

Research Synthesis in Applied Linguistics

ISSN: 2998-4475 (Print) 2998-4483 (Online) Journal homepage: www.tandfonline.com/journals/rsal20

The Effects of Different Approaches to Teaching Chinese Orthography on Character Acquisition and Reading Performance in Learners of Chinese as a Foreign Language: A Systematic Review

Sonia Qin & Hamish Chalmers

To cite this article: Sonia Qin & Hamish Chalmers (22 Aug 2026): The Effects of Different Approaches to Teaching Chinese Orthography on Character Acquisition and Reading Performance in Learners of Chinese as a Foreign Language: A Systematic Review, Research Synthesis in Applied Linguistics, DOI: 10.1080/29984475.2026.2716738

To link to this article: https://doi.org/10.1080/29984475.2026.2716738

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

View supplementary material

Published online: 22 Aug 2026.

Submit your article to this journal

Article views: 123

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rsal20

Research Synthesis in Applied Linguistics 2026

THE EFFECTS OF DIFFERENT APPROACHES TO TEACHING CHINESE ORTHOGRAPHY ON CHARACTER ACQUISITION AND READING PERFORMANCE IN LEARNERS OF CHINESE AS A FOREIGN LANGUAGE:

A SYSTEMATIC REVIEW

Sonia Qin, Hamish Chalmers



Key Terms

Chinese as a foreign language (CFL)
The teaching and learning of Chinese by students in contexts where Chinese is not the predominant language spoken or taught.

Mandarin
A spoken variety of Chinese, most commonly taught in formal educational settings.

Pinyin
The romanisation system used to represent Mandarin pronunciation using roman letters.

What can I learn?

Learn which approaches improve Chinese character learning and reading, and which are most effective at different learner proficiency levels.

What evidence is summarised?

30 STUDIES

Experimental and quasi-experimental studies evaluating the effects of different approaches to teaching Chinese orthography on literacy-related outcomes.

What does it find?

5 Results showed **five** categories of pedagogical approaches: **computer-assisted language learning**; **conventional (or traditional) techniques**; a **holistic curriculum** combining reading, writing, speaking, and listening; **various presentation methods** when introducing characters; and the **timing** of orthographic instruction.

Studies generally **preferred technology-enhanced methods**, a **holistic curriculum**, and **varied presentations** of character information over conventional techniques.

Findings on orthographic instruction timing were mixed, with some favoring immediate character teaching over a pinyin-first approach.

Participant samples **mainly included adult university learners**, primarily L1 English speakers. General **lack of consistency in the classification and reporting of CFL learner proficiency** across studies.

The evidence's **methodological quality was weak**, limiting causal inferences and strong pedagogical recommendations.

How can I use the findings?

Use Technology Purposefully

Use technology-enhanced teaching tools like online games, digital flashcards, and animated stroke-order demonstrations.

Teach in Meaningful Chunks

Teach characters in small groups, **focusing on radicals or semantic/phonetic components** to enhance pattern recognition and connections.

Introduce Characters Early

Introduce **characters early along with pinyin**, instead of pinyin first.

CC BY NC ND

RESEARCH SYNTHESIS IN APPLIED LINGUISTICS
https://doi.org/10.1080/29984475.2026.2716738

Routledge
Taylor & Francis Group

RESEARCH SYNTHESIS

OPEN ACCESS Check for updates

The Effects of Different Approaches to Teaching Chinese Orthography on Character Acquisition and Reading Performance in Learners of Chinese as a Foreign Language: A Systematic Review

Sonia Qin and Hamish Chalmers

Department of Education, University of Oxford, Oxford, UK

ABSTRACT

There is growing evidence investigating the effects of different instructional methods on character acquisition and reading performance in contexts where students are learning Chinese as a foreign language (CFL). However, comprehensive and transparent syntheses of the literature have been limited, with prior systematic reviews lacking quality appraisals and suffering from incomplete reporting. To address this gap, the current systematic review gathers and synthesises research on instructional approaches that have been empirically tested for effectiveness in CFL settings. Thirty eligible reports were identified through the search process, with data extracted for narrative synthesis and quality appraisal. Results showed five categories of pedagogical approaches: computer-assisted language learning; conventional techniques; holistic strategies (combining reading, writing, speaking, listening); various presentation methods when introducing characters; and the timing of orthographic teaching. Compared to conventional techniques, studies generally favoured technology-enhanced methods, holistic approaches, and different ways of presenting character information. Evidence regarding the timing of orthographic instruction was mixed, with some promoting immediate character teaching over a pinyin-first approach. Despite positive findings, the overall weak methodological quality of the evidence limits strong causal inferences and pedagogical recommendations. Future research must employ more robust and transparent study designs to support the development of evidence-informed CFL practice.

ARTICLE HISTORY
Received 18 February 2026
Revised 10 July 2026
Accepted 3 August 2026

KEYWORDS
Chinese as a foreign language; Chinese teaching and learning; Chinese characters; reading skills; orthographic acquisition

Introduction and Background

Recently, the teaching and learning of Chinese as a foreign language (CFL) has gained popularity among school- and university-aged students worldwide (British Council 2023; Gil 2024). Several factors have contributed to this popularity, including

CONTACT Hamish Chalmers hamish.chalmers@education.ox.ac.uk
Supplemental data for this article can be accessed online at https://doi.org/10.1080/29984475.2026.2716738.
© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

international migration; China's political and economic influence in recent decades; and the widespread use of Chinese varieties, such as Mandarin and Cantonese, as the lingua franca within the Chinese diaspora (Goh 2017). However, despite growing interest, many students encounter major linguistic hurdles when learning CFL. This is especially prevalent in Anglophone regions, where poor outcomes and slow acquisition have contributed to high attrition rates when pupils decide whether to pursue the language at more advanced levels (Yue 2017). Indeed, researchers have suggested that teaching and learning CFL offers a fundamentally distinct experience, especially for alphabetic first-language (L1) users, who are unaccustomed to the visual complexity of Chinese orthography (Han 2016; Zhao 2011). Consequently, it is critical to identify successful strategies in the current literature on CFL literacy instruction to better understand how practitioners can enhance CFL students' acquisition of the writing system and their reading performance. This study adopts a systematic review approach to present an evaluation of intervention research, examining the effects of different approaches to teaching Chinese orthography on school- and university-aged CFL learners' character acquisition and reading development. We begin by providing an overview of the writing system and its linguistic challenges, before analysing some popular CFL pedagogical strategies and previous systematic reviews.

The Writing System and Related Terminology

Chinese orthography is distinct from many alphabetic languages like English due to its logographic script. That is, instead of letters representing sounds, the writing system uses characters to convey meaning (H. C. Chen 1992). Each character contains various strokes and radicals, written in a prescribed order (Shen 2005). Strokes are the individual lines that shape each character, while radicals are the basic components of a character, which provide phonological and/or semantic cues (H. C. Chen 1992). A single radical can represent a whole (or integral) character, whereas combining two or more radicals creates a compound character.

Another factor to consider is that Chinese contains two different writing scripts: simplified and traditional. Simplified characters were introduced by the government of the People's Republic of China to reduce native speakers' illiteracy in China and enhance overall education (Snow 2013). While China has largely adopted the use of simplified characters since the 1950s, traditional characters remain the official script in Hong Kong, Taiwan, and Macau. CFL students typically learn simplified characters, as they are less visually complex and involve fewer strokes to produce and memorise (Ésik 2020).

In this review, the term 'Chinese' pertains to its orthography/writing system and is used more broadly when discussing CFL education. Meanwhile, 'Mandarin' refers to the spoken form most commonly taught in formal institutions globally (Goh 2017). Hanyu pinyin is used as the romanisation system for transcribing Mandarin. Examples of Chinese orthography are presented in simplified characters; however, when discussing experimental conditions in empirical studies, the specific script used will be clearly indicated. Moreover, although 'Hanzi' is increasingly used to refer to the writing system in Chinese language teaching and learning research, we use the term 'Chinese characters' throughout this review to maintain consistency with the search strategy and the terminology adopted in many of the included studies. Lastly, reading as an outcome can be

interpreted in a number of different ways. Broadly speaking, for the purposes of our review, 'reading' refers to the recognition of characters and/or the comprehension of written material, including words, sentences, or passages.

Linguistic Challenges

CFL learners must acquire over 3000 characters to become functionally literate in written Chinese (Norman 1988). This represents a heavy cognitive load on students. When compared to L1 English learners, native speakers of Mandarin undergo an additional two years of formal schooling to develop fundamental literacy skills (Lü and Zhang 1999). In fact, a large proportion of the Chinese national curriculum is dedicated to reading and writing: Chinese is a compulsory subject in the gaokao (equivalent to British A-levels), where it accounts for the highest weighting, alongside Mathematics and Science, in the exam (Fu 2024). The acquisition of characters is challenging for both native speakers and CFL learners. At the same time, orthographic knowledge is essential for achieving proficiency and assessment, making it crucial for students to master. For example, 55% of the total mark in a lower intermediate course at the University of Sydney (n.d.) is based on composition writing, and summative assessments on reading, writing, vocabulary, and grammar. Although educational contexts vary across regions, it is reasonable to assume that reading and writing characters occupy a substantial amount of both CFL students' and teachers' attention in the classroom. The cognitive effort involved in memorising such a large number of characters is echoed by British CFL secondary learners, who describe the task of recalling 'every single character' as nearly 'impossible' (Yang 2022, 8).

The sheer number of characters required for basic proficiency is further complicated by the fact that characters belong to several different structural categories. Many CFL beginner learners assume that most characters are pictographic, when in reality they are morpho-syllabic representations of meaning (Hu 2010; Huang 2022). Broadly speaking, characters can be grouped into the following categories (Boltz 1994):

1. Pictograms: visual representations of meaning
2. Ideographs: symbolic representations of abstract concepts
3. Ideographic compounds: a combination of pictograms
4. Phono-semantic compounds: contain both phonetic and semantic radicals
5. Phonetic loan words: borrowed from other languages (mainly English) that sound like their foreign counterparts; and
6. Derivative cognates: characters that share the same etymological root, but differ in pronunciation and meaning

These categories illustrate the inherent complexity of the writing system, complexity that is further compounded by the tonal nature of Chinese. Novice learners may find it difficult to differentiate the five tones of Mandarin and thus encounter a high number of apparent homophones, making it harder to identify the intended meaning during spoken interaction and the correct character when writing (Taylor and Taylor 2014). While Boltz's (1994) categorisation helps to illustrate the different ways in which characters can represent meaning, it should be noted that phonetic and semantic radicals do not always match the character's actual pronunciation or meaning (Honorof and Feldman

2006). Indeed, phonetic radicals are only 38% reliable in predicting the character's sound, excluding tone variations (Perfetti, Zhang, and Berent 1992). However, this falls to 26% when the tone of the phonetic radical corresponds to the tone of the entire character (Wong 2013). For example, the phonetic radical in 饿 (hungry) is 我 /wǒ/ (I, me), but the correct pronunciation is /è/. The semantic radical on the left indicates a connection to food, but it cannot function alone as an integral character. As phonemic cues in written Chinese are not always consistent, CFL learners who depend on such strategies to decode words/phrases may find this aspect of reading difficult.

CFL Pedagogy

Orthographic instruction in many CFL classrooms typically mimics the methods used with native Chinese children. The most common approach is repetitive handwriting practice, emphasising rote memorisation of vocabulary and visually-accurate reproductions of each character. While this strategy has been criticised for its inefficiency (Allen 2008), largely because CFL learners lack the orthographic foundation needed for successful memorisation and reading, other scholars (McBride 2015; Packard et al. 2006) contend that handwriting practice remains an essential component of character learning. This latter view is grounded in the assumption that attention to form connects sensory-motor information with mental representations, thereby supporting long-term retention (Tan et al. 2005).

