2005; Melby-Lervåg & Lervåg, 2011). Qureshi and Aljanadbah (2022) investigated the impact of TP on L2 reading comprehension among Arabic-speaking learners of English. But they found no significant difference between translanguaging and monolingual instructional conditions. When Teng and Fang (2024) examined Japanese learners acquiring Chinese, they found significant gains in morphological awareness under translanguaging instruction. Learners who engaged in translanguaging across Japanese, English, and Chinese outperformed those in the monolingual group. Given the theoretical plausibility that linguistic distance may affect the transferability of linguistic resources, this study considers linguistic distance as a potential moderator. In addition, as English occupies a distinct status as a global lingua franca with widespread exposure and pedagogical support, its instructional outcomes under TP may differ from those of other languages. Thus, target language type (English vs Languages other than English, LOTE) is also examined as a separate moderator.

Intervention duration constitutes another variable when assessing instructional effectiveness in second language acquisition (SLA). Existing studies on TP have adopted flexible intervention durations, ranging from a single session (e.g., Qureshi & Aljanadbah, 2022) to long-term interventions up to 6 months (e.g., Hopp & Thoma, 2021). This variation may reflect different instructional objectives and theoretical assumptions about how and why learners would benefit from translanguaging practices. Shorter interventions may be assumed to make immediate, task-related gains, whereas longer intervention durations are likely to focus on the development of metalinguistic awareness and the acquisition of cross-linguistic skills in the long run. But Plonsky and Oswald (2014) noted that although longer interventions generally yield larger effect sizes, the correlation between treatment length and learning outcomes tends to be small. These insights prompted the inclusion of treatment duration as a moderator to assess its moderating effect on TP's effectiveness.

TP produces gains in L2 learning, yet different language skills respond unevenly to its influence. As reviewed earlier, TP has been implemented to support learning outcomes in both receptive skills (e.g., reading and listening) and productive skills (e.g., speaking and writing), as well as in lexical and grammatical development. However, the mechanisms through which translanguaging supports these outcomes differ across skill domains. Reading may benefit from learners' capacity to draw on cross-linguistic resources and to transfer inferencing strategies across languages, as evidenced by improved text prediction outcomes among bilingual learners engaged in translanguaging-based reading instruction (Mgijima, 2021). Luo and Sun (2025) found that TP facilitates speaking production by enabling learners to use multiple linguistic resources during oral tasks, which can enhance communication efficiency and support idea exchange. As for lexical development, TP facilitates vocabulary discussion by enabling learners to draw on multilingual resources and collaboratively search for appropriate expressions (Teng & Fang, 2024). TP also improves listening comprehension through providing learners with multilingual and multimodal resources for note-taking, which may benefit encoding of listening input and meaning reconstruction (Zhou et al., 2024). These exemplar variations suggest that TP may function through different facilitative processes for the target language skills. Liao et al. (2025) conducted a qualitative meta-synthesis of ten English medium instruction (EMI) TP studies published between 2014 and 2024. Their review found a strong emphasis on qualitative methods, mainly case studies and thematic analyses, focused largely on tertiary Asian contexts

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

and on teachers' and students' translanguaging perceptions and practices. Noting limited quantitative, mixed-methods, and context-diverse research, the authors called for broader methodological approaches and expanded investigations across varied EMI settings and subject areas.

2.4. The present study

TP has gained recognition as a viable approach to promoting second and foreign language development. However, the field is still lacking a generalisable account of TP's impact on language achievement across diverse educational settings. The present study overcomes this limitation by applying a Bayesian multilevel meta-analytic approach to control for dependencies and decompose variance across sampling, within-study, and between-study levels (Assink & Wibbelink, 2016; Cheung, 2014). Unlike the conventional frequentist approaches used in previous syntheses, this approach allows for direct probability statements about effect magnitude and precision (e.g., Norouzian et al., 2018). Bayesian estimation may provide improved estimation accuracy and enhanced assessment of between-study heterogeneity, particularly when only a small number of studies are available (Reis et al., 2023). The Bayesian framework further allows for direct modelling of uncertainty through prior and posterior distributions, which is suitable for synthesising heterogeneous studies (Schmid & Mengersen, 2013; Sutton & Abrams, 2001). The present meta-analysis is guided by two research questions:

(1) What is the overall effectiveness of TP on L2 achievement?
(2) To what extent is the effectiveness of TP moderated by educational stage, target language, linguistic distance, intervention duration, and language skills?

3. Methodology

3.1. Literature identification and inclusion criteria

Literature identification was carried out following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Page et al., 2021). As bilingual researchers proficient in both English and Chinese, we searched English together with Chinese databases. They are Web of Science, Scopus, ERIC, ProQuest, and China National Knowledge Infrastructure (CNKI). The search for the literature began with Williams (1994), which is widely marked as one of the key contributions that brought translanguaging into scholarly attention. A three-stage search strategy was adopted in this study to guarantee comprehensive and systematic literature retrieval. Firstly, a database search was conducted based on two sets of keywords: one targeting language learning domains and the other focusing on translanguaging-related pedagogy. Within the English-language databases, the search string was: ('foreign language' OR 'second language' OR 'L2' OR 'FL' OR 'SL' OR 'additional language learning' OR 'language education') AND (translanguag* OR 'pedagogical translanguag*' OR 'translanguaging pedagogy' OR 'translanguaging practices' OR plurilingual*). 'AND' and 'OR' are Boolean operators that enable the combination of different terms. The asterisk (*) is a wildcard symbol, which means that it can be used to reflect lexical variants. To incorporate Chinese-language scholarship, equivalent keywords such as '超语', '超语言', '跨语言', '多语教学', and '超语言教学' were used in the CNKI. The term '超语' was used to capture studies on *translanguaging*. CNKI is the largest academic database in China, which indexes peer-reviewed journals, dissertations, and conference proceedings in Chinese. The language of the articles included in the search was limited to English and Chinese. The next search strategy involves the manual screening of high-impact journals in applied linguistics and language education, as shown in Table 1.

Moreover, backward reference searches were conducted on studies retained after full-text screening to identify earlier works not captured through database queries. Forward citation tracking via Google Scholar was also used to locate recent publications citing the same studies and meeting

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Table 1. List of journals selected for the analysis

Journal title	Index	Impact factor (JCR 2024)	Publisher
TESOL Quarterly	SSCI	3.9	Wiley
Applied Linguistics	SSCI	4.2	Oxford University Press
Language Teaching Research	SSCI	3.8	Sage
System	SSCI	5.6	Elsevier
Modern Language Journal	SSCI	4.0	Wiley
Journal of Second Language Writing	SSCI	4.5	Elsevier
International Journal of Bilingual Education and Bilingualism	SSCI/AHCI	2.5	Routledge (Taylor & Francis)
Linguistics and Education	SSCI/AHCI	2.1	Routledge (Taylor & Francis)
Language Learning	SSCI	4.2	Wiley
Language Teaching	SSCI/AHCI	5.1	Cambridge University Press
Journal of Multilingual and Multicultural Development	SSCI/AHCI	3.9	Routledge (Taylor & Francis)
International Journal of Multilingualism	SSCI/AHCI	2.0	Routledge (Taylor & Francis)
Language, Culture and Curriculum	SSCI/AHCI	2.2	Routledge (Taylor & Francis)
Language and Education	SSCI/AHCI	2.8	Routledge (Taylor & Francis)
Computer Assisted Language Learning	SSCI/AHCI	6.6	Routledge (Taylor & Francis)
Studies in Second Language Learning and Teaching	SSCI	5.4	Adam Mickiewicz University Press
RELC Journal	SSCI	6.2	Sage
ELT Journal	SSCI/AHCI	2.6	Oxford University Press
Journal of Language, Identity & Education	SSCI/AHCI	1.8	Routledge (Taylor & Francis)

