ReCALL (2026), 1–20
doi:10.1017/S0958344026100603

RESEARCH ARTICLE

Methodological homogeneity in informal digital learning of English: A systematic review and roadmap for diversification

Xian Liu
The Education University of Hong Kong, Hong Kong Special Administrative Region, China (s1158164@s.eduhk.hk)

XinWei Pan
The Education University of Hong Kong, Hong Kong Special Administrative Region, China (s1158175@s.eduhk.hk)

Ju Seong Lee
The Education University of Hong Kong, Hong Kong Special Administrative Region, China (jslee@eduhk.hk)

Abstract

Research on informal digital learning of English (IDLE) has grown substantially, yet a systematic analysis of its methodological foundations is lacking. To map current practices, this methodological review examines 104 peer-reviewed studies (2015–2025) indexed in the *Arts & Humanities Citation Index* or *the Social Sciences Citation Index*. The findings reveal a dominant pattern of *methodological homogeneity*, characterized by heavy reliance on non-experimental designs (94.23%), self-report instruments (notably questionnaires; 69.49% of quantitative tools), and descriptive and correlational analyses (80.44% of quantitative analysis strategies). This profile is understandable, as the informal, nonlinear, and learner-driven nature of IDLE aligns naturally with exploratory approaches. However, this focus may limit the investigation of causal mechanisms and theoretical development. To support the field's evolution, we propose a strategic diversification of methods. Integrating longitudinal designs to trace development, complementing self-reports with ethically collected and activity-sensitive behavioral and trace data, and leveraging new analytical tools such as learning analytics and, where appropriate, AI-assisted analysis can enrich inquiry and help build more robust, explanatory models of informal digital learning.

Keywords: informal digital learning of English (IDLE); informal language learning; methodological review; research design; methodological homogeneity; research rigor

1. Introduction

The rapid advancement of technology – including, more recently, developments in generative AI (GenAI) – coupled with the widespread availability of portable digital devices and an abundance of engaging online English learning resources, has profoundly reshaped the way English is learned today (Lai & Sundqvist, 2026; Liu, Soyoof, et al., 2025). Increasingly, English as a foreign language (EFL) learners are independently using these digital tools and resources to pursue informal, self-directed English learning beyond classroom boundaries (Guo & Lee, 2023; Soyoof et al., 2023;

Corresponding author: Ju Seong Lee; Email: jslee@eduhk.hk

Cite this article: Liu, X., Pan, XW. & Lee, J.S. (2026). Methodological homogeneity in informal digital learning of English: A systematic review and roadmap for diversification. *ReCALL* FirstView, 1–20. https://doi.org/10.1017/S0958344026100603

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

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Zou et al., 2026). This distinct phenomenon is known as informal digital learning of English (IDLE).

To precisely define IDLE, we draw on Benson's (2011) framework for learning beyond the classroom, which considers dimensions of *formality*, *location*, *pedagogy*, and *locus of control*. Within this framework, IDLE refers to self-initiated, naturalistic English learning activities that occur in out-of-class settings. Learners employ digital tools (e.g., smartphones or computers) and resources (e.g., apps, games, videos, podcasts, social media, and AI-mediated tools) to learn English without direct teacher intervention or formal assessment (Lee, 2022b).

To situate IDLE among related constructs, we conceptualize it as part of a nested hierarchy. At the broadest level, *informal second language learning (ISLL)* refers to unstructured, out-of-class learning of any second language (Kusyk et al., 2025). Within ISLL, *extramural English* (*EE*; Rød & Calafato, 2025; Sundqvist & Sylvén, 2016) narrows the focus to informal contact with English outside the classroom, spanning both digital (e.g., online gaming) and non-digital (e.g., reading paperback books) activities. *IDLE*, in turn, can be understood as the digital-specific subset of EE, focusing mainly on online and technology-mediated informal English practices (Zou et al., 2026). However, as IDLE learners tend to self-select their activities, platforms, and intensity in digital environments, IDLE is difficult to standardize and often partly unobservable. This challenge applies across diverse forms of IDLE, including gaming, video viewing, podcast listening, social media participation, online chatting, app-based practice, and AI-mediated interaction. As a result, researchers often rely on exploratory, cross-sectional, and qualitative approaches to capture naturally occurring practices and generate initial explanatory hypotheses (Lee, 2026; Sundqvist, 2024).

To date, researchers have identified over 40 factors – both antecedents and outcomes – associated with IDLE engagement. The antecedents can be broadly grouped into three domains: (a) *demographic variables*, such as gender and access to digital devices (Zhang & Liu, 2025); (b) *internal attributes*, including ideal L2 self, enjoyment, basic psychological needs, and perceived usefulness of IDLE (Gao et al., 2025; Liu, Darvin, & Ma, 2024; Liu, Zhang, & Zhang, 2024); and (c) *external influences*, such as parental support (Soyoof et al., 2025). On the outcome side, participation in IDLE has been linked to a wide range of learning benefits. These include *affective outcomes*, such as reduced anxiety and boredom (Rezai et al., 2025; Taherian et al., 2024) and increased enjoyment and motivation (Fu, 2025; Rezai et al., 2025), as well as *psychological gains* like improved well-being and a greater willingness to communicate (Fu, 2025; Lee & Chiu, 2024; Taherian et al., 2024). *Cognitive benefits* include enhanced self-efficacy (Fang et al., 2025), while *linguistic improvements* encompass richer vocabulary and more fluent speaking skills (Rezai et al., 2026; Tam & Reynolds, 2023). *Cultural benefits* have also been noted, including greater intercultural competence (Liu, Ma, et al., 2025). Finally, *pedagogical outcomes*, such as enhanced digital competence and technological pedagogical content knowledge, further demonstrate the transformative potential of IDLE (Rezai et al., 2024b). However, findings on IDLE remain inconsistent across studies, often because of differences in sampling, measurement – especially the use of frequency-only IDLE scales – and the predominance of cross-sectional designs. A methodological review is therefore needed to examine the design, measurement, and analytic choices underlying these claims.