Harvey and Brooks (2025) provide evidence for the benefits of handwriting practice on reading comprehension and character production. They recruited 70 Grade 4 pupils from a Chinese immersion public school in America. Participants' L1s were not provided, though all students had received three years of prior Chinese-language instruction. Intact classes were assigned to conditions based on teachers' decisions: two groups engaged in typing via text-messaging, while another two groups completed pencil-and-paper tasks. Students attended two sessions per week for eight weeks (15–20 min per lesson), taught by different instructors and using simplified characters. Reading skills and character production were evaluated pre- and post-intervention using Level Chinese (a standardised reading test) and free-writing exercises. While the handwriting group scored slightly higher than the texting group on the reading task, this difference was not statistically significant. Performance on the free-writing task was statistically significantly poorer for the typing group compared to the handwriting group. The authors suggest that learners' poor vocabulary and pinyin spelling hindered meaningful engagement in the texting group, leading them to discuss unrelated topics rather than focusing on the target items. They concluded that handwriting practice is a core component of CFL teaching and learning, but that it is unclear how different testing formats might have influenced outcomes. They suggest that handwriting demands greater cognitive effort, requiring learners to memorise strokes in their correct order, but by contrast, typing is considerably easier, as students only need to type the pinyin spelling and select the corresponding character in the writing system, thus developing receptive rather than productive skills. Nonetheless, handwriting practice remains a dominant teaching approach, largely due to its long-standing practice in native literacy instruction and the lack of evidence-informed alternatives suitably designed for CFL learners.

Another popular strategy involves teaching characters through radicals. Drawn from native Chinese-language education, it is assumed that when students perceive characters as structured units, it is easier for them to visually encode and process the entire character (Chang et al. 2014). Y. Xu, Chang, and Perfetti (2014) investigated the effects of grouping characters by common radicals on students' orthographic skills. They recruited 48 first-year beginners and 40 second-year intermediate-level CFL students at an American university. Participants were evenly divided into the following: 24 beginner and 20 intermediate learners were assigned to the grouped condition (characters were taught using shared radicals), while the remaining half were placed in the distributed condition (instruction was spread across eight radical groups). They attended four one-hour sessions delivered by their regular teachers, using identical learning materials in simplified characters. Pre-, post-, and two-week delayed post-tests involved English-to-Chinese translations, pinyin transcriptions, lexical decision and matching tasks, and radical recognition and semantic awareness exercises. Findings showed that beginners in the grouped condition outperformed their peers in the distributed condition across all measures with statistically significant differences. However, this difference disappeared two weeks later. Intermediate learners in the grouped condition, compared with their counterparts in the distributed condition, only illustrated short-term gains in pinyin, radical awareness, and translations. This was not statistically significant, and no lasting benefits were found across any measure. The authors offer three explanations for these findings: (1) the teaching materials contained high-frequency radicals that intermediate learners may have previously encountered; (2) these learners likely had greater exposure to radicals, having studied the language for two years; and (3) the intervention was too short to demonstrate statistically significant effects. The findings suggest that radical-grouping might be more effective for beginner learners, but this may only lead to temporary orthographic gains.

The above research indicates that radical-grouping and handwriting practice may improve reading and orthographic knowledge among beginner learners, though such methods remain rooted in approaches developed for native speakers. Given the challenges outlined previously, it is important to ascertain whether the findings of these illustrative studies are replicated elsewhere, and to understand the nature and effects of other approaches to teaching character acquisition and reading among CFL learners. In the next section, we identify gaps in prior systematic reviews and present our research questions.

Prior Systematic Reviews

S. Zhang (2024) conducted a systematic review of studies published between 1952–2023, investigating research trends and the effects of different strategies for teaching characters on the writing outcomes of L1 English learners' of Chinese. The resulting synthesis of 22 studies showed that diverse strategies have been employed to support students in acquiring characters, such as handwriting practice, teaching radicals in chunks, various ways of presenting character information, timing of orthographic instruction, and haptic teaching through body gestures. The majority of the research was administered in university settings (17 out of 22), mainly involving adult beginners in their first year of tertiary education. Conversely, far fewer studies were

administered in middle- ($n = 2$) and high-schools ($n = 3$). Research was also distributed across three geographic regions: America ($n = 14$); Australia ($n = 5$); and Ireland ($n = 3$), with over half implementing experimental designs ($n = 12$), seven using mixed-methods, and three being exploratory. The outcomes measured across studies were thematically categorised into character recognition, production, radical awareness, and pinyin recall.

Based on these findings, S. Zhang (2024) suggests that more empirical research is needed on adolescent learners who remain an under-represented population in the existing literature. She also proposes that orthographic instruction should be treated as a separate stream within the curriculum, rather than being fully integrated with other language skills, although the rationale and efficacy of this approach requires further empirical investigation. S. Zhang (2024, 22) concludes that several methods are effective in improving students' writing outcomes, stating that there is no 'single best method' to teach characters. However, this claim is weakened by the lack of clarity regarding which 'several methods' the author is explicitly referring to, as well as the failure to acknowledge that many of the included studies featured methodological shortcomings, such as no comparison group, which complicate interpretation of the findings. Zhang did not evaluate the trustworthiness of the evidence or report the search terms used when preparing her review. Incomplete reporting and selection bias hinder an informed understanding of the overall strength of the findings (Chalmers, Brown, and Koryakina 2023). In addition, Zhang's review is limited to only L1 English students, and thus does not address CFL learners from other language backgrounds. Therefore, more work is needed to fully understand the state-of-the-art.

Our review aimed to locate, appraise, and synthesise evidence from intervention studies that examine the effects of different approaches to teaching Chinese orthography on school- and university-aged CFL learners' character acquisition and reading performance. According to S. Zhang's (2024) review, a wide range of instructional strategies have been explored in L1 Anglophone regions. By focusing more broadly on CFL contexts, our review allows for a more precise conceptualisation of both instructional settings and learning outcomes. Like S. Zhang (2024), we include only studies conducted within formal educational programmes, as our aim is to provide evidence relevant to the practices of CFL teachers and learners, rather than those in other programme settings (e.g., tutoring or self-study). However, unlike Zhang, our synthesis includes a trustworthiness assessment of the literature, enabling readers to better understand the strength of the evidence and to interpret the overall findings accordingly. Our review is reported in adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al. 2020) for the sake of completeness and transparency. In contrast to Zhang, our review also aims to provide practitioners with evidence-based guidance on pedagogical strategies that might be more effective for CFL learners at different proficiency levels.

Review Questions

This systematic review addresses the following questions:

(1) What instructional methods or pedagogical approaches to teaching Chinese orthography have been evaluated for effectiveness among school- and university-aged CFL learners?
(2) What are the effects of these methods on character acquisition and reading performance?
(3) What methods or approaches to teaching Chinese orthography are most effective for CFL learners at different proficiency levels?

Method

Protocol

The review protocol was registered with IDESR (International Database of Education Systematic Reviews) in January 2025 (https://idesr.org/article/IDESR000185; Qin and Chalmers 2025). The protocol was enacted unamended in the conduct of the review.

Inclusion and Exclusion Criteria

We included studies based on:

- Participants: the population was typically developing CFL learners in primary, secondary, and university contexts – that is, participants were not exclusively learners with developmental language disorders or fine motor issues. Studies that compared non-native and heritage students were included, though the focus was on non-native speakers' outcomes
- Intervention: the Chinese writing system and/or Mandarin was the subject of instruction, where part of the teaching could have involved independent study (e.g., homework). Studies were included even if the spoken variety or script (simplified or traditional) were not explicitly mentioned
- Outcomes: quantitative measures of reading performance and character acquisition (e.g., character/word recognition and production, pinyin transcriptions, reading comprehension)
- Study design: experimental designs (e.g., randomised controlled trial (RCT), quasi-experimental design, regression discontinuity design, matched-pair design)
- Date and type of publication: all published articles and grey literature, regardless of language or publication date

Studies were excluded based on:

- Participants: the population consisted exclusively of heritage learners, students in Chinese-speaking regions (e.g., Singapore, Hong Kong, Taiwan), or those studying non-Mandarin varieties
- Intervention: studies conducted in informal settings (e.g., parents teaching at home), involving learners not enrolled in taught programmes (e.g., learning for pleasure), or lacking an instructional period/focus on teaching methods (e.g., studies examining learner strategies only)

- Study design: studies that did not examine causal effects or report primary research (e.g., ethnographies, systematic reviews, non-intervention studies)

Information Sources

We searched the following databases on 25 January 2025: ProQuest Social Science Premium Collection (including ERIC), British Education Index, Linguistics Collection (including LLBA), Web of Science, PsycINFO, SCOPUS, and ProQuest Dissertations and Theses.

For SCOPUS, the search was further restricted to records indexed under the 'Social Sciences only' category. For Web of Science, the search was additionally restricted to the following categories:

- Education Educational Research
- Linguistics
- Language Linguistics
- Social Sciences Interdisciplinary

Search Strategy

The following search terms were used to identify relevant papers in the databases listed above: ab('Chinese as a foreign language' OR 'Mandarin as a foreign language' OR 'Chinese character*' OR 'Chinese orthograph*' OR 'Chinese writing system' OR 'Chinese learning') AND ab('Chinese reading' OR 'Chinese writing' OR 'radical knowledge' OR 'Chinese literacy' OR 'character recognition' OR 'character recall' OR 'character retention' OR 'character knowledge' OR 'character acquisition' OR 'character improvement' OR reading development OR 'character recall').

Selection Process

All records were uploaded to Rayyan, an online management tool that allows multiple authors to compare and evaluate studies based on the eligibility criteria (Ouzzani et al. 2016). After deduplication, abstracts lacking sufficient information to be confidently ruled out at this stage, along with those that clearly met the eligibility criteria, were included for full-text screening. Screening ceased for abstracts that breached any one of the inclusion criteria. To ensure the reliability of the screening process, a second reviewer studying a master's degree in applied linguistics was recruited and independently screened a random 10% sample of titles and abstracts ($n = 136$). The extent of agreement was calculated using Cohen's Kappa (McHugh 2012). The Kappa value was 0.76. Although this was below the threshold for strong agreement (0.81–1.00), only four discrepancies were identified between the first and second reviewer. Conflicts were resolved through discussion. Full reports were retrieved via the relevant databases and screened for eligibility by the first author. Only those that met all inclusion criteria were retained for subsequent data extraction and analysis.

Data Extraction and Data Items

We recorded four categories of data: (1) administrative: bibliographic information, publication type, language, database source, funding; (2) contextual: research location, participant demographics (e.g., age, gender, L1, CFL proficiency, socioeconomic background), research question(s), educational level, type of learning and institution, writing script; (3) intervention: study design and duration, sample size, method of participant allocation, measurement tools, teaching approach(es), pedagogical materials; and (4) outcomes: assessment measures, data analysis, effect size, findings, author's conclusions, implications, and strengths and weaknesses.

Trustworthiness Appraisal

This review followed Slavin's (1986) best-evidence synthesis approach; that is, methodological quality did not determine eligibility for the review, but methodological quality was appraised and reported as part of evaluating the strength and reliability of the assembled evidence. To assess the quality of individual studies, we applied Gorard's (2014) sieve, a framework for appraising methodological quality, and thus trustworthiness, of educational intervention research. Reports are rated on a scale of zero to four on six categories: design, scale, dropout, outcomes, fidelity, and validity, with the overall rating corresponding to the lowest classification received in any of the categories (Gorard 2014). Examples of completed data extraction forms can be found in the Supplementary Materials. Per the tool's guidance, if the rating for a particular category could not be determined from the reports (e.g., because of unclear reporting or omission), that category was automatically assigned 0*. As there were no guidelines for rating cutoffs in *Scale* and *Dropout*, we developed the following criteria to ensure consistent scoring:

Scale (per group):

- Trivial: $n \leq 5$
- Very small: $5 < n \leq 15$
- Small: $15 < n \leq 20$
- Medium: $20 < n \leq 50$
- Large: $n \geq 50$

Dropout:

- Minimal: 0–10% dropout (over 90% completion)
- Some: 11–20% dropout (between 80–90% completion)
- Moderate: 21–30% dropout (between 70–80% completion)
- High: Over 30% dropout (less than 70% completion)

Design was rated on the extent to which baseline equivalence between comparison groups was established before the intervention: random allocation received the highest rating (4*); followed by quasi-experimental or matched-pairs designs (2–3*); poor comparison groups (1*); and studies without a comparator received the lowest (0*). *Outcomes* were rated according to the type of assessment measures used: standardised tests received the highest rating (4*), even when authors did not provide details about the marking criteria or whether raters were blind to participant allocation; studies in which all outcomes were validated and reported received 3*; a rating of 2* was assigned to studies that combined validated tools with self-designed rubrics; and studies that offered little or no information regarding how outcomes were internally validated or how assessment measures were deemed reliable received 0–1*. *Fidelity* was rated according to the clarity and

consistency with which the treatment was implemented: studies with no reported instructor differences received the highest rating (4*); studies with some variation in teacher delivery received 3*; a rating of 2* was assigned to studies where the intervention lacked sufficient detail for replication and/or delivery varied greatly across contexts (i.e., participants did not engage with the intervention as intended); and studies with poorly described interventions received 0–1*. *Validity* was assessed based on the extent to which studies controlled for differences in the impact of the intervention. While it is acknowledged that controlling for all confounding variables is difficult without randomisation, we specifically focused on controlling known differences between research groups. Studies in this category were rated 1–2* when data were missing, reporting was unclear, or outcome measures included only pre- or post-tests rather than both. Studies received 4* when threats to validity were clearly acknowledged and their potential impact on the interpretation of results was addressed, while studies received 3* when threats were identified but only partially discussed.