Table 2. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
■ Reported original empirical findings on TP within L2 contexts	■ Did not include a clearly defined TP intervention
■ Employed experimental or quasi-experimental designs (e.g., treatment-control comparisons)	■ Conducted outside the context of second or foreign language education (e.g., mathematical education)
■ Targeted participants who were L2 learners	■ Used qualitative methodology
■ Published in English or Chinese	■ Duplicate reports based on the same data set
■ Provided adequate statistical information, such as means, standard deviations, sample sizes, or other statistics allowing for effect size calculation	■ Provided insufficient statistical information for calculating effect sizes

the inclusion criteria. We also included grey literature to reduce the risk of publication bias. This included unpublished master's theses, doctoral dissertations, and conference proceedings retrieved from dissertation and conference proceedings databases, such as ProQuest Dissertations and CNKI Dissertations. A total of 3,649 records were retained from the above searching process. After removing 762 duplicate records, we yielded 2,887 unique records. Following the assessment of titles, keywords, and abstracts, we excluded 2,835 records that did not meet the objectives of this study. A total of 52 full-text articles were retrieved and reviewed for eligibility using the inclusion and exclusion criteria listed in Table 2.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

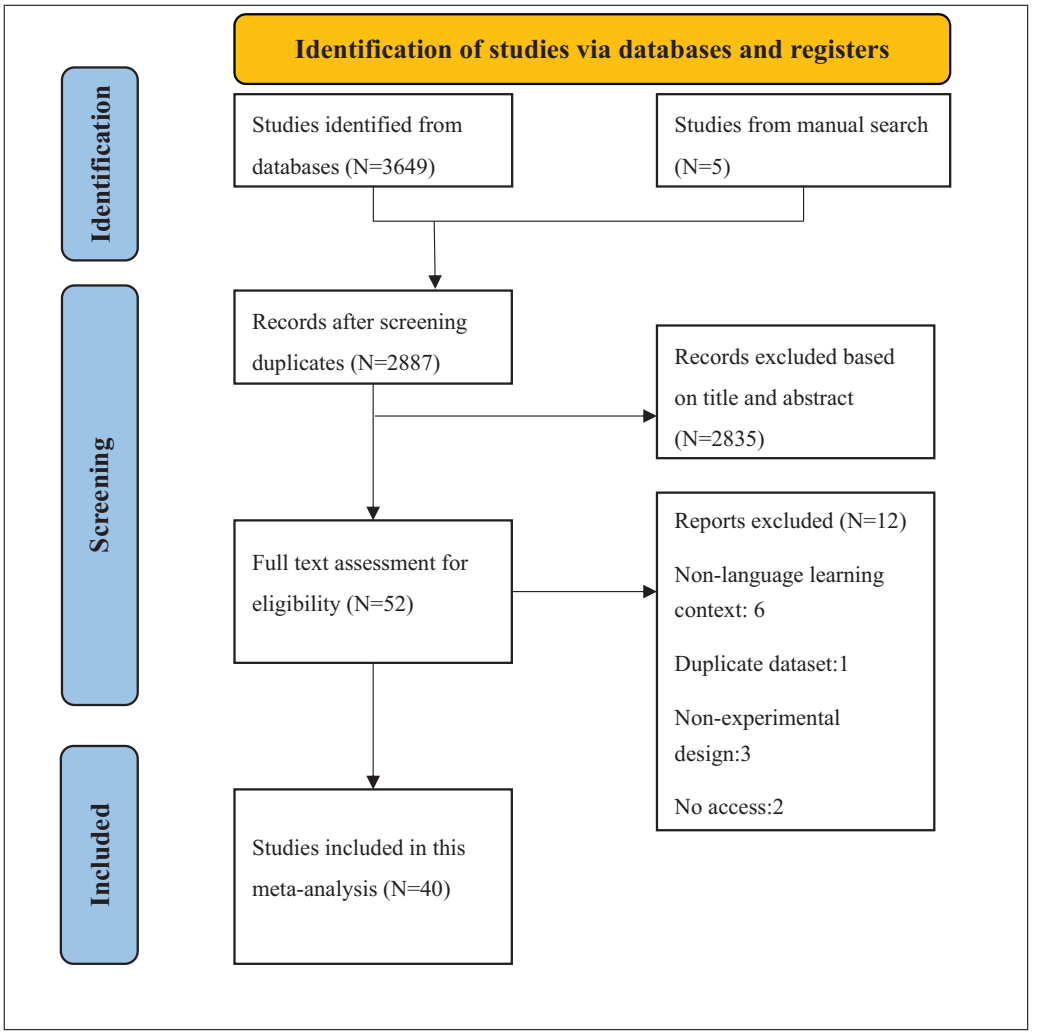


Figure 1. Flow diagram. *Source*: Adapted from Page et al. (2021).

We applied a set of eligibility criteria to determine whether each record was suitable for inclusion and excluded records that did not meet these criteria. A total of 40 studies were retained for the final analysis, comprising 34 journal articles, 3 Chinese master's theses, 2 English doctoral dissertations, and 1 book chapter (see Appendix A in the Supplementary Material). The flow diagram illustrating the study retrieval and screening process is presented in Figure 1.

3.2. Data extraction and effect size calculation

A data coding scheme was applied to extract bibliographic, sample, design, and effect size information from each eligible study (see Table 3). Two researchers independently coded all eligible studies after receiving training on the coding scheme. Inter-rater reliability was checked before the final coding results were obtained. For categorical variables, agreement between coders was measured using Cohen's kappa (k), and continuous variables were checked using intra-class correlation coefficients (see Supplementary Material Appendices B and C). Any differences between coders were

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Table 3. Coding scheme

Category	Variable	Operationalisation/description
Bibliographic information	Title	Study title
	Author(s)	Author name(s)
	Year of publication	Year in which the study was published
	Publication type	Journal article, dissertation, book chapter, conference proceeding, or others
Sample characteristics	Mean age	Average age of participants
	Sample size	Total number of participants included in the analysis
	Educational level	Primary, secondary, tertiary, or adult education
	First language	Participants' native language
Design features	Target language	Language being learned
	Study location	Country or region of data collection
	Linguistic distance	Typological similarity between the first language and the target language
	Intervention duration	Length of TP treatment in weeks
	Targeted language skills	Speaking, listening, reading, writing, grammar, vocabulary, or integrated skills
Effect size information	Hedges' *g*	Standardised mean difference adjusted for small-sample bias
	Standard error	Standard error associated with each effect size
	Variance	Sampling variance of each effect size

settled through discussion until full agreement was reached. Because many studies reported more than one effect size (e.g., multiple language skills, outcome measures, or assessment time points), we used a three-level meta-analytic model to account for the statistical dependence. This approach nested the data structure at three levels: sampling variance (Level 1), within-study variance (Level 2), and between-study variance (Level 3) (Assink & Wibbelink, 2016; Cheung, 2014). Hedges' *g* serves as the primary effect size measure, as it corrects for small-sample bias and provides less biased estimates of intervention effects (Hedges, 1981). We obtained 108 effect sizes from 40 eligible studies with 3,145 participants.

3.3. Statistical analysis

The multilevel Bayesian meta-analysis was conducted in RStudio 4.4.1 under the brms (Bürkner, 2017) and rstan (Stan Development Team, 2019) packages. The Bayesian framework integrates prior knowledge with observed data and estimates effect sizes through posterior distributions, which tend to be more robust than those produced by frequentist methods (Schmid & Mengersen, 2013; Sutton & Abrams, 2001).