As research on IDLE has grown rapidly, several comprehensive reviews have attempted to synthesize the growing body of work. Among the earliest, Zhang et al. (2021) conducted a scoping review of extramural language learning. Their review demonstrated the positive influence of extramural language learning on both linguistic proficiency and affective outcomes, while highlighting the mediating variables such as gender and target language proficiency. However, the study's scope was limited by its narrow keyword search, which produced a relatively small dataset ($n = 33$). Moreover, by incorporating both offline (non-digital) and online (internet-based) activities across multiple languages other than English (LOTE; e.g., Chinese), the review's breadth diluted its capacity to capture the unique characteristics of IDLE.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Building on this work, Kusyk et al. (2025) conducted a more methodologically rigorous scoping review of 206 papers published between 2000 and 2020, aiming to identify key trends in ISLL. Using targeted search terms such as "outside the classroom," "extracurricular," and "digital wild," they observed a growth of ISLL research, especially in Europe and East Asia. Most studies employed cross-sectional designs, relying heavily on questionnaires and interviews. While the review confirmed a positive link between ISLL engagement and L2 proficiency, it noted that causal relationships remained unclear. Another important insight was the field's *terminological diversity* and frequent definitional inconsistencies – what Kelley (1927) called the "jingle-jangle" problem – which make cross-study comparisons more difficult. For instance, some studies use the terms IDLE and EE interchangeably, even though EE is a broader construct that also includes non-digital activities. This lack of clear operationalization creates a methodological challenge, as it becomes unclear whether findings are specific to technology-mediated learning or apply more broadly to informal learning. Although the review was broad in scope and analytically rigorous, its inclusion of both offline and online learning, as well as LOTE, reduced its focus on IDLE specifically. In addition, because the review only covered studies up to 2020, it did not capture more recent developments associated with the rise of GenAI, including tools such as ChatGPT, which emerged in 2022 (Zou et al., 2026).

A more focused review was provided by Soyoof et al. (2023), who examined 30 IDLE studies between 1980 and 2019. Their analysis identified three main research themes: (a) *linguistic outcomes*, (b) *affective and cultural dimensions*, and (c) *digital literacy and agency*. The review also showed that most studies were conducted in Europe and Asia, focused on young adult learners, and mainly used qualitative or mixed-methods designs. While the review offered a useful thematic synthesis (Chong & Reinders, 2021), its contribution was limited by the absence of a formal review protocol, which reduces the reproducibility and systematicity of knowledge synthesis in IDLE.

Liu, Soyoof, et al. (2025) sought to address these gaps by conducting a thematic review of IDLE research published between 2014 and 2024, with a focus on Asian EFL contexts where exposure to English remains relatively limited. They identified three broad themes: (a) IDLE as an individual difference variable shaping English development; (b) teachers' perceptions of, and readiness to integrate, IDLE into instruction; and (c) the role of technology access and digital literacy in contributing to unequal participation. While the review offered useful insights, it – like Soyoof et al. (2023) – did not follow a formal review protocol. In addition, its focus on Asia meant that perspectives from other important regions, such as Europe, were not included, which limits its global scope.

Addressing some of these limitations, Guo and Lee (2023) conducted a systematic review of IDLE based on Bronfenbrenner's (1979) ecological systems theory. Their analysis showed that most IDLE research has focused on individual and microsystem factors, while broader contextual dimensions, such as the mesosystem, exosystem, macrosystem, and chronosystem, have received less attention. The authors therefore called for a more holistic ecological approach to IDLE research, highlighting the importance of examining how parental support, institutional settings, and cultural contexts shape learner engagement. They also encouraged the use of more varied methodologies and cross-cultural comparisons to better understand the role of context in IDLE. Although the study was systematic and theoretically grounded, its main focus was on influencing factors rather than methodological issues. As a result, it did not directly address questions of research design quality, instrument validity, or analytical rigor that affect the strength of the field's cumulative evidence.

Taken together, these reviews show that IDLE research has grown and become more diverse in scope. At the same time, they point to a continuing gap: there is still no systematic account of the field's methodological practices. Most previous reviews focus on themes, contexts, or influencing factors, while paying less attention to how evidence is actually produced. This matters because choices about research design and measurement, such as the heavy reliance on self-report and cross-sectional data, shape how confidently the field can speak about causality, underlying

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

mechanisms, and ecological validity. Without a clearer picture of current practices, repeated calls for greater rigor and transparency remain difficult to put into practice (Chong et al., 2022). A methodological review is therefore needed to map current approaches, identify key methodological bottlenecks, and support more robust and replicable future research.