Given the subjective nature of quality assessment, the second reviewer involved in the selection process also conducted a blind review of four randomly chosen studies. There was full agreement on overall ratings, though less consistency was found when scoring individual criteria on outcomes, fidelity, and validity. In such cases, both reviewers discussed the discrepancies and successfully reached a consensus. The first author then proceeded with independently evaluating the remaining full-texts.

Synthesis

A meta-analysis was deemed inappropriate due to the heterogeneity of interventions, comparison groups, and outcome measures across the included studies. Thus, a narrative synthesis approach was adopted (Popay et al. 2006). To address RQ1, we summarised the instructional strategies identified across the body of eligible literature. Regarding RQ2, studies were grouped according to their pedagogical approach, and their findings were synthesised by describing these approaches and summarising their relative effectiveness as reported in the original reports. To address RQ3, teaching methods were classified by proficiency level, and their level of trustworthiness was interpreted alongside the overall rating using Gorard's (2014) sieve.

Results

After deduplication, 1355 titles and abstracts were screened, of which 1243 were excluded for not meeting one or more of the inclusion criteria. Citation chaining identified 12 additional eligible papers. This resulted in 112 records being sought for retrieval; 14 of these were inaccessible despite interlibrary loan requests and attempts to contact the authors. Following full-text screening, 30 reports (representing 32 studies) met the eligibility criteria and were included in the synthesis (see Figure 1, PRISMA flow diagram).

Summary of Study Characteristics

Table 1 summarises the characteristics of each included study. The earliest eligible study was published in 1990, with peaks in 2014 and 2020 ($n = 4$ for each year; see Figure 2 for

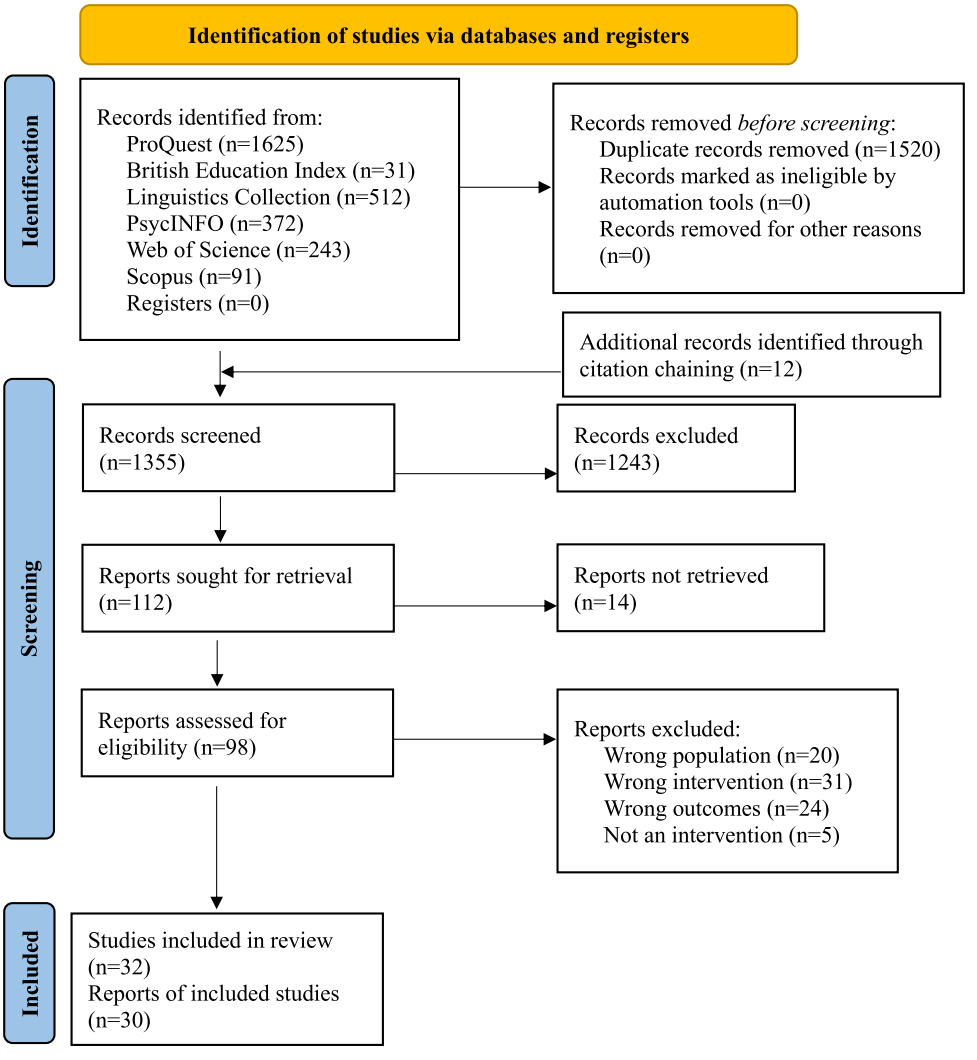

Figure 1. PRISMA flow diagram (adapted from Page et al. 2020).

the distribution of included studies across five-year intervals). There were 23 peer-reviewed journal articles, four doctoral theses, two master's dissertations, and one conference proceeding. Studies were conducted in America ($n = 20$), Australia ($n = 5$), Ireland ($n = 4$), and Thailand ($n = 1$). Sixteen studies recruited university students, 10 explored outcomes in secondary education, and four took place in primary schools. Study IDs (e.g., S1), which correspond to the information in Table 1, have been used to refer to individual studies in the commentary that follows.

Research designs were evenly distributed between RCTs ($n = 13$) and quasi-experimental studies ($n = 13$). Among the RCTs, four employed cluster randomisation, two used a randomised crossover design, one incorporated pair-matching (i.e., matching subjects with similar scores on pre-test results and randomly assigning one from each pair to the treatment or control condition), and in the remaining RCTs ($n = 6$), students were

Table 1. Study characteristics.

Study ID	Location	Education level	Design	Treatment	Comparator (or control group)	Writing script	Sample size	Length	Assessment focus	Results	Quality rating
1. Chang et al. (2014)	USA	University	Quasi-experimental crossover (pre-post and delayed post-test)	Chunking	(a) Pencil-and-paper handwriting (b) Passive reading (c) Stroke-order writing	Simplified	41	4 days	Character-recall and retention	Favored treatment group	0*
2. R. Chen (2020)	Thailand	Secondary	Single-group (pre-post-test)	Interactive learning platform	None	Simplified	12	4 weeks	Reading	Favoured treatment group	0*
3. Chiu (2024)	USA	Secondary	Quasi-experimental crossover (pre-post test)	Typing with pinyin input	Pencil-and-paper handwriting	Simplified	23	5 weeks	Character knowledge, vocabulary	Favoured treatment group	1*
4. Chung (2007)	Australia	Secondary	RCT crossover (pre-post and delayed post-test)	Study 1: Flashcards (various ways of presenting character information)	Study 2: Flashcards (with and without colour-coding)	Simplified	32 in Study 1 and Study 2	2–3 lessons per week (until pupils learned all characters)	Meaning-recall and retention	Favoured flashcards in both studies	0* for Study 1 and 2
5. Harvey and Brooks (2025)	USA	Primary	Quasi-experimental (pre-post-test)	Typing in chat-groups	Pencil-and-paper handwriting	Simplified	56	8 weeks	Reading	No significant differences between groups	2*
6. He and Huang (2014)	Australia	University	RCT (pre-post-test)	(a) Digital materials with English meaning and animated etymological information (b) Paper materials with English meaning and etymological	Paper materials with English meaning and character	Simplified	21	1 × 30-minute lesson	Character-recognition	Favoured digital materials group	1*

(*Continued*)

Table 1. Continued.

Study ID	Location	Education level	Design	Treatment	Comparator (or control group)	Writing script	Sample size	Length	Assessment focus	Results	Quality rating
				information via images							
7. Ke and Dubravac (2021)	USA	University	Quasi-experimental (mid-post- and delayed post-test)	Early exposure to characters	Delayed exposure to characters	Simplified	2	4 h per week over 14 weeks	Character recognition, retention	Favoured treatment group	0*
8. Knell and West (2017)	USA	Secondary	RCT (post-test)	Early exposure to characters	Delayed exposure to characters	Traditional	48	1 academic year	Character-recognition, reading	Favoured treatment group on one measure	2*
9. W. Li (2004)	USA	University	Quasi-experimental (pre-post-test)	Grouping characters by shared radicals	Rote memorisation	Simplified	50	8 weeks	Character-recognition, reading	Favoured treatment group	1*
10. J. T. Li and Tong (2020)	USA	Primary	RCT (pre-post and delayed post-test)	Verbal-coding* with etymological information	Verbal-coding* with pictures	Traditional	100	19 weeks	Character-recognition, retention	Favoured treatment group	0*
11. Osborne (2016)	Ireland	Secondary	Quasi-experimental (mid-post-test)	(a) Rote learning; (b) Delayed character instruction; (c) Colour-coding characters	Unity curriculum (focusing on listening, reading, writing, and speaking)	Not specified	98	1 academic year	Character-recognition and recall, reading, pinyin spelling	Effects differed across measures	0*
12. Osborne (2018)	Ireland	Secondary	RCT (pre-post-test)	(a) Rote learning (b) Delayed character instruction (c) Colour-coding characters	Unity curriculum (focusing on listening, reading, writing, speaking)	Not specified	85	16 weeks	Character-recognition and recall, pinyin spelling	Effects differed across measures	0*
13. Osborne, Zhang, and Zhang (2020)	Ireland	Secondary	Quasi-experimental (mid-post-test)	(a) Rote learning (b) Delayed character instruction (c) Colour-coding characters	Unity curriculum (focusing on listening, reading, writing, speaking)	Simplified	98	8 weeks	Character-recognition and recall, reading, pinyin spelling	Effects differed across measures	0*

14. Osborne, Zhang, and Adamson (2022)	Ireland	Secondary	Quasi-experimental (mid-post-test)	(a) Rote learning (b) Delayed character instruction (c) Colour-coding characters	Unity curriculum (focusing on listening, reading, writing, speaking)	Simplified	80	28 weeks	Character-recognition and recall, reading, pinyin spelling	Effects differed across measures	1*
15. Packard (1990)	USA	University	Quasi-experimental (post-test)	Early exposure to characters	Delayed exposure to characters	Traditional	23	26 weeks	Character production	Favoured the 'delayed' group	1*
16. Poole and Sung (2015)	USA	University	Quasi-experimental (post-test)	(a) Pinyin-only (b) Passive reading	Pencil-and-paper handwriting	Not specified	9	4 × 30-min lessons	Character-recognition, speaking	Effects differed across measures	0*
17. Poole, Clarke-Midura, and Ji (2022)	USA	Primary	Single-group (pre-post-test)	Digital game	None	Simplified	32	4x 25-min lessons over 4 weeks	Character recognition, reading	Favoured treatment group	0*
18. Ren (2004)	Australia	Secondary	Cluster RCT (post-test)	OVAL-writing**	PowerPoint	Not specified	19	10 weeks	Character retention	Favoured treatment group	0*
19. Shen (2010)	USA	University	Single-group (post- and delayed post-test)	Study 1: Verbal-coding with images; Study 2: Verbal-coding only	None	Simplified	40 in Study 1, 45 in Study 2	12 lessons	Concrete and abstract words	Effects differed across measures	0* for Study 1 and 2
20. Taft and Chung (1999)	Australia	University	Quasi-experimental (post- and delayed post-test)	Radicals introduced at various times: (a) Before the first lesson, (b) Early in the first lesson, (c) After the third lesson	No radicals were introduced	Simplified	40	3 × 15-min lessons over one week	Character-recall and meaning-pairings	Favoured 'radicals early' group	0*
21. Tsai (2014)	USA	University	Quasi-experimental	Character information with audio:	Pencil-and-paper handwriting	Simplified	63	10 lessons over 16 weeks	Character-recognition	Favoured control group	1*

(*Continued*)

Table 1. Continued.