The Bayesian framework is often recommended in meta-analyses with few included studies, as conventional estimators of between-study variance may yield boundary estimates of zero and thereby distort inference for the overall effect (Williams et al., 2018). The Bayesian framework yields a posterior distribution, and uncertainty is represented by highest-density credible intervals (CrIs), which allow researchers to quantify the probability that the parameter falls within a specified range (Kruschke & Liddell, 2018; Norouzian et al., 2018). This approach accounts for uncertainty in three steps: (a) specifying a prior distribution, which may come from external sources, (b) updating this

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

distribution based on observed data, and (c) deriving the posterior distribution via Bayes' theorem (Schmid & Mengersen, 2013). These properties have been demonstrated in recent applications of Bayesian methods in applied linguistics research synthesis (e.g., Brown et al., 2023; Norouzian et al., 2018).

Given that multiple effect sizes were nested within individual studies, we compared a two-level model and a three-level model to determine the optimal structure for handling dependent effect sizes. Model comparison was conducted using leave-one-out cross-validation information criterion (LOOIC) and the widely applicable information criterion (WAIC) as implemented in the loo package (Vehtari et al., 2023, 2017). A lower WAIC and LOOIC indicate a better-fitting model.

Because the included studies were heterogeneous, and following the previous Bayesian meta-analyses, such as Griffin and Oswald (2022), we placed a weakly informative Normal (0, 5) prior distribution for the overall effect size to reflect uncertainty regarding its magnitude and direction of TP's effectiveness. For estimating variance parameters, we employed a non-negative half-Cauchy distribution (Cauchy $\sim$ [0, 0.3]), as this distribution ensures non-negative estimates and is commonly used in psychological and educational meta-analyses due to its favourable properties for modelling variability (Williams et al., 2018). For the moderator analysis, we specified weakly informative Normal (0, 5) priors to stabilise estimation as well. All models were fit via Hamiltonian Monte Carlo (HMC) sampling implemented in the brms package (Bürkner, 2017), using four chains and 20,000 iterations (including 5,000 warm-up samples). Stan's adaptation period is specified as a target average proposal acceptance probability of 0.99 and a maximum tree depth of 12 (Griffin & Oswald, 2022). Before examining the parameter estimates, we ensured the HMC simulation exhibited reasonable properties by visually inspecting trace plots, evaluating chain convergence through the potential scale reduction factor ($\widehat{R}$), and confirming that the effective sample size for each parameter was sufficiently large. For model interpretability, trace plots should show good mixing (i.e., well-overlapping chains), while $\widehat{R}$ values close to 1 (Gelman & Rubin, 1992), and sufficiently large effective sample sizes were taken as indicators of satisfactory convergence (Griffin & Oswald, 2022).

Heterogeneity was tested with the I^2 statistic (Higgins & Thompson, 2002). The total, between-study, and within-study variance components were estimated at the three levels. Categorical moderators were translated into dummy indicators so that differences between subgroups could be examined. We calculated the posterior probability of a positive or negative difference for each comparison. For continuous moderators, meta-regression coefficients were estimated, and posterior probabilities exceeding 0.95 were taken as evidence for credible effects (Griffin & Oswald, 2022). For publication bias analysis, we used the PublicationBias package in RStudio (Braginsky et al., 2023; Mathur & VanderWeele, 2020) for a sensitivity analysis framework that does not rely on funnel plot symmetry. This method gauges how strong publication bias would need to be to bring the observed effect, or its CrI, down to a given threshold. Assuming that statistically significant positive results were more likely to be published (favor_positive = TRUE), we generated a significance-based funnel plot. This plot allows the comparison between affirmative and non-affirmative studies. Affirmative studies demonstrated significant positive outcomes, whereas non-affirmative investigations yielded non-significant or negative results. Using non-affirmative data alone, the worst-case estimate established a conservative robustness baseline. We also computed the *s*-value to identify how strong selection (η) would need to be to drive the effect to null, with 'not possible' indicating that bias cannot fully account for the findings.

In addition, we used traditional methods from the metafor package as a supplement to this method. Egger's regression test offers another test for possible publication bias (Egger et al., 1997). When Egger's test indicated publication bias ($p < 0.05$), the trim-and-fill procedure was applied to estimate missing studies and adjust the pooled effect size. Leave-one-out sensitivity analysis was conducted at both the study level and the effect size level to examine whether removing individual studies or specific effect sizes would substantially alter overall results.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

4. Results

4.1. Overall effectiveness

This synthesis brings together evidence from 40 studies ($k = 108$ effect sizes; $N = 3{,}145$ learners). Before pooling effect sizes, we conducted a model comparison to evaluate the suitability of each model. The results favoured the three-level Bayesian meta-analytic model, which showed a lower LOOIC (113.1, SE = 20.8) than the two-level model (124.5, SE = 26.9). WAIC results also preferred the three-level model (75.7, SE = 17.0) over the two-level model (97.1, SE = 24.1). Trace plots indicated consistent chain mixing with highly overlapping paths. Potential scale reduction factor values ranged from 1.00 to 1.002, and effective sample sizes were above 1,000 (see Supplementary Material Appendix D). These findings showed that the HMC produced reliable posterior estimates (Bürkner, 2017; Gelman & Rubin, 1992). The results show a positive effect of TP on L2 achievement (Hedges' $g = 1.014$, SE = 0.138, 95% CrI [0.748, 1.291]). Most of the variation came from differences between studies (85.8%), while smaller portions were due to sampling error (10.8%) and differences within studies (3.3%). The related heterogeneity estimates were τ_within = 0.157 (95% CrI [0.057, 0.243]) and τ_between = 0.794 (95% CrI [0.610, 1.054]). The posterior probability that the effect size exceeded zero was close to 100%, which supports a positive effect of TP. Figure 2 shows the posterior distributions of effect sizes for individual studies together with the overall pooled effect. The horizontal axis displays the size of the effect (Hedges' g), and the dashed vertical line marks the zero-effect point. The vertical axis lists the studies included in the meta-analysis by study identifier, with the overall effect shown separately at the bottom to distinguish it from the individual study estimates. Each row represents one study, and the ridge-shaped curve shows the posterior distribution of that study's effect size. Narrower curves indicate more stable estimates, while wider curves reflect greater uncertainty. Dots placed on the curves show the effect size estimates for each study.

4.2. Publication bias and sensitivity analysis

The funnel plots produced from the PublicationBias and metafor packages are shown in Figures 3 and 4, respectively. A multilevel Egger's test detected clear asymmetry in the data ($p < 0.0001$), and the result points to the possibility of publication bias. The significance funnel plot generated by the PublicationBias package (Figure 3) showed that studies with non-affirmative results, such as non-significant or negative findings, outnumbered studies reporting significant positive effects. The grey diamond, which summarises estimates based only on non-affirmative studies, closely overlapped with the black diamond that represents all included studies. This suggests that the overall effect of TP is relatively stable even in the presence of possible selection bias. In addition, the s-value analysis indicated no severe degree of publication bias (s-value = not possible).

Traditional funnel plot analysis conducted via the metafor package (Figure 4) provided converging evidence. The trim-and-fill analysis imputed zero studies, indicating no clear asymmetry requiring adjustment. Next, leave-one-out sensitivity analysis at both the study level and effect size level demonstrated that the overall results remained stable across all iterations. At the effect size level, removing individual effect sizes yielded estimates ranging from 0.958 to 1.043, while at the study level, removing individual investigations produced estimates ranging from 0.958 to 1.047. All pooled estimates remained statistically significant during at each removal, indicating that no single data point disproportionately influenced the synthesised result.