To respond to this need, the present study uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to conduct a methodological review of IDLE research published between 2015 and 2025. We recognize that PRISMA was originally developed for health-care intervention reviews and may not fully accommodate a heterogeneous and conceptually diverse field such as IDLE. Its structured inclusion and exclusion procedures may also leave out some nuanced work. Even so, we adopt its core principles for three main reasons (Page et al., 2021): first, to support a transparent and rigorous search and selection process; second, to provide a replicable audit trail that helps reduce selection bias in a fragmented body of literature; and third, to offer a common structure that improves the clarity and usefulness of the review. More broadly, this study aims to map the methodological landscape of IDLE research, identify prevailing research practices, and offer practical recommendations for improving rigor, innovation, and methodological transparency in future studies. Specifically, it addresses the following research questions:

RQ1. What research designs are most commonly used in IDLE research?

RQ2. What research instruments are most frequently employed in IDLE research?

RQ3. What data analysis methods are most frequently used in IDLE research?

2. Methodology

2.1 Data sources and search strategies

We searched *Web of Science* and *Scopus* for IDLE-related studies published between January 1, 2015, and February 6, 2025, with the search conducted in early February 2025. Given the field's persistent "jingle-jangle" issue – where overlapping constructs are used interchangeably (Kelley, 1927; Kusyk et al., 2025) – we intentionally adopted a comprehensive and inclusive search strategy. Our keyword list encompassed a wide range of related terms, including "language learning beyond the classroom," "LBC," "informal language learning," "autonomous language learning," "self-directed language learning," "extramural English," "EE," "naturalistic CALL," "CALL in the digital wild," "Language learning in digital wild," "informal language learning with technology," "online informal language learning," "informal digital language learning," "online informal learning of English," "OILE," "informal digital learning of English," and "IDLE."

2.2 Inclusion and exclusion criteria

The initial database search identified 1,412 articles, which was reduced to 1,079 after duplicates were removed. To maintain focus and relevance, we included studies that (a) explicitly examined informal digital English learning, (b) reported original empirical data, (c) were written in English, and (d) were published in journals indexed in the *Arts & Humanities Citation Index* (A&HCI) or *Social Sciences Citation Index* (SSCI). For example, studies referring to "informal language learning with technology" were excluded if they focused on LOTE, such as French. By contrast, studies framed as "extramural English" were included when they involved a clear digital component, even if they also considered offline activities (Guo & Lee, 2023). We also excluded non-empirical publications, such as reviews, conceptual papers, and commentaries. The corpus

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

was limited to A&HCI/SSCI-indexed journals to support transparency and comparability, as these indexes apply relatively consistent selection and peer-review standards and thus provide a more stable and auditable sampling frame for a methodological review (see Figure 1).

2.3 Operationalizing the "jingle-jangle" problem in screening

Because terms such as IDLE, EE, and learning beyond the classroom are used inconsistently in the literature, we did not base inclusion solely on titles or keywords. Given the scope of this methodological review, we also do not attempt to resolve broader definitional debates, such as those concerning intentionality, frequency or duration thresholds, or passive versus active engagement, which have been discussed elsewhere (e.g., Arndt & Kusyk, 2026).

Instead, during abstract and full-text screening, we applied a content-based criterion aligned with our working definition. Studies were considered eligible if they examined informal English learning that involved a clearly identifiable digital or technology-mediated component taking place primarily outside the classroom, regardless of whether the authors referred to it as IDLE or EE. This meant that (a) studies on EE were included if they examined digital activities such as online gaming, social media, streaming, online chatting, app-based practice, or, where relevant, AI-mediated interaction, even when offline activities were also reported; (b) studies focusing only on offline exposure were excluded; and (c) studies in which digital activity was embedded within formal instruction, such as required or assessed platform use, were excluded as formal digital learning. Borderline cases were discussed among the authors and judged using a practical criterion: whether the study provided enough information about the digital component through its measures and/or findings to allow meaningful extraction and coding of its design, instruments, and analysis.

Applying these criteria, we excluded 831 irrelevant or non-English records at the abstract-screening stage. A full-text review of 242 papers led to further exclusions, including studies on LOTE, formal-only studies, and studies without a digital component, as well as non-empirical publications and papers published outside A&HCI/SSCI-indexed journals. The final corpus consisted of 104 studies for in-depth analysis (see Appendix A in the supplementary material).

2.4 Data extraction and analysis

Each study was coded across three main categories: design (quantitative, qualitative, or mixed methods, including subtypes), instruments (e.g., questionnaire, test, interview, diary/artifact, observation), and analysis (e.g., correlation, regression, structural equation modeling [SEM], thematic analysis, content analysis, discourse analysis). For studies using hybrid designs, we applied a clear coding rule: we identified the overall design logic (e.g., explanatory sequential or convergent) while also coding each instrument and analytic technique at the level at which it was reported (see Appendix B in the supplementary material). For example, a study that combined a survey with learner diaries, and used the diary data to help explain statistical results, was coded as explanatory sequential mixed methods. In that case, "questionnaire" and "diary/personal artifact" were coded as instruments, and the quantitative and qualitative analyses were coded separately.