Study ID	Location	Education level	Design	Treatment	Comparator (or control group)	Writing script	Sample size	Length	Assessment focus	Results	Quality rating
			crossover (post-test)	(a) Digital writing without feedback; (b) Typing input only (c) Digital writing with feedback					and production		
22. L. Wang (2005)	USA	University	Cluster RCT (post- and delayed post-test)	(a) Text plus animation aids. (b) Text with audio. (c) Text plus animation and audio.	Printed text	Traditional	72	1 × 50-minute lesson	Character-recognition, recall, and retention	Favoured text-plus-animation and audio group	0*
23. T. Wang (2024)	USA	University	Cluster RCT (pre-post-test)	Digital game	PowerPoint	Simplified	49	4 × 20-min lessons over 4 weeks	Radical knowledge	Favoured treatment group	0*
24. Wu (2012)	USA	University	RCT (pre-post-test)	Text-plus-audio	Printed text	Simplified	65	4 weeks	Character-recognition and recall, reading	Favoured treatment group	3*
25. P. Xu and Jen (2005)	USA	University	Cluster RCT (mid-post-test)	Typing with pinyin input	Pencil-and-paper handwriting	Not specified	138	3 academic years	Character-recognition	Favoured treatment group	0*
26. S. Xu and Ke (2020)	USA	University	RCT (pre-post- and delayed post-test)	Imitate character strokes through body movement	Viewing stroke-order sequences of characters onscreen	Simplified	53	1 × 40-minute lesson	Character-recognition, recall, and retention	Favoured treatment group	2*
27. X. Xu and Padilla (2013)	USA	Secondary	Matched-pairs RCT (pre-post- and delayed post-test)	Meaningful interpretation and chunking	Rote repetition and stroke-order writing	Simplified	108	4 days	Character-recognition, recall, and retention	Favoured treatment group	3*

28. Y. Xu et al. (2013)	USA	University	RCT crossover (pre-post- and delayed post-test)	(a) Digital writing (b) Animated demonstrations of stroke-order writing	Passive reading (online)	Traditional	36	6 × 20–40-min lessons	Character-meaning, recall, and retention	Favoured digital writing group	2*
29. Y. Xu, Chang, and Perfetti (2014)	USA	University	Quasi-experimental (pre-post- and delayed post-test)	Grouping characters by shared radicals	Learning characters across eight radical groups	Simplified	88	4 lessons	Character and radical knowledge	Favoured treatment group	3*
30. T. Zhang (2019)	Australia	Primary	Single-group (pre-post-test)	Online comics	None	Simplified	60	10 weeks	Character production	Favoured treatment group	0*

*Verbal encoding involved the instructor teaching characters through verbal definitions and/or explanations.
**OVAL-writing approach: observe, visualise, articulate, listen, write. The instructional sequence involved students observing the character's structure, mentally visualising its form, practising pronunciation, listening to the instructor's pronunciation, and concluding with handwriting practice.

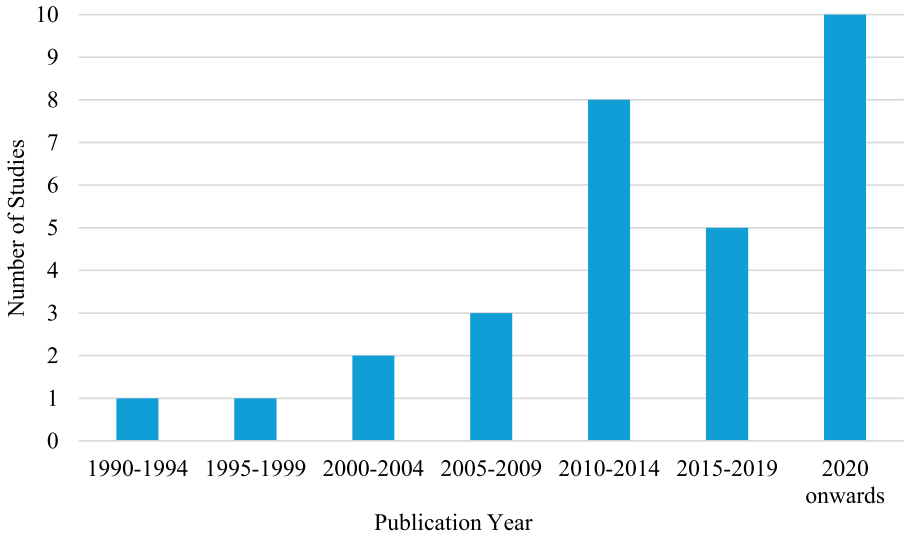

Figure 2. Included studies by publication year.

Table 2. Summary of research designs.

Research design	Assessment timing	Quantity	Study ID
Cluster RCT	Pre-post-test	1	S23
	Post-test only	1	S18
	Post- and delayed post-test	1	S22
	Mid-post-test	1	S25
Crossover RCT	Pre-post- and delayed post-test	2	S4, S28
Individual RCT	Pre-post-test	3	S6, S12, S24
	Pre-post- and delayed post-test	2	S10, S26
	Mid-post-test	1	S8
Matched-pair RCT	Pre-post- and delayed post-test	1	S27
Quasi-experimental	Pre-post-test	3	S3, S5, S9
	Pre-post- and delayed post-test	2	S1, S29
	Mid-post- and delayed post-test	1	S7
	Mid-post-test only	3	S11, S13, S14
	Post-test only	2	S15, S16
Quasi-experimental crossover	Post-test only	1	S21
	Post- and delayed post-test	1	S20
Single-group	Pre-post-test	3	S2, S17, S30
	Post- and delayed post-test	1	S19

individually randomised into groups, following a parallel design. Four studies adopted single-group designs. Regarding outcome measures, six studies evaluated outcomes mid-way through the intervention and upon its completion, while another six studies relied solely on post- and/or delayed post-test data. Table 2 summarises the research designs. Twenty studies used simplified characters and four focused on traditional characters. Six studies did not clarify which writing script was employed.

All included studies implemented quantitative measures of character acquisition and/or reading performance. Twenty-eight reports focused on orthographic knowledge, which encompassed a range of subskills, including character or radical recognition, recall, retention, form and meaning recall, and pinyin transcriptions. Two studies focused on reading only. Assessments involved a variety of receptive and productive

tasks, such as reordering Chinese sentences, character-picture matching, and translating Chinese to English and vice versa.

The duration of interventions varied widely, from a single 30-minute lesson to treatments spanning three years. However, five reports lacked sufficient information to confidently judge their duration.

CFL proficiency was mostly determined by researchers, classroom teachers, or through students' self-reports based on the total number of characters they could recognise prior to the intervention. Descriptions of proficiency were often vague, using terms like 'novice', 'beginner' or 'non-fluent'. Eight university-based studies identified participants as first-year beginners, with reported character acquisition ranging between 180–400. One study measured proficiency by instructional exposure, categorising students as beginners after 150 h. Three studies involved bilingual programmes (English-Chinese) in American primary school settings, with no details on the percentage or number of hours taught in each language. Only one study used a standardised measurement for CFL proficiency, the Hanyu Shuiping Kaoshi, for classifying participants as Level 1 (complete beginners). Lastly, six studies reported no prior exposure to Mandarin. Consequently, these findings highlight a general lack of consistency in the classification and reporting of CFL learner proficiency across studies.

The final column of Table 1 reports the quality ratings assigned to each study using Gorard's (2014) sieve (see Supplementary Materials for individual trustworthiness assessments). Altogether, 19 (out of 32) studies received 0*, five were rated 1*, four were 2*, and three were 3*. Studies with low trustworthiness mainly resulted from the use of unvalidated or non-standardised instruments to measure outcomes, insufficient control of confounding variables in study design and data analysis, and small numbers of cases (especially in cluster RCTs).

Findings of RQ1: What Pedagogical Approaches to Teaching Chinese Orthography Have Been Evaluated for Effectiveness among School- and University-Aged CFL Learners?

This review identified five broad pedagogical approaches: computer-assisted language learning (CALL), conventional techniques, character component instruction, timing of orthographic exposure, and holistic methods. CALL studies differed in interactivity and learner engagement, from passive approaches using animated stroke-order demonstrations with audio, to formats that required linguistic output (e.g., creating online comics). Several studies also explored various ways of presenting orthographic knowledge. Other studies compared instructional timing, contrasting immediate orthographic teaching with delayed approaches where students exclusively learned pinyin before transitioning to characters. Conventional techniques largely reflected native Chinese teaching practices, emphasising repetition and rote memorisation. These were typically evaluated against other methods like CALL. Holistic strategies balancing speaking, listening, reading, and writing were additionally implemented as comparators to explicit orthographic instruction. Lastly, one study explored the effects of an OVAL-writing approach, combining the character's visual structure, its oral articulation, and handwriting practice. Table 3 displays the breakdown of pedagogic methods reported across the studies.

Table 3. Included studies by instructional approach.

Instructional category	Specific method	Quantity	Study ID*
CALL	Digital games	2	S17, S23
	Chat-based texting	1	S5
	Animated orthographic input with audio	4	S6, S21, S22, S28
	Online comics	1	S30
	Kinaesthetic-haptic learning with images	1	S26
	Typing characters with pinyin input	3	S3, S21, S25
	Animated demonstrations of stroke-order	1	S28
	Interactive learning platform	1	S2
	Printed text plus audio	1	S24
	Digital writing	1	S28
Conventional techniques	Pencil-and-paper instruction (including handwriting, stroke-order writing)	9	S1, S3, S5, S6, S16, S21, S25, S27, S28
	Rote memorisation/repetition	7	S9, S11, S12, S13, S14, S17, S27
	PowerPoint	2	S18, S23
	Passive reading	5	S1, S16, S22, S24, S28
	Pinyin only	1	S16
	Verbal definitions of characters	1	S19
Character components	Grouping the same radicals to form compound characters	3	S9, S20
	Colour-coding characters and/or tone marks	5	S4, S11, S12, S13, S14
	Visual chunking	3	S1, S27, S29
	Flashcards (different sequences of information presented, e.g., pinyin-character-English meaning vs. character-pinyin-English meaning)	1	S4
	Etymological explanations of radicals	2	S6, S10
	Verbal definitions of characters with pictures	1	S19
	Learning characters across eight different radical groups	1	S29
Timing of instruction	Early vs. delayed character introduction	3	S7, S8, S15
Holistic approach	Unity curriculum (balancing speaking, reading, writing, and listening)	4	S11, S12, S13, S14
	OVAL-writing	1	S18

*Study IDs are listed more than once in cases where authors evaluated multiple instructional approaches within a single study.

Findings of RQ2: What are the Effects of These Methods on CFL Learners' Character Acquisition and Reading Performance?

An overview of the included studies with their characteristics and reported effects is provided in Table 1, with detailed findings presented below.