4.3. Moderator analysis

Considering the moderating role of educational level (see Figure 5), the posterior estimates indicated that the tertiary group tended to show a larger estimated effect relative to other educational levels (Hedges' $g = 1.326$, SE = 0.188, 95% CrI [0.962, 1.702], $k = 44$), followed by the secondary group (Hedges' $g = 1.165$, SE = 0.332, 95% CrI [0.509, 1.815], $k = 7$) and the primary group (Hedges'

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

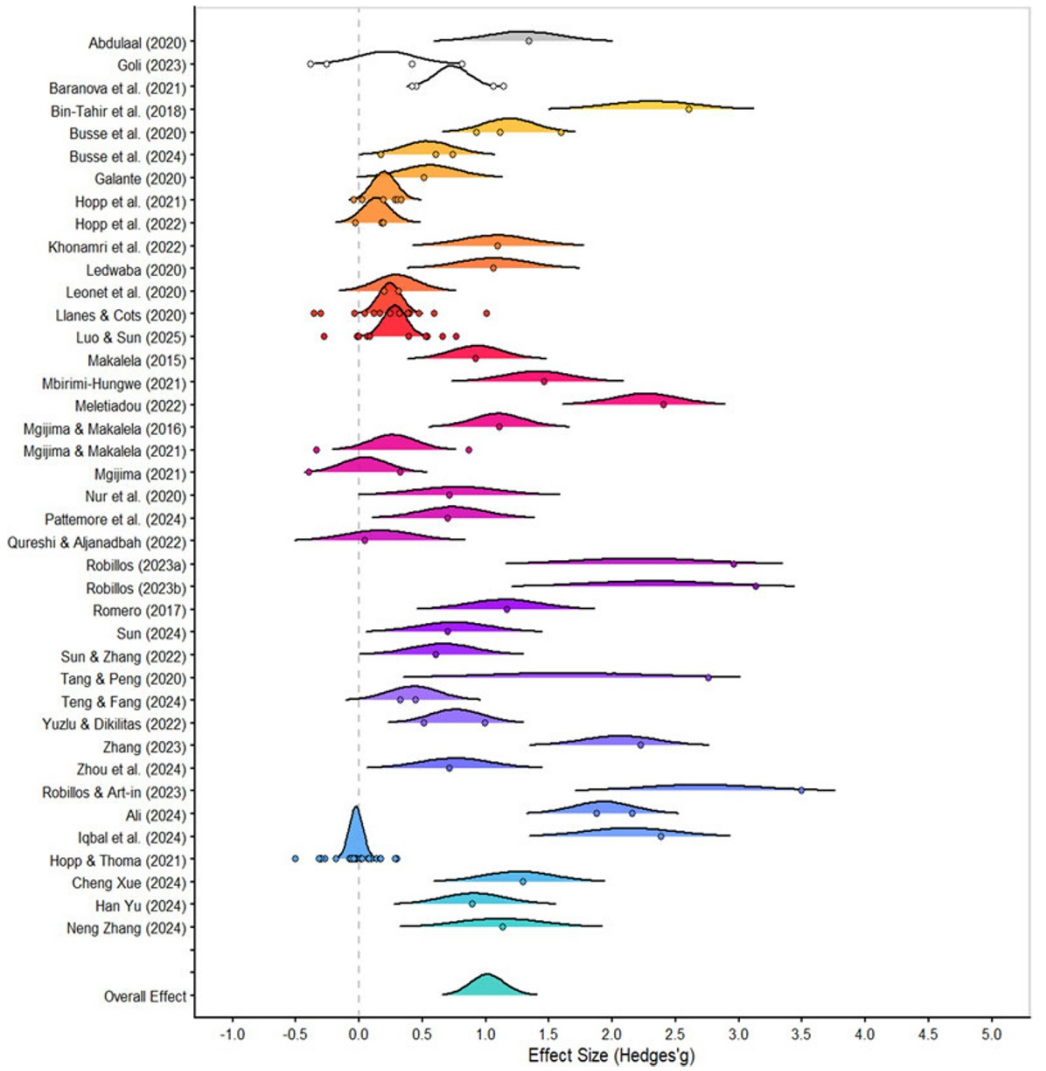


Figure 2. Forest plot of study-level posterior distributions ($N = 40$).

$g = 0.602$, SE $= 0.220$, 95% CrI [0.178, 1.042], $k = 51$). By contrast, the adult group demonstrated the smallest and non-credible effect (Hedges' $g = 0.263$, SE $= 0.549$, 95% CrI [−0.824, 1.344], $k = 6$). Pairwise contrasts showed that the difference between tertiary and primary education reached a posterior probability exceeding the 0.95 benchmark (99.3%), which provides credible evidence for a stronger effect in the tertiary group. While other contrasts showed different effect sizes, they did not reach the 95% thresholds for credible inference. For the target language (English vs language other than English, LOTE), the results yielded comparable effects for both groups. The effect size for the English group (Hedges' $g = 1.010$, SE $= 0.148$, 95% CrI [0.726, 1.309], $k = 100$) was slightly lower than that of the LOTE group (Hedges' $g = 1.064$, SE $= 0.432$, 95% CrI [0.211, 1.916], $k = 8$). The probability that the effect for the English group was greater than that for the LOTE group was 45.2%, thus providing weak evidence for a credible difference between the two groups. The comparison of the two groups is visually presented in Figure 6. Linguistic distance was also not a credible moderator. The regression coefficient was −0.998 (SE $= 1.540$, 95% CrI [−4.076, 2.029], $k = 44$). The

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

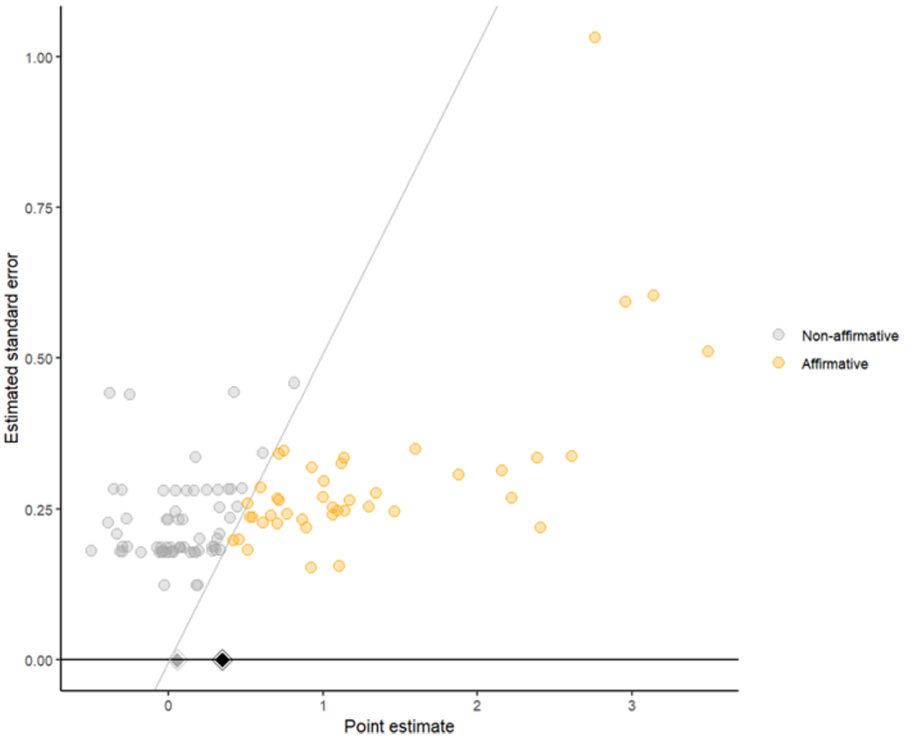


Figure 3. Funnel plot from the PublicationBias package.