Using the finalized scheme, the first author coded all included studies. To assess inter-coder reliability, the second author independently coded a randomly selected 10% of the dataset. Agreement between the two coders was high, with a Cohen's kappa of 0.83 (McHugh, 2012). The full set of coded data was then reviewed and confirmed by a third author with expertise in IDLE and CALL. Finally, we calculated the frequencies of the different methodological approaches. This quantitative synthesis helped identify dominant trends and emerging patterns, providing an overview of methodological practices in IDLE research.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

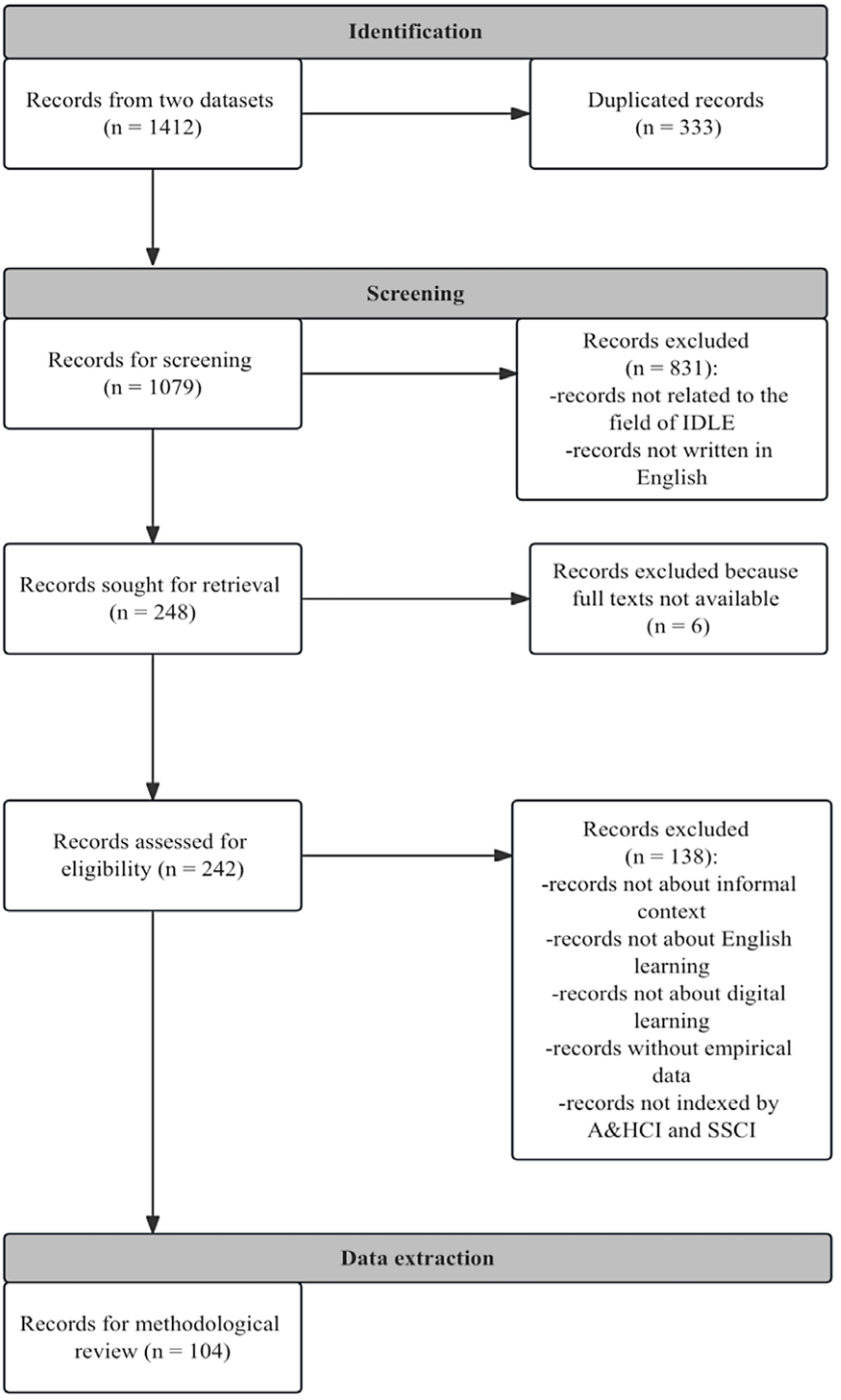


Figure 1. PRISMA flow diagram for data extraction.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Table 1. Categorization of non-experimental research designs

Category	Subcategory	Frequency ($N = 98$)
Quantitative ($n = 48$; 48.98%)	Correlational design	39 (39.8%)
	Causal-comparative design	5 (5.1%)
	Repeated-measures design	2 (2.04%)
	Longitudinal design	2 (2.04%)
Mixed methods ($n = 32$; 32.65%)	Explanatory sequential mixed methods	22 (22.45%)
	Convergent mixed methods	8 (8.16%)
	Exploratory sequential mixed methods	2 (2.04%)
Qualitative ($n = 18$; 18.37%)	Case study	9 (9.18%)
	Descriptive method	6 (6.12%)
	Ethnography	2 (2.04%)
	Phenomenology	1 (1.02%)

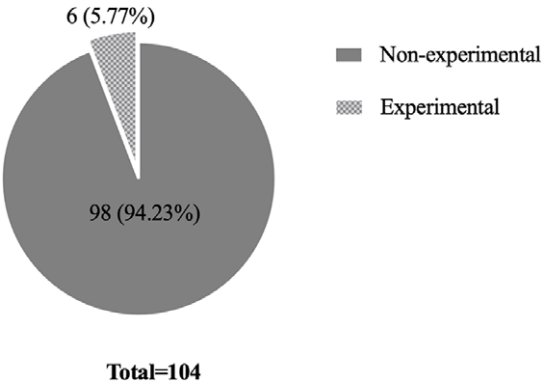


Figure 2. Research design types.