Timing of Orthographic Exposure

Three studies compared two methods that manipulated the timing of orthographic exposure: an 'early' group learned characters from the beginning of the intervention, and a 'delayed' group initially focused on pinyin with characters introduced later in the intervention. Knell and West (2017) found that immediate exposure facilitated better reading skills, but no statistically significant between-group differences emerged in character recognition. Ke and Dubravac (2021) identified statistically significant advantages in character recognition for students exposed early rather than later. Conversely, Packard (1990) found significant benefits in pinyin transcriptions and oral proficiency for delayed instruction only. While Osborne (2016, 2018) and Osborne, Zhang,

and Zhang (2020, 2022) did not directly compare these approaches, they reported statistically significant gains in pinyin knowledge with delayed exposure. This was evaluated against holistic methods, rote memorisation, and colour-coding.

Collectively, findings on the optimal timing for orthographic learning were mixed and are complicated by methodological and reporting shortcomings. Ke and Dubravac (2021) conducted a 14-week intervention, with the early group starting in Week 1 and the delayed condition in Week 5, meaning groups had different quantities of overall instructional time. Similar was true of Packard (1990), who attempted to account for the imbalance by providing three additional one-hour lessons per semester. Osborne (2016, 2018) and colleagues (2020, 2022) do not specify when character teaching began post-delay or the total number of instructional hours. These inconsistencies are noteworthy, as differences in outcomes may reflect variations in exposure rather than instructional timing. Study quality does little to moderate these effects: despite Knell and West (2017) obtaining the highest rating of this subsample of the literature (2*) for trustworthiness, the early group were introduced to characters in Week 2 (around 30 contact hours), while their counterparts began three months later (18–20 h). The remaining studies were rated 1* or 0*. Therefore, the available evidence does not offer confident conclusions regarding the relative effectiveness of early versus delayed exposure for orthographic acquisition and reading development.

Digital Games

Two studies employed digital games. Poole, Clarke-Midura, and Ji (2022) adopted a single-group design to examine the effects of an online game, where CFL learners completed language tasks across China as adventurers. Participants used in-game glossaries to understand content and complete missions independently. This was supplemented with workbook exercises (e.g., sentence construction, handwriting practice). Findings showed statistically significant gains in both vocabulary and reading skills from pre- to post-test. Workbook completion was statistically significantly correlated with improvements in reading performance.

T. Wang (2024) designed a digital escape-room game using Google Forms to teach radicals, comparing its effectiveness to PowerPoint instruction. Game progression depended on correctly identifying radical functions and decoding compound characters through multiple-choice questions. Compared to their peers, game-based instruction resulted in statistically significant gains in orthographic knowledge from pre- to post-test. These studies, however, are undermined by methodological limitations: Poole, Clarke-Midura, and Ji (2022) lacked a control group and T. Wang's (2024) cluster RCT involved only two cases.

Texting, Interactive Platform, Online Comics

Harvey and Brooks (2025) compared text-messaging with handwriting instruction to analyse its impact on reading comprehension and orthographic production. The texting condition engaged in online discussions on everyday topics using prompts provided by the authors, whereas their peers completed various pencil-and-paper exercises. No statistically significant differences on any measure were found between groups from

pre- to post-test. Limited effects in texting were linked to poor pinyin skills and vocabulary, resulting in many students ignoring the prompts and discussing unrelated topics. The authors conclude that typing should only supplement instruction, not replace it.

R. Chen (2020) examined an interactive platform designed to improve reading comprehension, employing a single-group design. The platform included videos on grammar and group-chat features, both accessible on smartphones and computers. Statistically significant gains were reported in reading performance. How this intervention performed relative to alternative teaching methods is impossible to say because the study did not include a comparator. The author does not provide access to the platform, and thus impedes precise understanding of the nature of the intervention.

T. Zhang (2019) also adopted a single-group design to assess whether online comics benefit character recognition and production. Each weekly, two-hour lesson involved 40 minutes of instruction via online comics from the author, with the remaining time delivered by the classroom teacher. Participants showed statistically significant gains, with 23% achieving full marks in the post-test, possibly suggesting a ceiling effect.

However, both R. Chen (2020) and T. Zhang (2019) lacked comparator groups, limiting the ability to attribute improvements to the teaching methods alone. T. Zhang's (2019) study also lacked sufficient control over confounding variables, as only 40-minutes of each two-hour lesson was devoted to the intervention. It is unknown whether orthographic gains stemmed from the intervention or were additionally influenced by the regular teacher's instruction. The above evidence suggests that texting, online comics, and interactive platforms offer no discernible advantage for character learning and reading. Harvey and Brooks (2025) were rated the highest of this subsample (2*), and the remaining studies obtained 0*.

Digital Handwriting and Typing Characters

One study examined digital handwriting, two investigated typing, and one compared both approaches. In their randomised crossover study, Y. Xu et al. (2013) explored whether different CALL presentation formats increased orthographic knowledge, comparing animation, digital handwriting, and reading. Participants initially completed a reading task on computers, with the other conditions counterbalanced. Each session followed the same structure: 60 characters were presented with pinyin and audio (10 per session); students viewed each character for 15 seconds (reading-only); they watched stroke-by-stroke animations three times (animation group); and the digital handwriting condition involved form recall. All groups demonstrated statistically significant pre-post improvement: the reading group outperformed others in pinyin production, meaning recall, and character production, followed by the animation and digital writing groups respectively. No statistically significant effects were found at the one-month delayed post-test between groups. Although this study received 2*, as tests were not standardised or validated, the authors employed an RCT design, reported attrition, and showed strong validity and fidelity in the implementation of the intervention.

Chiu (2024) conducted a quasi-experimental crossover design to compare two instructional sequences: handwriting-to-typing and typing-to-handwriting. Teaching focused on radical knowledge, writing in the correct stroke order, copying vocabulary lists, and constructing Chinese sentences. The typing group performed similar activities on a computer.

While both groups improved in orthographic recognition, the handwriting-first condition outperformed on both pre-post tests, with statistically significant differences, particularly in sentence construction and grammar. However, this study was rated 1* as the author did not appropriately control for heritage learner background (18 of 23 pupils spoke L2 Chinese).

P. Xu and Jen's (2005) cluster RCT compared typing with handwriting to investigate the effects of a custom word-processing programme on orthographic acquisition. This software differed from standard input systems by offering real-time error feedback for pinyin or missing characters in sentences. The authors reported positive outcomes associated with typing but did not specify whether effects were statistically significant, nor provide the raw data to allow for this to be calculated. Details on participant demographics and assessments were also omitted. Given such incomplete reporting, this study was rated 0*.

Using a quasi-experimental crossover design, Tsai (2014) evaluated handwriting instruction against three conditions implementing orthographic animation with audio: digital writing without feedback; typing-only; and digital writing with immediate error-specific feedback. All groups began with orthographic animation and audio, followed by weekly writing exercises in their assigned condition, with treatment order counterbalanced. Only post-tests were administered after participants had completed all learning conditions, assessing character recognition, production, and stroke-order awareness. Findings revealed that handwriting outperformed typing-only in accuracy and stroke-order knowledge, showing statistically significant effects, while error-specific feedback illustrated no benefits across any measure. However, without pre-tests, the effects of each condition remain unclear, and the use of intact classes raises concerns about whether research groups were comparable (1*).

The above evidence offers minimal practical guidance on whether teachers should prioritise digital writing, typing, or handwriting instruction. Results are mixed, with digital writing performing worse than reading and animation (Y. Xu et al. 2013), Chiu (2024) and Tsai (2014) favouring handwriting-first, while P. Xu and Jen (2005) conclude that typing is more beneficial for character learning. Collectively, the evidence is limited and of relatively low quality.

Kinaesthetic-Haptic Learning

X. Xu and Ke's (2020) RCT was the only study investigating kinaesthetic-haptic learning, combining imagery and full-body movement to teach characters. Participants used a Microsoft Kinect V2 to mimic character forms, while the control group learned identical content using a laptop. Although immediate post-test results showed no statistically significant group differences, the treatment group still performed statistically significantly better at the three-day delayed post-test. However, learning was limited to simple characters with six or fewer strokes, as only these forms could be represented through body gestures. It is unclear if this approach would yield similar outcomes for characters with greater stroke complexity. CFL teachers should also consider the feasibility of replicating the required technological setup. The study was rated 2* due to moderate attrition.

Animation and Text Vocalisation

Two studies explored whether video demonstrations with etymological explanations resulted in orthographic gains. He and Huang's (2014) RCT compared animated

presentations of character evolution with two paper-based conditions: one showing the character and its English meaning, and the other adding etymological images. Each animation included pictures depicting the character's meaning and its English definition, without pinyin or audio. The treatment lasted one 30-minute lesson. Pre-test scores illustrated no statistically significant differences between groups. Post-tests revealed that animations outperformed both paper-based conditions in character recognition with statistically significant effects. While etymological images outperformed text-only instruction, this effect was only statistically significant on picture-based tasks. However, these results apply to characters with fewer than eight strokes, and long-term gains remain unknown. This study was rated 1* due to the small number of cases per group.

L. Wang (2005) evaluated text-with-audio, text-with-animation, and text with both audio-and-animation, against traditional printed-text instruction. Pre-existing textbook software was adopted to teach characters, which provided explanations through both words and images, and included audio narration instead of on-screen text. Twenty-one compound characters were taught in one 50-minute session. The audio-and-animation group scored the highest on post-tests in character recognition, followed by text-plus-animation, printed-text (control), and text-plus-audio. Only text-with-animation and audio-and-animation were significantly correlated with post-test results. Although differences were non-significant, audio-and-animation illustrated the highest retention after one week, while printed-text showed the lowest. The remaining groups displayed equal retention. Due to the trivial scale (only two cases) of Wang's cluster RCT, the study was rated 0*.

Wu (2012) explored the effects of text vocalisation on reading performance, character/word recognition, and recall. Over four weeks, participants completed online reading activities featuring clickable audio links recorded by native speakers, while the comparison group accessed text-only versions. At post-test, the intervention group achieved statistically significantly better scores than the control group in word recognition, recall, and reading speed, but not in overall comprehension. Mean score analysis suggested that lower-performing learners benefited most from vocalised text, while high-achievers demonstrated no additional gains. This study was among the most methodologically rigorous (3*), adopting random allocation and reporting the internal reliability of assessment tools.

These findings suggest that animated orthographic input (with or without audio) can enhance reading and orthographic outcomes. This approach appears particularly successful for characters with fewer than nine strokes, though long-term benefits remain unclear. Text vocalisation seems to support lower-proficiency CFL readers. Given that this approach has been evaluated in only one study, this promising finding deserves attention in future research.

Radical Components

Four studies implemented this approach. W. Li's (2004) quasi-experiment investigated whether grouping radicals enhanced reading comprehension and vocabulary compared to rote memorisation. Students learned five radicals with etymological images and their English meaning. They subsequently used these radicals to form 16 characters and

completed paper-based tasks that involved combining radicals into compound characters. After eight weeks, the grouped-radicals group scored statistically significantly higher than the rote memorisation group across all tests. However, assessment tools lacked internal validity, and participants were self-selected into beginner programmes, thereby obtaining 1* for study quality.

In their quasi-experimental crossover design, Chang et al. (2014) adopted a similar pedagogical strategy, which was explored against handwriting practice, reading-only, and stroke-order writing, spanning four days (one lesson per day). The principal author initially taught 48 characters and facilitated comprehension through guided questions. Participants then engaged in independent study on laptops: each item was presented chunk-by-chunk on-screen, with each demonstration lasting nine seconds. This procedure was repeated three consecutive times for each character in every session. Pre-tests showed no statistically significant between-group differences in form recall and text production. While radical-grouping achieved statistically significantly greater gains than other methods in the short-term, these advantages were not sustained after two weeks. This study received 0* due to incomplete reporting on attrition.