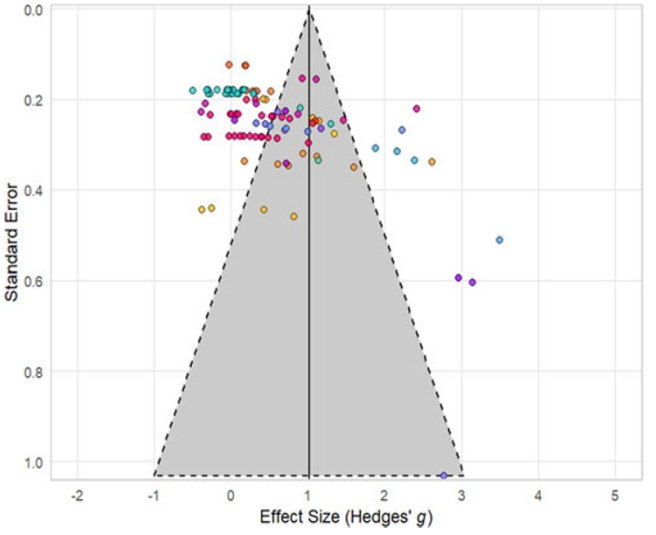


Figure 4. Funnel plot from the metafor package.

posterior probability of a negative association was 74.7%, which did not reach the 95% threshold for credible inference (Figure 7). For treatment duration, the estimated regression coefficient was −0.017 (SE = 0.022, 95% CrI [−0.060, 0.027], k = 86), with a posterior probability of a negative association of 78.3% (Figure 8). Moreover, the effects of TP varied across language skill domains. Speaking was

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

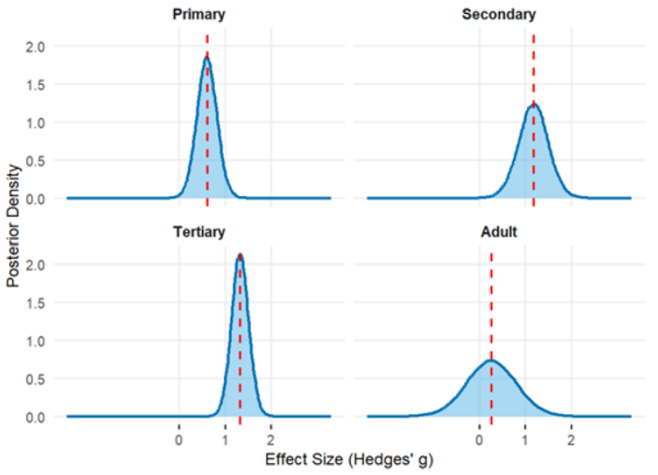


Figure 5. Posterior distributions of effect sizes across educational levels.

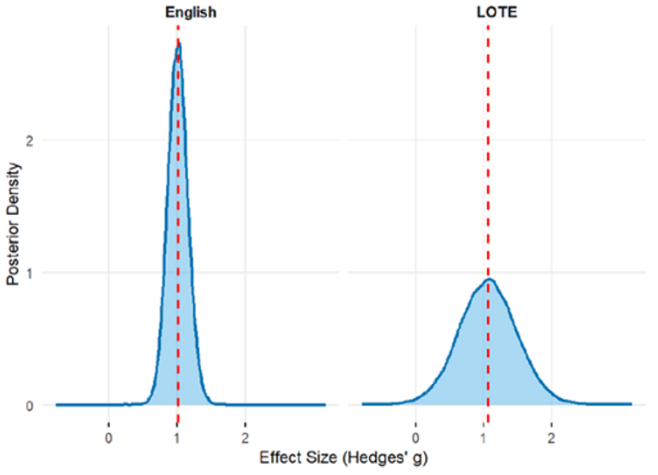


Figure 6. Posterior distributions of effect sizes across the target language.

associated with the largest estimated effect (Hedges' $g = 1.364$, SE = 0.234, 95% CrI [0.908, 1.825], $k = 19$), followed by reading (Hedges' $g = 1.143$, SE = 0.229, 95% CrI [0.704, 1.602], $k = 15$), writing (Hedges' $g = 1.084$, SE = 0.218, 95% CrI [0.657, 1.516], $k = 16$), and listening (Hedges' $g = 1.061$, SE = 0.300, 95% CrI [0.477, 1.654], $k = 4$). Grammar showed a moderate effect (Hedges' $g = 0.826$, SE = 0.259, 95% CrI [0.316, 1.339], $k = 29$), while vocabulary exhibited the smallest effect (Hedges' $g = 0.693$, SE = 0.236, 95% CrI [0.232, 1.162], $k = 16$). The general language outcome also demonstrated a positive effect (Hedges' $g = 0.919$, SE = 0.271, 95% CrI [0.393, 1.461], $k = 9$). Among the pairwise contrasts, two comparisons reached the conventional posterior probability threshold of 0.95. The effect of speaking was credibly greater than that of vocabulary (97.8%) and writing (96.0%). A visual representation of the effect size distribution across language skills is presented in Figure 9.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

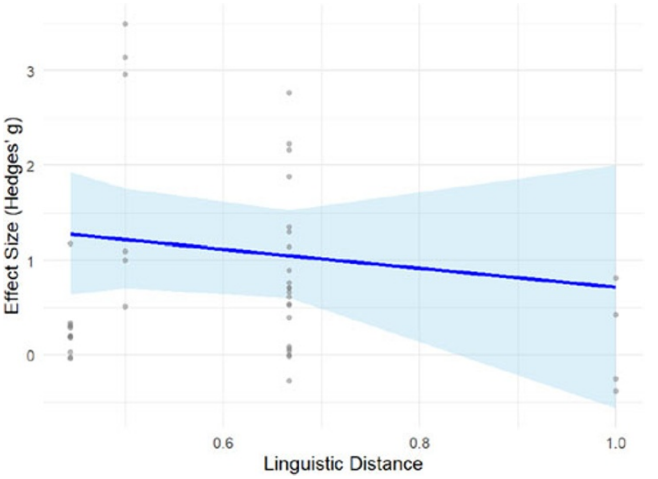


Figure 7. Posterior estimates of effect sizes across linguistic distance. Shaded areas represent 95% CrIs.

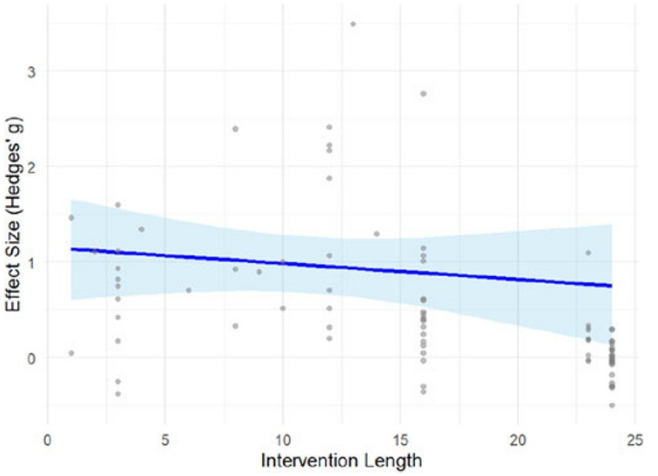


Figure 8. Posterior estimates of effect sizes across treatment length. Shaded areas represent 95% CrIs.