3. Findings

3.1 Research designs

The overwhelming majority of IDLE research relies on non-experimental designs, representing 94.23% (98 out of 104) of the studies reviewed (see Figure 2). Experimental research remains scarce, with only six studies – three randomized and three quasi-experimental – attempting causal investigation.

Among the 98 non-experimental studies, three primary approaches emerged (see Table 1): (a) quantitative designs, used in nearly half the studies ($n = 48$; 48.98%); (b) mixed-methods designs, combining qualitative and quantitative inquiry ($n = 32$; 32.65%); and (c) qualitative designs, employed in a smaller portion ($n = 18$; 18.37%).

3.1.1 Quantitative designs

Within the quantitative research, four major subtypes were identified. First, correlational studies dominated ($n = 39$; 39.8%), exploring relationships between IDLE engagement and outcomes like vocabulary growth and intercultural competence (Liu, Ma, et al., 2025; Rezai et al., 2026) or

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

examining predictors of IDLE participation, such as the "ideal L2 self" or "enjoyment" (Liu, Darvin, & Ma, 2024). Second, causal-comparative designs ($n = 5$; 5.1%) examined whether differing IDLE experiences led to distinct language outcomes. For example, Calafato and Clausen (2025) examined digital gaming among Norwegian students and found that language references, inferencing, and note-taking behaviors correlated with productive vocabulary, while only inferencing predicted receptive vocabulary. Third, repeated-measures designs ($n = 2$; 2.04%) traced the same learners over time to monitor fluctuations in IDLE activity. For example, Lee and Chiu (2024) collected experience-sampling data five times per day over two weeks, revealing that more frequent productive IDLE activities was associated with higher students' willingness to communicate. Lastly, longitudinal designs ($n = 2$; 2.04%) were rare. Cadierno et al. (2022) followed Danish children for a year and demonstrated that sustained exposure to English beyond the classroom was associated with gains in receptive language skills.

3.1.2 Mixed-methods designs

Explanatory sequential, convergent, and exploratory sequential designs differed not only in timing but also in how the two strands were brought together. In explanatory sequential studies ($n = 22$; 22.45%), quantitative findings typically informed the qualitative phase. For example, statistical patterns were used to shape interview protocols and/or participant selection, and the qualitative findings were then used to help explain or add context to those results (e.g., Gao et al., 2025).

In convergent studies ($n = 8$; 8.16%), the two datasets were brought together during interpretation through direct comparison or triangulation. For instance, app-use statistics were considered alongside learner reflections to support or refine the same claims (e.g., Rosell-Aguilar, 2018). In exploratory sequential studies ($n = 2$; 2.04%), the qualitative phase came first and informed the later quantitative stage, such as when themes from diaries or interviews were used to develop constructs, items, or hypotheses for subsequent testing (e.g., Lai et al., 2022).

3.1.3 Qualitative research

Qualitative studies in the corpus fell into four broad approaches. Case studies were the most common ($n = 9$; 9.18%). These were coded when the research focused on a clearly bounded case, such as a learner, family, class, program, or community, and examined it in depth, often using multiple sources to build a context-sensitive understanding. For example, Soyoof et al. (2025) examined Iranian families to explore how cultural and familial dynamics shape maternal scaffolding of children's IDLE engagement.

Qualitative descriptive studies ("descriptive method"; $n = 6$; 6.12%) were coded when the study aimed to provide a relatively close, low-inference account of learners' IDLE experiences in their everyday terms, without being organized around a bounded case. In our coding, this emphasis on staying close to participants' accounts was used as an indicator of analytic orientation or level of abstraction, rather than as a judgment of theoretical sophistication or study quality. For instance, Lee (2023) described how learners used emerging social media platforms for informal language learning, focusing on reported experiences rather than a case-based structure.

Less commonly, ethnography appeared in two studies (2.04%), with attention to community practices and interactional norms (e.g., Vazquez-Calvo, 2021), while phenomenology appeared in one study (1.02%), focusing on learners' lived experiences with GenAI tools such as ChatGPT (Li et al., 2024).

3.2 Research instruments

As shown in Figure 3, among the 193 research tools identified, quantitative instruments dominated, appearing in 118 cases (61.14%), while qualitative tools were used less often, in 75 instances (38.86%).

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Table 2. Categorization of quantitative instruments

Category	Subcategory	Frequency ($N = 118$)
Self-report ($n = 83$; 70.34%)	Questionnaire	82 (69.49%)
	Self-assessment	1 (0.85%)
Objective measure ($n = 33$; 27.97%)	Researcher-constructed test	27 (22.88%)
	Standardized test	4 (3.39%)
	Course-based test	2 (1.69%)
Computer tracking ($n = 2$; 1.69%)	–	2 (1.69%)

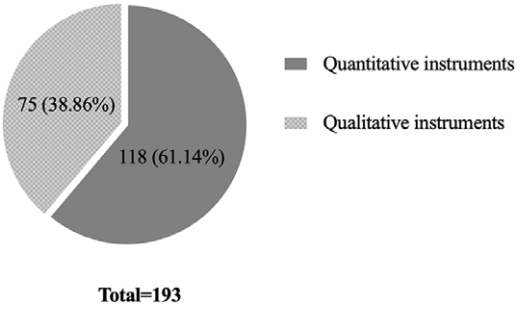


Figure 3. Research instrument types.

3.2.1 Quantitative instruments

Of the 118 quantitative tools, self-report measures were by far the most prevalent, used in 83 instances (70.34%). Objective measures followed with 33 occurrences (27.97%). Only two studies (1.69%) employed computer tracking (see Table 2).