Taft and Chung's (1999) quasi-experiment examined the impact of radical instruction at various stages on orthographic knowledge. Learners were presented with a set of 24 characters on three occasions. Sixteen radicals were used to create 72 items, each appearing in three characters and occupying identical spatial positions (e.g., 口/kǒu/ is on the left of 听/tīng/ and 叶 /yè/). These were displayed individually with English translations. Four conditions were compared: (1) participants in the radicals-before group initially learned 16 radicals. Subsequently, students were told that radicals form the structural basis of characters and completed 15-minute of writing practice; (2) in the radicals-early condition, radicals were explained as each character was introduced; (3) the radicals-late group received no instruction until their third exposure; and (4) the no-radicals (control) group learned characters without orthographic information. The radicals-early group performed better than other groups on character recall at both immediate- and one-week delayed post-tests, with statistically significant differences. This study obtained 0* due to unreported attrition.

Y. Xu, Chang, and Perfetti (2014) compared the effectiveness of radical-grouping with an approach in which students learned 48 compound characters from eight separate radical groups (i.e., distributed condition). Forty-eight beginners and 40 intermediate CFL learners were divided into intact classes: half from each proficiency-level received radical-grouping instruction, while the remaining students followed the distributed approach. Across four one-hour sessions, students spent 20 min on teacher-led vocabulary tasks without etymological input and then independently viewed each character on laptops for around 30 seconds. The treatment group learned two radical groups at a time, while the distributed condition studied across eight different radical groups. First-year CFL pupils displayed better recall and radical awareness on post-test scores compared to their peers in the distributed group, with statistically significant effects. Intermediate learners in both groups showed no statistically significant gains in radical awareness, and lexical improvements declined after two weeks for both proficiency-levels. The authors conclude that radical-grouping is most effective for beginners. The study was rated 2* as some assessments were not internally validated.

Low trustworthiness in most of these studies limits reliable conclusions about whether radical-grouping outperforms traditional methods like repetitive handwriting instruction.

Colour-Coding and Flashcards

Five studies employed colour-coding techniques, one of which examined its use specifically on flashcards. In addition to investigating delayed character exposure, Osborne (2016, 2018) and Osborne, Zhang, and Zhang (2020, 2022) compared colour-coding with rote memorisation and a 'unity curriculum' (control) approach (focusing on listening, speaking, reading, writing) to explore its impact on beginner learners' reading, form recognition, and recall. Lexical items were colour-coded by tone marks: neutral in pencil, and green, black, blue, and red for tones 1–4 respectively. In Osborne (2016), rote memorisation produced the highest recall and recognition scores at both mid- and post-tests, followed by colour-coding, the control group, and delayed exposure. Conversely, Osborne (2018) found colour-coding most effective, yielding the highest gains in orthographic knowledge from pre- to post-test, followed by delayed instruction, rote memorisation, and the control group. It is unclear whether these findings were statistically significant. In both studies, she attributes higher scores to a combination of the instructional approach and students' independent study outside class, making it difficult to isolate intervention effects.

Osborne, Zhang, and Zhang (2020) observed that different teaching strategies produced distinct linguistic outcomes. Rote memorisation demonstrated statistically significant greater gains than other methods in character recognition at both mid- and post-tests, while colour-coding was most effective for pinyin transcription. Delayed instruction performed statistically significantly worse across all measures in recall, recognition, and production, which the authors attribute to insufficient teaching time. The control group also demonstrated limited gains, displaying a slight but non-significant increase in lexical diversity during sentence production. Similar findings were observed in Osborne, Zhang, and Adamson (2022), with results in orthographic knowledge favouring rote memorisation and colour-coding from mid- to post-test, showing statistically significant differences. While the authors recommend combining rote memorisation and colour-coding for character teaching, they acknowledge that assessment measures primarily targeted character-centred rather than sentence-level tasks. However, the methodological quality of these experiments are weak, with Osborne, Zhang, and Adamson (2022) receiving the highest rating of this group of studies (1*). Only Osborne (2018) adopted an RCT design, whereas others were quasi-experimental and lacked pre-tests, assessing progress mid-way and post-intervention. This restricts the ability to determine instructional effectiveness due to unknown baseline proficiency.

In Chung's (2007) randomised crossover study, he conducted two experiments evaluating the effects of different orthographic presentation formats on character acquisition. Sixteen two-character Chinese nouns were shown on flashcards under four conditions: character-pinyin-English; character-English-pinyin; pinyin-English-character; and English-pinyin-character. Participants completed all presentation conditions in a counterbalanced order and received one-on-one instruction from the author until they had been exposed to all lexical items. Results showed that placing characters to the left of

the English word significantly enhanced both short- and long-term retention. Pronunciation improved when characters preceded its pinyin, and placing English words beside the character demonstrated better results in meaning acquisition than other formats.

In a follow-up experiment, Chung (2007) examined if colouring the pinyin or English word would increase students' orthographic awareness. Sixteen two-character words were written on flashcards, following the same teaching procedure as the first experiment. Colour-coding pinyin and English words significantly facilitated pronunciation and meaning acquisition compared to no colouring. However, allowing unlimited learning sessions poses a significant limitation in both experiments. The author does not specify how many sessions each student attended. This discrepancy in treatment intensity raises concerns about whether such positive results reflect the actual efficacy of colour-coding and presentation formats or merely differences in instructional exposure. Both experiments were rated 0* for trustworthiness due to unreported attrition.

Collectively, there is weak evidence suggesting that colour-coding or rote learning is better than an integrated or delayed approach. Similarly, positioning characters before English words or after pinyin on flashcards does not show greater benefits beyond those explained by time and exposure.

Meaningful Interpretation and Chunking

X. Xu and Padilla (2013) was the only study comparing meaningful interpretation and chunking with handwriting practice. Beginners, intermediate, and advanced learners were matched by pre-test scores and randomly assigned to treatment or control groups. Students completed a pre-intervention lesson on character origins and structures, where the principal author explained how shared radicals convey meaning. The four-day intervention involved teaching eight characters per day: Day 1 covered memorisation strategies; Day 2 included learners practising visual chunking (i.e., recalling characters with shared radicals); and Days 3–4 focused on individual study. The control group followed the same structure but focused on stroke-order writing on Day 2. Post-test findings revealed that the treatment group performed statistically significantly better than the comparator in character recognition and recall, though gains declined after two months. However, it is noteworthy that only transparent characters were taught, which may have favoured the treatment condition. These characters are easier to decode since their phonetic and semantic radicals align with their pronunciation and meaning. Nonetheless, the study was robust (3*) methodologically, employing an RCT design and accounting for confounding variables like heritage learner background. Indeed, no significant interactions emerged between proficiency, heritage status, and treatment at either pre- or post-test, though beginners and non-heritage learners consistently underperformed, indicating the need for targeted support.

Etymological Explanations

Two studies compared teaching characters through verbal and pictorial etymological explanations. In J. T. Li and Tong's (2020) RCT, treatment lessons consisted of a 5-minute overview, 25 minutes of vocabulary instruction with pictograms and etymological details, and 10 minutes of review where the instructor (also the principal author) read a

story featuring target characters. The comparison group received identical instruction, except characters were taught with images rather than pictograms or etymological explanations. Lessons were administered 2–4 times per week over 19 weeks, with 2–5 weeks of in-class writing practice to consolidate learning. Despite both groups demonstrating statistically significant improvements in character recognition from pre- to post-test, some knowledge diminished after one week. The treatment group consistently outperformed the control group at both post-tests. Only 4–5 weeks of repetition boosted vocabulary recall to a statistically significant effect, whereas shorter intervals showed no additional benefit. Raw scores also revealed that students' ability to recall characters increased from 23% after two weeks of repetition to 53% after five weeks. The authors conclude that pictograms and etymological explanations offer distinct benefits for orthographic acquisition and retention. However, it is unclear whether positive results were from the intervention, repeated practice, or a combination of both. The study received 0* for incomplete reporting on attrition.

In a single-group design, Shen (2010) conducted two experiments to compare the effects of etymological explanations and image-based instruction on learning concrete and abstract words. Experiment 1 involved presenting 10 concrete items with etymological explanations via PowerPoint. The teaching procedure involved students repeating each word aloud three times, practising in pairs, and selecting matching flashcards after hearing the instructor's pronunciation. The process was then repeated with 10 new concrete words using images-only. This procedure was implemented in Experiment 2 to teach abstract items, recruiting the same pupils. Post-tests on pinyin transcription and form recognition were conducted immediately and one day later. Experiment 1 found no statistically significant differences between approaches, whereas Experiment 2 showed that image-only instruction resulted in statistically significant gains in recall. The author argues that concrete items are inherently imageable and likely linked to visual representations already present in learners' mental lexicon, thereby reducing the added benefit of pictures. Both experiments obtained 0* for lack of comparison groups.

Evidence supporting etymological explanations over image-based instruction is inconclusive. J. T. Li and Tong's (2020) findings may reflect greater exposure and increased repetition, while Shen's (2010) single-group design limits causal interpretation. The relative effectiveness of either approach is therefore unclear.

Holistic Approaches

Ren (2004) conducted a cluster RCT to investigate the effectiveness of an OVAL-writing approach against PowerPoint instruction. This approach involved five steps: observing character formation; visualising its form; articulating/explaining its structure aloud; listening to the teacher's pronunciation; and practising writing while repeating the word in English and Chinese. The intervention involved three stages: (1) introducing 27 target items; (2) completing OVAL-writing at spaced intervals (10 min, 24 hours, eight days, 13 days, and 15 days); (3) and assessing retention through written cue-dependent tasks. The control group received PowerPoint instruction without spaced practice. Post-tests followed each interval for the treatment condition, while the control group was tested only at 10-minutes (T1) and on Days 15 (T2) and 16 (T3). OVAL-writing

participants scored higher than the control group at all stages (T1–T3), but no statistical analyses were reported. Ren concludes that this approach improves recall, though prior CFL proficiency and pre-tests were not included, in addition to details on the writing script and types of characters taught. The observed differences in outcomes, however, may be attributable to the effects of spaced practice rather than the OVAL-writing approach. With only two cases in its cluster RCT, the study is severely underpowered, thereby receiving 0* for trustworthiness.

Conventional Techniques

Poole and Sung (2015) were the only study to compare conventional strategies with each other. Their quasi-experiment evaluated the effects of pinyin-only, handwriting, and reading-only instruction on character recognition and oral proficiency. The pinyin-only group used PowerPoint slides focusing on tones and pronunciation without orthographic instruction. The handwriting condition practised stroke-order writing, while the reading group focused on form recognition and oral fluency. Instruction for reading was delivered in Mandarin, whereas other groups were taught in English. No rationale is provided for this difference in the language of instruction. Post-tests were conducted after the third and fourth sessions. In oral proficiency, the pinyin-only group achieved the highest, followed by reading-only, and then handwriting. This trend was reversed for character recognition. The authors attribute these findings to differences in instructional focus/time allocated to oral and orthographic exposure across conditions, though it should be noted that no statistical analyses were conducted. Due to small sample size and variations in instructional exposure across groups, this study obtained 0*. The evidence does not clearly favour one approach for improving orthographic and/or reading outcomes.

Findings of RQ3: What Specific Methods or Pedagogic Approaches to Teaching Chinese Orthography Are Most Effective for CFL Learners at Different Proficiency Levels?

As is evident from our summary of the eligible evidence, approaches to teaching Chinese orthography to CFL learners are varied, as are the conditions to which they are compared. Teasing out how different approaches relate to each other and what, if any, implications exist for particular groups from this multifaceted literature is not straightforward. Moreover, while some studies in the assembled literature adopted designs that are more or less well suited to evaluating causal relationships between teaching approaches and learning outcomes, the overarching theme is methodological and reporting shortcomings. These complicate assessment of the effects of these approaches relative to each other and for different groups of learners, and do not inspire confidence in the cumulative evidence. Nonetheless, we have interpreted the findings of the assembled literature in terms of what it suggests for two groups: beginner learners and intermediate and advanced learners. Given the overall low trustworthiness of the evidence base, any reports of effectiveness presented in the following synthesis should be interpreted in light of this shortcoming in the primary research.