5. Discussion

5.1. The overall effectiveness of TP

We combined 108 effect sizes from 40 independent studies and found a strong positive effect in the multilevel Bayesian meta-analysis ($g = 1.014$). These findings suggest that TP has a positive effect on language development across the educational settings and learner groups represented in this synthesis. The $\widehat{R}$ values were 1.00, and effective sample sizes were above 1,000, showing reliable posterior estimation in the HMC simulations (Griffin & Oswald, 2022). Our findings are consistent with theoretical views that regarded translanguaging as a teaching response to the structural limits imposed by monolingual language ideologies (Li, 2018). They also reflect the 'multilingual turn' in language education (May, 2014) and substantiates the rationality of viewing learners' full linguistic resources as learning foundations. When learners are encouraged to use their entire linguistic and communication repertoires, translanguaging supports the building of new linguistic knowledge with less cognitive effort (Cenoz & Gorter, 2017, 2020; García & Li, 2015; Luo & Sun, 2025). This movement

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

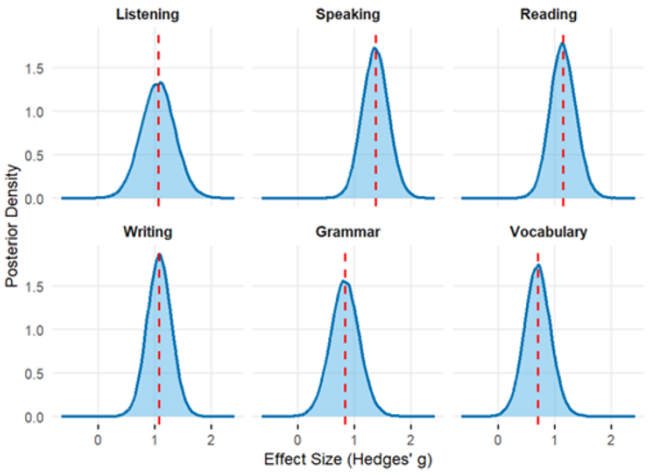


Figure 9. Posterior estimates of effect sizes across language skills. Shaded areas represent 95% CrIs.

among linguistic repertoires supports both linguistic development and learner agency in multilingual environments. These teaching effects match the findings of Yasar Yuzlu and Dikilitas (2022), who showed that students taught through translanguaging-based approaches performed better than those taught through traditional methods in the development of language skills. The benefits of TP fit within sociocultural models of language learning that view meaning construction as interactional and mediated by cultural and semiotic tools (Lantolf, 2000; Vygotsky, 1978). The synthesised evidence establishes TP as a credible instructional practice, offering empirical grounding for existing theoretical discussions of its pedagogical implications.

5.2. The interpretation of each moderator of TP

The moderator analysis for education level showed the strongest effects in tertiary education (Hedges' $g = 1.326$), followed by secondary ($g = 1.165$) and primary education ($g = 0.602$). In contrast, the smallest effect was observed in adult education ($g = 0.263$). The contrast between tertiary and primary groups returned a posterior probability above 95%, which provides credible evidence that the effect in tertiary education is stronger than that in primary education. The stronger effect in tertiary settings may be due to instructional contexts that allow for more flexibility in language use and provide opportunities for learners to engage with discipline-related content with different linguistic resources. This is exemplified in a study of Zhou et al. (2025). Zhou and her colleagues found that in a translanguaging-multiliteracies learning program, university students tend to assess multiple linguistic and semiotic resources and exhibit more agency and active participation in meaning-making practices. In contrast, the small effect sizes in primary education and adult education may reflect developmental and institutional constraints. For primary education, the limited effectiveness may be attributed to the diverse home language competences among learners and the absence of language peers for some students (Foster et al., 2023). Unless sufficient support and gradual exposure are provided, pupils in this age group are likely to struggle to engage actively with multilingual practices. In adult education, Teng and Fang's (2024) study using TP to promote morphological awareness among Japanese adult learners of Chinese revealed a mixed response. Some of them are engaging productively with the multilingual approach, but others reported confusion towards multilingual learning environments and a preference for monolingual instruction. The limited influence in adult learning

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

contexts implies that TP may be less effective in informal contexts than in formal school-based settings. Therefore, we call for more attention to explore how distinctive curriculum designs, teacher styles, and institutional conditions influence both the acceptance and the effectiveness of TP.

The moderator analysis showed that neither the target language nor the linguistic distance affected the effectiveness of TP. Learners demonstrated comparable gains regardless of whether the target language was English (Hedges' $g = 1.010$) or a non-English language (Hedges' $g = 1.064$). Similarly, the slope for linguistic distance was negative but not credibly different from zero ($b = -0.998$, SE = 1.540), which means that typological distance did not condition the TP's effectiveness. These results challenge the assumption that linguistic proximity serves as a reliable pathway to efficient language development (e.g., Chiswick & Miller, 2005). One interpretation is that TP works across language differences because learning through translanguaging is organised around meaning rather than language boundaries (Li, 2011). Within such a space, learners are positioned to make use of their full linguistic repertoires in a fluid manner, unbound by fixed language distinctions (Li, 2011). Another possible explanation for this finding may lie in the measurement of linguistic distance. Our study adopted the scale developed by Chiswick and Miller (2005), which is based on the difficulty English speakers face in learning other languages. When using this scale, we did not consider the language pairs in which the target language was not English. Although only four studies in our sample focused on non-English target languages, it is possible that the linguistic distance scale based on English may have obscured the possible moderating effects of those studies. We therefore recommend that future research develop linguistic distance measures that cover a wider range of target languages and explore how typological distance influences pedagogical outcomes.

The posterior estimate of the regression coefficient for intervention duration was −0.017 (SE = 0.022), with a posterior probability of a negative association being 78.3%. Although the results showed a weak trend towards smaller effects for longer interventions, the association is not statistically credible. This finding partially supports the empirical findings from Sun and Zhang (2022), who found that the effectiveness of TP in online peer feedback was confined to the initial phase of L2 writing instruction. This suggests that extending implementation duration alone may not be sufficient to sustain continuous learning gains from TP. Given the inconclusive evidence, future research could benefit from adopting a dosage–response perspective. For example, this would involve looking at how long TP is used, along with how often and how consistently it appears in classroom instruction. This may help move beyond one-time measures to better understand how TP works in practice.

For target language skills, the speaking skill had the largest effect size (Hedges' $g = 1.364$) and was credibly greater than vocabulary ($g = 0.693$) and writing ($g = 1.084$), with posterior probabilities above the 0.95 threshold. This is consistent with the empirical results that TP may be more effective for speaking development. For instance, in Baranova et al.'s (2021) work for teaching foreign language students Spanish through a translingual model, statistically significant gains were observed across all four language skills (listening, speaking, reading, and writing), with the gains for speaking being the greatest. The large effect size for speaking may be explained by the overlap between the nature of oral communication and the operational logic of TP. For example, speaking tasks are dialogic, interactive, and immediate, and learners are required to negotiate meaning-making in real-time interaction. TP may be well-suited to the demands of oral skills because it matches the immediacy, interactivity, and meaning-negotiation required in speaking tasks. Although the gains for other skills were not as large as those for speaking, the positive consistency across skills suggests that TP supports development in both productive and receptive skills. However, vocabulary ($g = 0.693$) and grammar ($g = 0.826$) showed smaller effect sizes. This difference may be attributed to the cognitive and instructional differences. These two skill types are often taught as discrete units, focusing on decontextualised patterns such as word lists, morphological rules, or syntactic structures. Such components of language tend to require accuracy, memorisation, and rule-based manipulation. It may not fully align with the demands of the meaning-focused and learner-driven nature of translanguaging.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

6. Implications

The large, pooled effect size ($g = 1.014$) provides empirical support for moving beyond traditional SLA models that treat languages as separate systems, showing that language development benefits when learners can draw on their full linguistic resources. Pedagogically, teachers can ask students about their home languages, literacy backgrounds, and daily language use at the beginning of a course. This provides information about what the language resources students bring to the classroom, which can then inform instructional design that fits the students in that particular class. For example, students can develop ideas in their native language before expressing them in the target language. They may also be permitted to annotate texts using multiple languages to support meaning-making and learning. The construction of multilingual glossaries can also help students relate new vocabulary to familiar words from their linguistic repertoires. Methodologically, this study illustrates the strengths of Bayesian meta-analysis in handling the complexity and uncertainty in applied linguistics research. With the growing use of more complex data in applied linguistics, Bayesian meta-analysis offers a powerful statistical framework for synthesising findings across studies.