Within the subcategories, questionnaires stood out as the most frequently used tool ($n = 82$; 69.49%) to collect three main types of data: demographic information, IDLE behaviors, and variables related to antecedents or outcomes of IDLE. Only one study utilized self-assessment, in which participants rated their English proficiency based on the Common European Framework of Reference for Languages (Jurkovič, 2019).

Researcher-constructed tests ($n = 27$; 22.88%) were typically tailored to specific contexts and designed to evaluate English skills (e.g., Tam & Reynolds, 2023). Standardized tests, like the English section of a nationwide entrance exam (Ghorbani & Golparvar, 2020), were less common ($n = 4$; 3.39%). Course-based assessments, including university English exams measuring learning outcomes (Ghorbani & Golparvar, 2020), were rare ($n = 2$; 1.69%).

A small number of studies employed computer tracking ($n = 2$; 1.69%), leveraging digital systems to trace learners' online English activities. For example, Zhao et al. (2024) employed a web platform to capture participants' task scores and behavioral metrics like revision frequency. Such indicators can provide useful evidence of interaction within a digital environment, although they should not be interpreted on their own as direct evidence of language development.

3.2.2 Qualitative instruments

Among the 75 qualitative tools, interviews were the most widely used ($n = 46$; 61.33%). Artifact analysis followed with 15 instances (20%), while open-ended questions appeared in nine studies (12%). Observation was the least common, used in only five studies (6.67%; see Table 3).

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Table 3. Categorization of qualitative instruments

Category	Subcategory	Frequency ($N = 75$)
Interview ($n = 46$; 61.33%)	Individual interview	40 (53.33%)
	Stimulated-recall interview	5 (6.67%)
	Focus group interview	1 (1.33%)
Artifact ($n = 15$; 20%)	Personal artifact	9 (12%)
	Digital artifact	4 (5.33%)
	Organizational artifact	2 (2.67%)
Open-ended question ($n = 9$; 12%)	–	9 (12%)
Observation ($n = 5$; 6.67%)	–	5 (6.67%)

Within the interview category, individual interviews predominated ($n = 40$; 53.33%), serving as the primary means of gathering in-depth insights into learners' IDLE experiences and perspectives. For example, Tajeddin et al. (2024) used individual interviews to explore how various out-of-class online activities supported English learning among Iranian adolescents. Stimulated-recall interviews ($n = 5$; 6.67%) engaged participants with visual prompts to recall past experiences (Gass & Mackey, 2000). Soyoof et al. (2025) employed videos to stimulate children's perceptions of past IDLE activities. Only one study (1.33%) employed focus group interview. Jensen (2019) gathered students who engaged in informal learning with friends to research peer-based language learning.

Personal artifacts were employed in nine studies (12%), such as diaries (Lai et al., 2022), journal entries (Lee & Roger, 2023), or demographic profiles (Soyoof et al., 2025). Digital artifacts appeared in four studies (5.33%), involving social media content such as posts, comments, or messages (e.g., Lyu & Lai, 2024). Organizational artifacts, which refer to documents created by or for institutions such as teacher records (Brevik & Holm, 2023), were the least frequent ($n = 2$; 2.67%).

Open-ended questions ($n = 9$; 12%) were used to elicit participants' narrative responses. For instance, Dizon (2026) employed written open-ended responses to explore learners' experiences and attitudes toward using ChatGPT for English learning, thereby treating AI use as one form of IDLE experience rather than as evidence of learning in itself. Lastly, observation appeared in only five studies (6.67%). Observations in these studies were conducted indirectly through video recordings or screenshots, enabling detailed post hoc analysis while minimizing disruption (Brevik & Holm, 2023).

3.3 Data analysis strategies

Across the 285 data analysis methods identified, quantitative methods overwhelmingly dominated ($n = 225$; 78.95%), while qualitative methods made up a smaller portion ($n = 60$; 21.05%; see Figure 4).

3.3.1 Quantitative data analysis

Five main quantitative techniques emerged (see Table 4). Correlational analysis was the most frequently used ($n = 94$; 41.78%). Next came descriptive statistics ($n = 87$; 38.67%), followed by

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Table 4. Categorization of quantitative data analysis strategies

Category	Subcategory	Frequency ($N = 225$)
Correlation ($n = 94$; 41.78%)	Correlation analysis	28 (12.44%)
	Regression analysis	28 (12.44%)
	Structural equation modeling	38 (16.89%)
Descriptive statistics ($n = 87$; 38.67%)	–	87 (38.67%)
Group comparison ($n = 36$; 16%)	ANOVA	15 (6.67%)
	t tests	11 (4.89%)
	Mann–Whitney *U* test	5 (2.22%)
	Kruskal–Wallis test	2 (0.89%)
	MANOVA	3 (1.33%)
Factor analysis ($n = 6$; 2.67%)	–	6 (2.67%)
Cluster analysis ($n = 2$; 0.89%)	–	2 (0.89%)

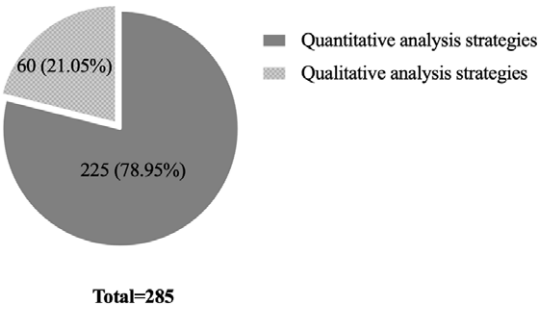


Figure 4. Data analysis strategies.

group comparison ($n = 36$; 16%). By contrast, factor analysis ($n = 6$; 2.67%) and cluster analysis ($n = 2$; 0.89%) appeared only occasionally, suggesting their specific research use.