Beginners

Current research indicates considerable variation in the effectiveness of CFL teaching strategies. Studies on orthographic timing suggest that delayed instruction enhances oral proficiency and pinyin skills (Osborne 2016, 2018; Osborne, Zhang, and Zhang 2020, 2022; Packard 1990). Conversely, early exposure benefits reading (Knell and West 2017) and character recognition (Ke and Dubravac 2021). Harvey and Brooks (2025) conclude that typing offers no advantage over handwriting for younger learners with limited vocabulary. This suggests that CFL learners may need to achieve a foundational level of proficiency before illustrating measurable gains. Chiu's (2024) handwriting-typing comparison offers evidence that secondary students require more orthographic support prior to typing, though results should be interpreted cautiously as heritage learner effects were not controlled for. P. Xu and Jen (2005) reported positive gains from their custom word-processing programme, but incomplete reporting and lack of software access limit replicability and practical application. Their conclusion favouring typing over handwriting practice contrasts with Harvey and Brooks (2025) and Chiu (2024), likely due to their inclusion of adult CFL learners who may possess greater cognitive abilities than adolescents. X. Xu and Ke (2020) found that kinaesthetic-haptic learning facilitated recognition and retention of simple characters. Other CALL methods, including animated orthographic input and text-audio pairing, appear more successful than handwriting practice or reading-only approaches for both short- and long-term acquisition (He and Huang 2014; L. Wang 2005; Wu 2012). While online platforms, digital games, and comics improved learning, these studies adopted single-group designs, thereby limiting causal inferences (Poole, Clarke-Midura, and Ji 2022; R. Chen 2020; Zhang 2019). Only one study compared digital games to PowerPoint slides, with results favouring the former (T. Wang 2024).

Findings further suggest that teaching characters through common radicals can improve outcomes (Chang et al. 2014; W. Li 2004; Y. Xu, Chang, and Perfetti 2014), especially when radicals are explained as they are presented (Taft and Chung 1999). While meaningful interpretation and chunking may yield short-term gains, this method does not appear to enhance knowledge retention, highlighting the need for greater exposure/repetition (X. Xu and Padilla 2013). J. T. Li and Tong (2020) found that combining etymological explanations with pictures was better than merely presenting the character's form and English translation. These results were reaffirmed by Shen (2010), noting greater recall for abstract items only. Colour-coding and rote learning were also associated with improved orthographic knowledge and oral skills, though long-term benefits were not tested (Osborne 2016, 2018; Osborne, Zhang, and Zhang 2020, 2022). Chung's (2007) experiments show that placing English translations on the left of characters on flashcards facilitates meaning acquisition, and colour-coding enhances orthographic awareness. But, inconsistent treatment intensity may have influenced outcomes, as participants attended as many sessions as required to master the vocabulary. Ren (2004) reported increased character retention with an OVAL-writing method compared to PowerPoint slides, but statistical power is hindered by the study's small sample. Poole and Sung's (2015) investigation of traditional techniques revealed that pinyin-only instruction was most successful for speaking, whereas handwriting practice yielded the greatest gains in character recognition, followed by reading.

Intermediate and Advanced Learners

Research on CFL pedagogy involving intermediate and advanced learners was scarce, with evidence drawn from only two studies. Y. Xu, Chang, and Perfetti (2014) found no significant advantage to radical-grouping when teaching characters to second-year intermediate CFL learners. This was compared with a distributed approach, where subjects learned across eight radical sets. The authors suggest that participants' prior exposure to radicals and combinational patterns likely influenced outcomes. Conversely, results from X. Xu and Padilla (2013) demonstrated that meaningful interpretation and chunking can benefit both intermediate and advanced students, leading to short-term linguistic gains, but these effects diminished two months later. Differences in outcomes may derive from the types of characters/radicals tested: Xu et al. incorporated high-frequency radicals and identical teaching materials for all proficiency levels, while Xu and Padilla controlled for memory load, number of strokes, character frequency, and phonetic transparency to enable more reliable within-subject comparisons. The evidence, however, is insufficient to provide strong pedagogical recommendations.

Conclusion and Implications

This review synthesised and summarised a diverse range of approaches to teaching Chinese orthography, such as CALL, timing of instruction, character component presentation, and holistic approaches, often against conventional techniques like rote memorisation. The findings present a broadly positive picture of the pedagogical effects of many of these approaches on character learning and reading comprehension, with only one study reporting non-significant outcomes. However, given the heterogeneity of the approaches and the methodological characteristics of the literature, there is little reliable evidence that any single approach enhances acquisition over alternatives. Indeed, any approach appears better than no approach, which is a finding with limited practical value for CFL teachers and learners wanting to improve outcomes. This also contradicts S. Zhang's (2024) systematic review, which, despite focusing solely on L1 English CFL learners and lacking trustworthiness assessments, argues that several methods have 'proved' (22) successful for writing skills. Research on intermediate and advanced learners is also sparse and inconsistent, with a notable bias towards studies examining foundational character teaching at American universities. Consequently, little is known about how to support higher-proficiency learners or those from non-Anglophone backgrounds. This raises concerns for school-based teaching and learning, especially given the growing global interest among primary and secondary students in studying Mandarin.

These findings, however, can help CFL practitioners revisit and reflect upon their current strategies, as well as be inspired to experiment with different ways of promoting orthographic acquisition and reading development. Nonetheless, CFL pedagogy is still an under-researched phenomenon, especially in primary- and secondary- settings and among higher-proficiency learners in diverse geographic contexts. Indeed, the current sample represents a narrow learner group: educated adults, most of whom are native English speakers. A learner's L1 plays an important role in foreign language learning due to factors such as cross-linguistic transfer effects. As such, the narrow participant

samples in the eligible studies limit the applicability of the pedagogical implications to learners outside these groups, and further restrict our understanding of effective strategies for students from diverse educational levels and language backgrounds.

Therefore, future research should focus on (1) CFL learners from different L1 backgrounds within taught programmes and those at pre-university level; (2) more longitudinal research to investigate knowledge retention; (3) objective measures of learner proficiency through the implementation of validated and/or standardised tests, such as the HSK; and (4) adopting designs that are well-suited for making confident causal inferences (i.e., the inclusion of comparison groups and random allocation of participants). Researchers should clearly report the methodology of their studies to maximise replicability and ensure accurate interpretations of findings. Finally, the results may encourage teachers to adjust their approaches while considering factors such as resource availability and logistical constraints.

One limitation of this review was confining the search process to Western educational databases only. It is worth noting, though, that the aim of this review was to identify methods effective for CFL learners, who are typically based outside China, and thus may not be the focus of studies published in Chinese-language academic journals. Future updates to this review can consider broadening the range of databases in the search strategy to capture a wider selection of relevant literature. Another limitation pertains to the search terminology employed. Although the search strategy incorporated widely-used terms such as 'Chinese characters', it did not include the term, 'Hanzi', which has become increasingly common in Chinese language teaching and learning research. Consequently, it is possible that some eligible studies using Hanzi-specific terminology were not identified and excluded in this review. In addition, while eligible studies implemented some writing measures (e.g., translations), writing was not examined as a distinct outcome alongside reading performance to maintain a manageable scope for the review. Future reviews are encouraged to investigate writing development to provide a more in-depth understanding of Chinese orthographic acquisition. Lastly, variations in personal judgement may have influenced how the level of trustworthiness was evaluated when interpreting Gorard's (2014) sieve. It is equally important to acknowledge the limitations of using Gorard's sieve for quality appraisals: (1) the organisation of the sieve means that overall ratings often obscure the nuanced strengths of individual studies; and (2) while its intentional vagueness is designed to accommodate diverse research contexts, the lack of precise criteria (e.g., in Scale and Dropout) can complicate the rating process.

Collectively, this systematic review has found that, while numerous pedagogical strategies have been examined in the literature to enhance CFL reading and orthographic outcomes, robust evidence is severely lacking. The strategies evaluated in this review may serve as a valuable starting point, but teachers should be wary and avoid placing undue confidence in their efficacy.

Author contributions

CRediT: **Sonia Qin:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing; **Hamish Chalmers:** Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

The author(s) confirm that no AI tools were used in the preparation, writing, formatting, and editing of this work.

ORCID

Sonia Qin http://orcid.org/0009-0008-2948-0315
Hamish Chalmers http://orcid.org/0000-0003-2687-9722

References

*Denotes Publications Selected as Included Studies in This Review.

Allen, J. R. 2008. "Why Learning to Write Chinese Is a Waste of Time: A Modest Proposal." *Foreign Language Annals* 41 (2): 237–251. https://doi.org/10.1111/j.1944-9720.2008.tb03291.x.

Boltz, W. 1994. *The Origin and Early Development of the Chinese Writing System*. New Haven, CT: American Oriental Society.

British Council. 2023. *Language Trends*. https://www.britishcouncil.org/sites/default/files/language_trends_england_2023.pdf.

Chalmers, H., J. Brown, and A. Koryakina. 2023. "Topics, Publication Patterns, and Reporting Quality in Systematic Reviews in Language Education. Lessons from the International Database of Education Systematic Reviews (IDESR)." *Applied Linguistics Review* 15 (4): 1645–1669. https://doi.org/10.1515/applirev-2022-0190.

*Chang, L. Y., Y. Xu, C. A. Perfetti, J. Zhang, and H. C. Chen. 2014. "Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language." *International Journal of Disability, Development and Education* 61 (3): 288–305. https://doi.org/10.1080/1034912X.2014.934016.

Chen, H. C. 1992. "Reading Comprehension in Chinese: Implications from Character Reading Times." *Advances in Psychology* 90 (1): 175–205. https://doi.org/10.1016/S0166-4115(08)61892-0.

*Chen, R. 2020. "The Use of UMU Interactive Platform in Chinese Language Reading Skills of Grade 10 Thai Students." Master's thesis, Rangsit University, Thailand. https://rsuir-library.rsu.ac.th/handle/123456789/524.

*Chiu, C. 2024. "How Does Writing Instruction Impact Chinese as a Foreign Language Learners' Literacy?" Doctoral thesis, University of Delaware, United States of America. https://www.proquest.com/openview/c9341259df9fd7c16384ea18866df8bb/1?pq-origsite = gscholar&cbl = 18750&diss = y.

*Chung, K. H. 2007. "Presentation Factors in the Learning of Chinese Characters: The Order and Position of Hanyu Pinyin and English Translations." *Educational Psychology* 27 (1): 1–20. https://doi.org/10.1080/01443410601061306.

Ésik, S. 2020. "Teaching Chinese Characters to Second Language Learners." *Researching and Teaching Chinese as a Foreign Language* 3 (1): 1–22. https://doi.org/10.1558/rtcfl.40506.

Fu, Y. 2024. "The Impact of Gaokao High-Stakes Testing on Student Mental Health in China: An Analysis of Stress Levels and Coping Mechanisms among Senior High School Students." *Research and Advances in Education* 3 (5): 23–32. https://doi.org/10.56397/RAE.2024.05.03. https://www.paradigmpress.org/rae/article/view/1136.

Gil, J. 2024. "Confucius Institute and Confucius Classroom Closures: Trends, Explanations, and Future Directions." *Applied Linguistics Review* 15 (2): 699–712. https://doi.org/10.1515/applirev-2022-0111.

Goh, Y. S. 2017. *Teaching Chinese as an International Language: A Singapore Perspective*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781107280472.

Gorard, S. 2014. "A Proposal for Judging the Trustworthiness of Research Findings." *Radical Statistics* 110:47–60. http://www.radstats.org.uk/no110/Gorard110.pdf.

Han, Z. 2016. "Research Meets Practice: Holding off and Holding on." *Chinese as a Second Language* 51 (3): 236–251. https://doi.org/10.1075/csl.51.3.01han.

*Harvey, R. E., and P. J. Brooks. 2025. "Effects of Text Messaging Using Digital Pinyin Input on Literacy Skills of Elementary Chinese Immersion Learners." *Language Teaching Research* 29 (4): 1753–1779. https://doi.org/10.1177/13621688221099909.

*He, J., and H. Huang. 2014. "Learning Chinese Characters with Animated Etymology." *International Journal of Computer-Assisted Language Learning and Teaching* 38 (1): 99–118. https://doi.org/10.4018/ijcallt.2014040105.