7. Conclusions

This study represents an attempt to apply multilevel Bayesian meta-analysis in exploring the effectiveness of TP within the field of SLA. The 40 empirical studies, including 108 effect sizes, returned a positive pooled effect of TP on L2 achievement (Hedges' $g = 1.014$, SE = 0.138, 95% CrI [0.748, 1.291]). Speaking was associated with the largest effect size across language skill domains, while effect sizes were larger among tertiary-level learners than among other educational levels. Subsequent analysis found limited moderating evidence from target language, treatment duration, or linguistic distance. These findings reinforce the pedagogical value of translanguaging as a context-sensitive approach to language instruction that reflects the linguistic realities of multilingual classrooms. In addition, the study illustrates how Bayesian hierarchical modelling can be applied to provide more reliable heterogeneity estimates and probabilistic interpretations of effect sizes, offering methodological insights for future quantitative work in SLA. This study intends to advance our understanding of how and when TP benefits L2 learning and provides an empirical basis for the targeted use of TP in language classrooms where learner backgrounds and educational contexts vary.

Despite the increasing use of technology in TP, the technology type was not examined as a moderator because scarce studies tested the same digital tools. An initial step would be to examine how different technologies, such as generative artificial intelligence (GenAI), are integrated into TP, which would increase the number of primary studies.. With more studies focusing on comparable forms of technology-meidated instruction, it would then be possible to compare different technology integration patterns using network meta-analysis.. Of the 40 studies included in this meta-analysis, only 3 used delayed post-tests, and most assessed learning outcomes immediately after instruction. This pattern makes it difficult to tell whether the observed gains endure beyond the immediate task or fade once instruction ends. An important direction for future research is the inclusion of delayed assessments several weeks or months after instruction to evaluate the durability of TP's effectiveness. Another limitation is that the analysis relied on quantitative effect sizes and did not incorporate qualitative insights. This limitation may be addressed through integrative mixed methods meta-analysis (Levitt, 2024), which combines quantitative effect estimation with qualitative analysis. If this approach is applied in future work, it might become clearer how learners experience, accept, and deal with the use of TP across instructional contexts.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0261444826101335.

Funding statement. This research did not receive any specific grant.

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Author contributions. The authors have contributed equally to the manuscript, with Lawrence Jun Zhang serving as the corresponding author.

Competing interests. The authors declare that there is no conflict of interest.

References

Abourehab, Y., & Azaz, M. (2023). 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

Assink, M., & Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Baker, C. (2001). *Foundations of bilingual education and bilingualism*. Multilingual Matters.

Baranova, T., Kobicheva, A., Tokareva, E., & Vorontsova, E. (2021). Application of translinguism in teaching foreign languages to students (specialty "Ecology"). In E3S Web of Conferences, (Vol. 244, Article 11034). EDP Sciences. https://doi.org/10.1051/e3sconf/202124411034

Blackledge, A., & Creese, A. (2014). *Heteroglossia as practice and pedagogy*. Springer Netherlands.

Braginsky, M., Mathur, M., & VanderWeele, T. J. (2023). *PublicationBias: Sensitivity analysis for publication bias in meta-analyses* (Version 2.4.0) [R package]. CRAN. https://doi.org/10.32614/CRAN.package.PublicationBias

Brown, D., Liu, Q., & Norouzian, R. (2023). Effectiveness of written corrective feedback in developing L2 accuracy: A Bayesian meta-analysis. *Language Teaching Research*, *30*(3), 1357–1389. https://doi.org/10.1177/13621688221147374

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

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

Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *Modern Language Journal*, *95*(3), 401–417. https://doi.org/10.1111/j.1540-4781.2011.01207.x

Cenoz, J., & Gorter, D. (2017). Minority languages and sustainable translanguaging: Threat or opportunity? *Journal of Multilingual and Multicultural Development*, *38*(10), 901–912. https://doi.org/10.1080/01434632.2017.1284855

Cenoz, J., & Gorter, D. (2020). Pedagogical translanguaging: An introduction. *System*, *92*, 102269. https://doi.org/10.1016/j.system.2020.102269

Cenoz, J., & Gorter, D. (2021). *Pedagogical translanguaging*. Cambridge University Press.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Chiswick, B. R., & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, *26*(1), 1–11. https://doi.org/10.1080/14790710508668395

Cummins, J. (2019). The emergence of translanguaging pedagogy: A dialogue between theory and practice. *Journal of Multilingual Education Research*, *9*(13), 19–36. https://fordham.bepress.com/jmer/vol9/iss1/13

Dryden, S., Tankosić, A., & Dovchin, S. (2021). Foreign language anxiety and translanguaging as an emotional safe space: Migrant English as a foreign language learners in Australia. *System*, *101*, 102593. https://doi.org/10.1016/j.system.2021.102593

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fisher, L., Evans, M., Forbes, K., Gayton, A., & Liu, Y. (2020). Participative multilingual identity construction in the languages classroom: A multi-theoretical conceptualisation. *International Journal of Multilingualism*, *17*(4), 448–466. https://doi.org/10.1080/14790718.2018.1524896

Foster, N., Auger, N., & Van Avermaet, P. (2023). Multilingual tasks as a springboard for transversal practice: Teachers' decisions and dilemmas in a Functional Multilingual Learning approach. *Language and Education*, *37*(1), 22–38. https://doi.org/10.1080/09500782.2021.1958835

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

García, O., Johnson, S. I., Seltzer, K., & Valdés, G. (2017). *The translanguaging classroom: Leveraging student bilingualism for learning*. Caslon.

García, O., & Li, W. (2015). Translanguaging, bilingualism, and bilingual education. In W. Wright, S. Boun & O. García (Eds.), *The handbook of bilingual and multilingual education* (pp. 223–240). Wiley-Blackwell.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Griffin, J. W., & Oswald, F. (2022). A multilevel Bayesian meta-analysis of the body inversion effect: Evaluating controversies over headless and sexualized bodies. *Psychonomic Bulletin and Review*, *29*(5), 1558–1593. https://doi.org/10.3758/s13423-022-02067-3

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

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

Hopp, H., & Thoma, D. (2021). Effects of plurilingual teaching on grammatical development in early foreign-language learning. *Modern Language Journal*, *105*(2), 464–483. https://doi.org/10.1111/modl.12709

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin and Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Lantolf, J. (2000). *Sociocultural theory and language learning*. Oxford University Press.

Lee, C. (2024). How do children go through a heteroglossic path to becoming bilingual? Comparison of Korean children's translanguaging performance in first and third grade classrooms. *International Journal of Bilingual Education and Bilingualism*, *27*(4), 612–630. https://doi.org/10.1080/13670050.2023.2232088

Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.

Levitt, H. M. (2024). How to conduct an integrative mixed methods meta-analysis: A tutorial for the systematic review of quantitative and qualitative evidence. *Psychological Methods*. https://doi.org/10.1037/met0000675

Li, W. (2011). Moment analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics*, *43*(5), 1222–1235. https://doi.org/10.1016/j.pragma.2010.07.035

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, *39*(1), 9–30. https://doi.org/10.1093/applin/amx044

Liao, S. J., Fang, F., & Zhang, L. J. (2025). Evaluating methodological features of research on translanguaging pedagogy in English medium instruction (EMI). *Research Methods in Applied Linguistics*, *4*, 100263. https://doi.org/10.1016/j.rmal.2025.100263

Liu, Y., & Fang, F. (2022). Translanguaging theory and practice: How stakeholders perceive translanguaging as a practical theory of language. *RELC Journal*, *53*(2), 391–399. https://doi.org/10.1177/0033688220939222

Luo, X., & Sun, P. P. (2025). Greater complexity and higher efficiency: Effects of translanguaging during collaborative pre-task planning on EFL learners' speaking performance. *Language Teaching Research*, 13621688241307029. https://doi.org/10.1177/13621688241307029

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *69*(5), 1091–1119. https://doi.org/10.1111/rssc.12440

May, S. (2014). *The multilingual turn: Implications for SLA, TESOL, and bilingual education*. Routledge.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114–135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

Mgijima, V. D. (2021). Advancing text prediction skills through translanguaging. *Reading & Writing*, *12*(1), 1–10. https://doi.org/10.4102/rw.v12i1.284