Descriptive statistics offered a snapshot of central tendencies and variability, serving as the groundwork for more complex analyses. Correlation analysis ($n = 28$; 12.44%) explored the strength and direction of relationships between IDLE and related constructs. For example, Rezai et al. (2026) found a strong positive correlation between IDLE and vocabulary knowledge among Iranian EFL learners. Regression analysis ($n = 28$; 12.44%) helped reveal directional effects. Tsang (2026) demonstrated that English exposure beyond the classroom could significantly predict children's English proficiency in Hong Kong. SEM, used in 38 studies (16.89%), allowed researchers to examine complex relationships and latent variables. For example, by modeling the complex relation between IDLE and *online motivation for learning* (OML), *online foreign language enjoyment* (OFLE), and *telecollaborative foreign language anxiety* (TFLA), Rezai et al. (2025) found that IDLE is significantly associated with OML, OFLE, and TFLA.

Group comparison techniques, such as ANOVA ($n = 15$; 6.67%) and *t* tests ($n = 11$; 4.89%), were frequently applied to compare learners with and without IDLE experiences (e.g., Guan et al.,

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Table 5. Categorization of qualitative data analysis strategies

Category	Subcategory	Frequency ($N = 60$)
Thematic analysis ($n = 25$; 41.67%)	–	25 (41.67%)
Content analysis ($n = 11$; 18.33%)	–	11 (18.33%)
Linguistic analysis ($n = 2$; 3.33%)	Syntactic-level	1 (1.67%)
	Lexical-level	1 (1.67%)
Discourse analysis ($n = 2$; 3.33%)	–	2 (3.33%)
Others ($n = 20$; 33.33%)	–	20 (33.33%)

2025) or among learners with differing IDLE exposure (e.g., Tsang & Lam, 2025). When normality assumptions were violated, researchers used nonparametric tests like the Mann–Whitney *U* ($n = 5$; 2.22%) and Kruskal–Wallis ($n = 2$; 0.89%). MANOVA ($n = 3$; 1.33%) enabled simultaneous comparison of multiple dependent variables, as in Leona et al. (2021), who compared motivation, vocabulary, and English exposure between learners with and without formal instruction.

Factor analysis ($n = 6$; 2.67%) was used to uncover underlying structures influencing observed variables, such as identifying factors of EE engagement types (e.g., gaming and social interaction) in Sundqvist and Uztosun (2024). Cluster analysis ($n = 2$; 0.89%) classified learners into typologies, as in Lee and Xie (2023), who discovered five learner profiles: minimal, maximal, gaming-entertainment, entertainment, and entertainment-socializing.

3.3.2 Qualitative data analysis strategies

Four main approaches characterized qualitative data analysis (see Table 5). Thematic analysis was the most common ($n = 25$; 41.67%) and was used to identify recurring patterns in learners' experiences (Braun & Clarke, 2006). For example, Liu, Darvin, and Ma (2025) examined learners' use of AI in IDLE and identified themes such as a greater willingness to use AI as a conversational partner. These findings illustrate learners' perceptions and interactional practices with AI, rather than demonstrating language development directly. Content analysis ($n = 11$; 18.33%) was also used to organize data into categories and identify patterns (Krippendorff, 2019). For instance, Rezai et al. (2024a) grouped Iranian teachers' perspectives into ecological layers – personal, micro-, macro-, and chrono-systems – that shaped IDLE engagement. Linguistic analysis ($n = 2$; 3.33%) focused on structural and lexical features of language. Soyoof et al. (2025), for example, examined sentence complexity and content words to explore how IDLE relates to children's language development. Discourse analysis ($n = 2$; 3.33%) was used to examine meaning-making through language and multimodal interaction. For example, Lee et al. (2019) used multimodal discourse analysis to study how learners employed technology and gestures to negotiate meaning in informal communication.

A further point to note is that one-third of the studies were grouped into the "Others" category because they did not clearly specify a recognizable qualitative analytic approach ($n = 20$; 33.33%). In our corpus, these less clearly reported cases appeared more often in earlier publications (pre-2019) and were especially common in explanatory sequential mixed-methods studies, where interview or open-ended data were added after an initial survey phase. For example, one

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

anonymized mixed-methods study noted that interview data were "transcribed, coded, and synthesized" to supplement the quantitative findings, but did not describe the analytic framework or key procedures, such as whether coding was inductive or deductive, how codes were developed, or how themes or categories were generated. We return to this pattern and its implications in the limitations and recommendations sections.

4. Discussion

The findings of this methodological review suggest a clear methodological tendency in IDLE research. The strong predominance of non-experimental designs (94.23%) shows that the field has been shaped mainly by *exploratory* and *correlational* approaches (Lee, 2022a; Sundqvist, 2024). This pattern is understandable, given the nature of the phenomenon itself. Informal language learning is often nonlinear, dynamic, and highly individualized, taking place in unstructured, out-of-school settings (Lee, 2026). Because of these features, controlled experiments or interventions can be difficult to implement. As a result, cross-sectional surveys and qualitative methods, such as interviews, have often been more suitable for mapping the phenomenon and developing an initial understanding of it (Sockett, 2013; Soyoof et al., 2025).