Honorof, D., and L. Feldman. 2006. "The Chinese Character in Psycholinguistic Research: Form, Structure, and the Reader." In *The Handbook of East Asian Psycholinguistics*, edited by P. Li, L. H. Tan, E. Bates, and O. J. L. Tzeng, 195–217. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511758652.

Hu, B. 2010. "The Challenges of Chinese: A Preliminary Study of UK Learners' Perceptions of Difficulty." *Language Learning Journal* 38 (1): 99–118. https://doi.org/10.1080/09571731003620721.

Huang, S. 2022. "A Tale of Two Less Successful CSL Readers." In *Reading in Chinese as an Additional Language*, edited by L. Li and D. Zhang, 156–178. London: Routledge. https://doi.org/10.4324/9781003029038.

*Ke, S., and S. Dubravac. 2021. "When Should Characters Be Introduced to Novice-Level Chinese in a Blended Learning Setting?" *Studies in Chinese Learning and Teaching* 6 (1): 1–26. https://drive.google.com/file/d/1nzmMwVw_m3evfMEEtdfw8y8DsedixU8D/view?usp = share_link.

*Knell, E., and H. West. 2017. "To Delay or Not to Delay: The Timing of Chinese Character Instruction for Secondary Learners." *Foreign Language Annals* 50 (3): 519–532. https://doi.org/10.1111/flan.12281.

*Li, J. T., and F. Tong. 2020. "The Effect of Cognitive Vocabulary Learning Approaches on Chinese Learners' Compound Word Attainment, Retention, and Learning Motivation." *Language Teaching Research* 24 (6): 834–854. https://doi.org/10.1177/1362168819829025.

*Li, W. 2004. "The 'Grapheme Combination Method': Teaching and Learning Chinese Characters Through Associative Links." Doctoral thesis, George Mason University, United States of America. https://www.proquest.com/openview/f3614ac5a9b4a52564fa023c771c0612/1?pq-origsite = gscholar&cbl = 18750&diss = y.

Lü, X., and J. Zhang. 1999. "Reading Efficiency: A Comparative Study of English and Chinese Orthographies." *Literacy Research and Instruction* 38 (4): 301–317. https://doi.org/10.1080/19388079909558298.

McBride, C. A. 2015. "Is Chinese Special? Four Aspects of Chinese Literacy Acquisition That Might Distinguish Learning Chinese from Learning Alphabetic Orthographies." *Educational Psychology Review* 28 (3): 523–549. https://doi.org/10.1007/s10648-015-9318-2.

McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochemia Medica* 22 (3): 276–282. https://doi.org/10.11613/BM.2012.031. https://pmc.ncbi.nlm.nih.gov/articles/PMC3900052/.

Norman, J. 1988. *Chinese*. New York: Cambridge University Press.

*Osborne, C. 2016. "Chinese in the Classroom: Initial Findings of the Effects of Four Teaching Methods on Beginner Learners." *Journal of Second Language Teaching and Research* 5 (1): 202–225. https://pops.lancashire.ac.uk/index.php/jsltr/article/view/443.

*Osborne, C. 2018. "Examining Character Recognition and Recall Skills of CFL Beginner Learners under Four Different Approaches." *Journal of the Irish Association for Applied Linguistics* 25 (1): 52–73. https://doi.org/10.35903/teanga.v25i0.49.

*Osborne, C., Q. Zhang, and B. Adamson. 2022. "The Next Steps for Teaching Characters in CFL: Investigating the Effects of Four Character Teaching Methods on Beginner Learners." *International Journal of Chinese Language Education* 11 (1): 45–82. http://ijcle.eduhk.hk/en/catalog/No.11/05/.

*Osborne, C., Q. Zhang, and G. X. Zhang. 2020. "Which Is More Effective in Introducing Chinese Characters? An Investigative Study of Four Methods Used to Teach CFL Beginners." *Language Learning Journal* 48 (4): 385–401. https://doi.org/10.1080/09571736.2017.1393838.

Ouzzani, M., H. Hammady, Z. Fedorowicz, and A. Elmagarmid. 2016. "Rayyan – a Web and Mobile App for Systematic Reviews." *Systematic Reviews* 5 (1): 1–10. https://doi.org/10.1186/s13643-016-0384-4.

Packard, J. L., X. Chen, W. Li, X. Wu, J. S. Gaffney, H. Li, and R. C. Anderson. 2006. "Explicit Instruction in Orthographic Structure and Word Morphology Helps Chinese Children Learn to Write Characters." *Reading and Writing* 19 (1): 457–487. https://doi.org/10.1007/s11145-006-9003-4.

*Packard, J. L. 1990. "Effects of Time Lag in the Introduction of Characters into the Chinese Language Curriculum." *Modern Language Journal* 74 (1): 167–175. https://doi.org/10.2307/328120.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hossmann, C. D. Mulrow, L. Shamseer, et al. 2020. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *Research Methods and Reporting* 372 (71): 1–9. https://doi.org/10.1136/bmj.n71.

Perfetti, C. A., S. Zhang, and I. Berent. 1992. "Reading in English and Chinese: Evidence for a 'Universal' Phonological Principle." *Advances in Psychology* 94 (1): 227–248. https://doi.org/10.1016/S0166-4115(08)62798-3.

*Poole, F., J. Clarke-Midura, and S. Ji. 2022. "Exploring the Affordances and Effectiveness of a Digital Game in the Chinese Dual Language Immersion Classroom." *Journal of Technology and Chinese Language Teaching* 13 (1): 46–73. https://www.researchgate.net/publication/363233499_Exploring_the_Affordances_and_Effectiveness_of_a_Digital_Game_in_the_Chinese_Dual_Language_Immersion_Classroom_tansuodianziyouxizaizhongwenchenjinshiketangzhongdekeyonggongnengheyouxiaoxing.

*Poole, F., and K. Sung. 2015. "Three Approaches to Beginning Chinese Instruction and Their Effects on Oral Development and Character Recognition." *Eurasian Journal of Applied Linguistics* 1 (1): 59–75. https://doi.org/10.32601/ejal.460599.

Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, M. Rodgers, N. Britten, K. Roen, and S. Duffy. 2006. "Guidance on the Conduct of Narrative Synthesis in Systematic Reviews." *A Product from the ESRC Methods Programme Version* 1 (1): b92. https://doi.org/10.13140/2.1.1018.4643.

Qin, S., and H. Chalmers. 2025. "The Substantive Linguistic Effects of Alternative Instructional Methods on CFL Learners' Character Acquisition and Reading Outcomes. Protocol for a Systematic Review." *International Database of Education Systematic Reviews*. https://idesr.org/article/IDESR000185.

*Ren, G. 2004. "Introducing OVAL Writing: A New Approach to Chinese Character Retention for Secondary Non-Chinese-Speaking Background Learners." In *Australian Federation of Modern Language Teachers Associations National Conference*. https://search.informit.org/doi/epdf/10.3316ielapa.200409296.

Shen, H. H. 2005. "An Investigation of Chinese-Character Learning Strategies among Non-Native Speakers of Chinese." *System* 33 (1): 49–68. https://doi.org/10.1016/j.system.2004.11.001.

*Shen, H. H. 2010. "Imagery and Verbal Coding Approaches in Chinese Vocabulary Instruction." *Language Teaching Research* 14 (4): 485–499. https://doi.org/10.1177/1362168810375370.

Slavin, R. E. 1986. "Best-evidence Synthesis: An Alternative to Meta-analytic and Traditional Reviews." *Educational Researcher* 15 (9): 5–11. https://doi.org/10.3102/0013189X015009005.

Snow, D. 2013. "Revisiting Ferguson's Defining Cases of Diglossia." *Journal of Multilingual and Multicultural Development* 34 (1): 61–76. https://doi.org/10.1080/01434632.2012.699531.

*Taft, M., and K. Chung. 1999. "Using Radicals in Teaching Chinese Characters to Second Language Learners." *Psychologia* 42 (4): 243–251.

Tan, L. H., J. A. Spinks, G. F. Eden, C. A. Perfetti, and W. T. Siok. 2005. "Reading Depends on Writing in Chinese." *Proceedings of the National Academy of Sciences of the United States of America* 102 (24): 8781–8785. https://doi.org/10.1073/pnas.0503523102.

Taylor, I., and M. M. Taylor. 2014. *Writing and Literacy in Chinese, Korean, and Japanese*. Amsterdam: John Benjamins. https://www.torrossa.com/en/resources/an/5015596.

*Tsai, C. H. 2014. "Multimedia Mediation and Chinese Orthographic Character Learning among Non-Heritage CFL Beginners." Doctoral thesis, University of Iowa, United States of America. https://www.proquest.com/openview/3e5f9ba65425c30938c0af490e8d4e40/1?pq-origsite = gscholar&cbl = 18750.

University of Sydney. n.d. *CHNS2601: Chinese Lower Intermediate 1*. https://www.sydney.edu.au/units/CHNS2601.

*Wang, L. 2005. "The Impact of Multimedia on Chinese Learners' Recognition of Characters: A Quantitative and Qualitative Study." Doctoral thesis, Purdue University, United States of America. https://www.proquest.com/openview/bb0b89648be7ec45eda92cd19b891f5c/1?pq-origsite = gscholar&cbl = 18750&diss = y.

*Wang, T. 2024. "Designing a Digital Game for Chinese Character Learning: A Theory-Driven Practice Approach." *Education Sciences* 14 (12): 1–14. https://doi.org/10.3390/educsci14121366.

Wong, K. S. R. 2013. "Learning to Read Chinese: The Importance of Reading Skills in a Non-alphabetic Language." *Perspectives on Language and Literacy* 39 (1): 33–39. https://repository.eduhk.hk/en/publications/learning-to-read-chinese-the-importance-of-reading-skills-in-a-no-5/.

*Wu, Y. 2012. "Using External Text Vocalisation to Enhance Reading Development among Beginning Level Chinese Learners." *Journal of the Chinese Language Teachers Association* 47 (1): 1–23.

*Xu, P., and T. Jen. 2005. "'Penless' Chinese Language Learning: A Computer-Assisted Approach." *Journal of the Chinese Language Teachers Association* 40 (2): 25–42.

*Xu, X., and F. Ke. 2020. "Embodied Interaction: Learning Chinese Characters through Body Movements." *Language Learning & Technology* 24 (3): 136–159. https://doi.org/10.64152/10125/44744.

*Xu, X., and A. M. Padilla. 2013. "Using Meaningful Interpretation and Chunking to Enhance Memory: The Case of Chinese Character Learning." *Foreign Language Annals* 46 (3): 402–422. https://doi.org/10.1111/flan.12039.

*Xu, Y., L. Y. Chang, and C. A. Perfetti. 2014. "The Effect of Radical-Based Grouping in Character Learning in Chinese as a Foreign Language." *The Modern Language Journal* 98 (3): 773–793. https://doi.org/10.1111/modl.12122.

*Xu, Y., L. Y. Chang, J. Zhang, and C. A. Perfetti. 2013. "Reading, Writing, and Animation in Character Learning in Chinese as a Foreign Language." *Foreign Language Annals* 46 (3): 423–444. https://doi.org/10.1111/flan.12040.

Yang, J. 2022. "Teenage Beginners' Perceptions of Learning Chinese Characters: A Case Study." *Journal of Chinese Writing Systems* 6 (11): 3–15. https://doi.org/10.1177/25138502221050960.

Yue, Y. 2017. "Teaching Chinese in K-12 Schools in the United States: What Are the Challenges?" *Foreign Language Annals* 50 (3): 601–620. https://doi.org/10.1111/flan.12277.

Zhang, S. 2024. "Effective Character Teaching Methods for L1 English Chinese as a Foreign Language Learners: A Review of Empirical Research." *Chinese as a Second Language* 58 (2): 1–31. https://doi.org/10.1075/csl.00033.zha.

*Zhang, T. 2019. "Online Comics for the Teaching and Learning of Chinese Language in the Australian Context." Master's thesis, Western Sydney University, Australia.

Zhao, Y. 2011. "Review Article: A Tree in the Wood: A Review of Research on L2 Chinese Acquisition." *Second Language Research* 27 (4): 559–572. https://doi.org/10.1177/0267658311417836.