Nagashima, Y., & Lawrence, L. (2022). To translanguage or not to translanguage: Ideology, practice, and intersectional identities. *Applied Linguistics Review*, *13*(5), 735–754. https://doi.org/10.1515/applirev-2019-0040

Norouzian, R., De Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, *68*(4), 1032–1075. https://doi.org/10.1111/lang.12310

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qureshi, M. A., & Aljanadbah, A. (2022). Translanguaging and reading comprehension in a second language. *International Multilingual Research Journal*, *16*(4), 247–257. https://doi.org/10.1080/19313152.2021.2009158

Rajendram, S. (2023). Translanguaging as an agentive pedagogy for multilingual learners: Affordances and constraints. *International Journal of Multilingualism*, *20*(2), 595–622. https://doi.org/10.1080/14790718.2021.1898619

Reis, D. J., Kaizer, A. M., Kinney, A. R., Bahraini, N. H., Holliday, R., Forster, J. E., & Brenner, L. A. (2023). A practical guide to random-effects Bayesian meta-analyses with application to the psychological trauma and suicide literature. *Psychological Trauma: Theory, Research, Practice, and Policy*, *15*(1), 121–130. https://doi.org/10.1037/tra0001316

Richards, J. C., & Wilson, O. (2019). On transidentitying. *RELC Journal*, *50*(1), 179–187. https://doi.org/10.1177/0033688218824780

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

Robillos, R. J. (2023). Exploring translanguaging during metacognitive strategy use on L2 listening and writing skills. *Journal of Language and Education*, *9*(3), 110–128. https://doi.org/10.17323/jle.2023.14329

Schmid, C. H., & Mengersen, K. (2013). Bayesian meta-analysis. In J. Koricheva, J. Gurevitch & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 174–191). Princeton University Press.

Song, K. (2016). "Okay, I will say in Korean and then in American": Translanguaging practices in bilingual homes. *Journal of Early Childhood Literacy*, *16*(1), 84–106. https://doi.org/10.1177/1468798414566705

Stan Development Team. (2019). *RStan: the R interface to Stan. R package version 2.19.2*. http://mc-stan.org/

Sun, L. (2024). Translanguaging pedagogy on the digital stage: Exploring Chinese undergraduates' English grammar learning through DingTalk platform. *Humanities and Social Sciences Communications*, *11*(1), 1–14. https://doi.org/10.1057/s41599-024-03771-2

Sun, P. P., & Zhang, L. J. (2022). Effects of translanguaging in online peer feedback on Chinese university English-as-a-foreign-language students' writing performance. *RELC Journal*, *53*(2), 325–341. https://doi.org/10.1177/00336882221089051

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, *10*(4), 277–303. https://doi.org/10.1177/096228020101000404

Teng, M. F., & Fang, F. (2024). Translanguaging pedagogies in developing morphological awareness: The case of Japanese students learning Chinese in China. *Applied Linguistics Review*, *15*(4), 1477–1506. https://doi.org/10.1515/applirev-2022-0138

Ticheloven, A., Blom, E., Leseman, P., & McMonagle, S. (2021). Translanguaging challenges in multilingual classrooms: Scholar, teacher and student perspectives. *International Journal of Multilingualism*, *18*(3), 491–514. https://doi.org/10.1080/14790718.2019.1686002

Ulum, Ö. G. (2024). Empowering voices: A deep dive into translanguaging perceptions among Turkish high school and university language learners. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e39557

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2023). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. https://mc-stan.org/loo/

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics & Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wawire, B. A., & Barnes-Story, A. (2023). Translanguaging for multiliteracy development: Pedagogical approaches for classroom practitioners. *International Journal of Bilingual Education and Bilingualism*, *26*(2), 173–188. https://doi.org/10.1080/13670050.2022.2094702

Wen, Z. E., Zhang, L. J., Kong, H., & Han, L. (2022). Translanguaging genre pedagogy: Implications for teaching business communication in the Greater Bay Area. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 1–15. https://doi.org/10.1186/s40862-022-00161-6

Williams, C. (1994). *Arfarniad o ddulliau dysgu ac addysgu yng nghyd-destun addysg uwchradd ddwyieithog*. Bangor University.

Williams, D. R., Rast, P., & Bürkner, P. (2018). Bayesian meta-analysis with weakly informative prior distributions. https://doi.org/10.31234/osf.io/7tbrm

Yasar Yuzlu, M., & Dikilitas, K. (2022). 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

Zainuddin, N. M., & Zaki, N. N. N. M. (2023). Comparison of translanguaging practices and ideologies in secondary and tertiary ESL settings in Malaysia: Teachers' perspective. *Higher Education and Oriental Studies*, *3*(1), 4–23. https://doi.org/10.54435/heos.v3i1.90

Zhang, L. (2023). Empowering Chinese college students in English as a foreign language writing classes: Translanguaging with translation methods. *Frontiers in Psychology*, *14*, 1118261. https://doi.org/10.3389/fpsyg.2023.1118261

Zhang, L. J. (2022a). Deepening the understanding of translanguaging as a practical theory of language: A conversation with Professor Li Wei. *RELC Journal*, *53*(3), 739–746. https://doi.org/10.1177/00336882221139894

Zhang, L. J. (2022b). Translingual practices and English as an additional language: A conversation with Professor Suresh Canagarajah. *RELC Journal*, *53*(2), 460–468. https://doi.org/10.1177/00336882221104512

Zhao, S., & Flewitt, R. (2020). Young Chinese immigrant children's language and literacy practices on social media: A translanguaging perspective. *Language and Education*, *34*(3), 267–285. https://doi.org/10.1080/09500782.2019.1656738

Zhou, J., Lee, S. S., & Kew, S. N. (2025). Developing and evaluating a translanguaging-multiliteracies learning design (LDTMP) for student agency in a college English course in China. *Frontiers in Education*, *9*, 1504871. https://doi.org/10.3389/feduc.2024.1504871

Zhou, X., Chen, X., & Wang, Z. (2024). The effect of linguistic choices in note-taking on academic listening performance: A pedagogical translanguaging perspective. *International Review of Applied Linguistics in Language Teaching*, *62*(2), 775–806. https://doi.org/10.1515/iral-2022-0127

Xiaoqi Wang is currently a full-time Ph.D. student in the Faculty of Arts and Education, University of Auckland, Auckland, New Zealand. His research interests include multilingualism, translanguaging, individual differences in language learning, GenAI for language education, and informal digital learning of English. He has published several empirical articles and reviews

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press

in leading SSCI-indexed journals such as *System, Interactive Learning Environments, Innovation in Language Learning and Teaching, Journal of Computer Assisted Learning*, and CSSCI-indexed journals, including *Modern Foreign Languages, Foreign Language World*, and *Foreign Language Research*, among others.

Lawrence Jun Zhang, Ph.D., is Professor of Applied Linguistics and Associate Dean, Faculty of Arts and Education, University of Auckland, Auckland, New Zealand. His research focuses on the psychology of language learning and teaching, including learner metacognition, L2 reading and writing development, motivation, teachers' use of AI, and assessment literacy. He has published widely in leading journals such as *Applied Linguistics, Modern Language Journal, TESOL Quarterly, Language Teaching, Language Teaching Research, Computer Assisted Language Learning, Computers & Education*, and *System*. He serves on several editorial boards and was Co-Editor-in-Chief of *System* until December 2024. He received the International TESOL Association's '50@50' Award in 2016 and has been consistently listed among the top 2% most-cited scholars in linguistics and language education.

Cite this article: Wang, X., & Zhang, L. J. (2026). A multilevel Bayesian meta-analysis of the effects of translanguaging pedagogy on L2 achievement. *Language Teaching*, 1–22. https://doi.org/10.1017/S0261444826101335

https://doi.org/10.1017/S0261444826101335 Published online by Cambridge University Press