Although these dominant approaches have been useful for identifying consistent *associations* between IDLE engagement and learning outcomes (Kusyk et al., 2025), their concentration also brings some limitations. With relatively few experimental or longitudinal studies, the field still has only limited capacity to make causal claims or further develop explanatory theory (Burston et al., 2024; Choubsaz et al., 2024; Lee, 2026). As a result, it remains unclear whether IDLE itself contributes to development or whether it tends to occur alongside factors such as motivation and contextual advantage (Ballance, 2024; Lee, 2022a).

This pattern also creates a practical challenge. Cross-sectional and qualitative designs are well suited to capturing the complexity of informal learning in natural settings (Sockett, 2013), but decision-makers, such as policymakers and school leaders, often need stronger forms of evidence when considering institutional investment or curricular integration (Lee, 2026). For example, longitudinal or quasi-experimental evidence can provide a firmer basis than correlation alone for decisions about allocating curriculum time to guided out-of-class tasks, adopting school-supported digital resources including, where appropriate, AI speaking tools, or expanding a pilot beyond a single grade (Creswell & Creswell, 2023; Taherian et al., 2024). In a similar way, combining self-reports with behavioral or trace data, such as platform analytics, learner-exported activity records, or carefully interpreted AI interaction logs, may help educators identify what should be monitored and supported in practice, including spacing, feedback uptake, and revision routines (Lee, 2026; Sundqvist, 2024). However, AI logs should be understood as records of AI-mediated activity co-produced by learners and systems, not as transparent records of learner language development alone. Clearer reporting of sampling and subgroup differences may also help inform more equity-oriented policies, such as device or data subsidies and differentiated support (Lee, 2026).

Given the review's focus on methodological patterns, the strong reliance on self-reported instruments, especially questionnaires, is understandable, although it also raises some questions about validity (Creswell & Creswell, 2023; Sundqvist, 2024). Questionnaires appear to be widely used not only because they are practical but also because they align closely with the field's most common research aims (Arndt & Rose, 2023). In many studies, they are used to operationalize IDLE engagement, most often through learners' self-reported frequency or duration of specific digital activities (Arndt & Kusyk, 2026). They are also commonly used to measure antecedents or individual difference factors, such as motivation, enjoyment, ideal L2 self, and perceived usefulness, as well as outcomes, often in the form of learners' perceived gains in areas such as vocabulary, speaking confidence, or intercultural competence rather than performance-based

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

measures (Lee, 2026). This pattern is consistent with an evidence base that is largely correlational and often oriented toward SEM, where survey instruments are used to examine pathways linking engagement, motivational variables, and outcomes (Liu, Ma, et al., 2025).

At the same time, this instrument profile may help explain why our findings can sometimes seem somewhat "catalog-like." When most studies rely on cross-sectional survey data to examine associations, the resulting evidence is naturally strongest for perceived engagement and self-reported outcomes (Arndt & Rose, 2023; Lee, 2026). As questionnaires are susceptible to recall error and social desirability bias, some widely cited relationships may reflect learners' retrospective perceptions more than their actual behavior in digital environments (Creswell & Creswell, 2023). Another related issue concerns construct coverage (Arndt & Kusyk, 2026). Many IDLE scales capture how often or how long learners engage, but provide less insight into how that engagement unfolds, such as in terms of interaction, attention, or feedback uptake, even though different activity types may place quite different cognitive and linguistic demands on learners (Sundqvist, 2024). These limitations do not diminish the value of survey-based research; rather, they suggest ways in which the evidence base can be strengthened (Lai & Sundqvist, 2026; Lee, 2026). Interviews can provide added depth (Jensen, 2019), and self-reported data may become more convincing when combined with additional behavioral indicators, such as platform analytics, learner-exported logs, or screen-tracking data, which are still relatively uncommon in the corpus (Arndt & Rose, 2023; Lee, 2026). These indicators can help document patterns of engagement, but they vary considerably across activity types and should not be assumed to carry the same meaning in gaming, video viewing, podcast listening, online chatting, app-based practice, or AI-mediated interaction. Additionally, trace data also bring ethical and practical challenges, including issues of privacy, ownership, platform policy, interpretability, and learner consent (O'Grady et al., 2026). In some IDLE contexts, such as gaming, streaming, podcast listening, or online chatting, collecting meaningful and ethically defensible trace data may be difficult or even inappropriate. They may also reduce ecological validity when they depend on intrusive monitoring or highly controlled conditions (Arndt & Rose, 2023; Jensen, 2019; Lee, 2026). For this reason, consent-based and minimally intrusive approaches, such as participant-mediated exports and aggregated indicators, may offer a more realistic balance for IDLE research (Lee, 2022a; Sundqvist, 2024).

Also, our corpus shows meaningful variation within the broad correlational category, ranging from simple bivariate correlations to regression-based modeling and more complex SEM applications. The growing use of techniques such as SEM may indicate increasing methodological development, as these approaches can model latent constructs and examine more complex theoretical relationships (Liu, Ma, et al., 2025). Even so, techniques such as SEM cannot by themselves overcome the limitations of cross-sectional, perception-based designs, and their contribution to causal explanation therefore remains limited (Lee, 2026). On the qualitative side, there is also room for greater methodological transparency, as one-third of the studies did not fully specify their analytical procedures, which may affect judgments about replicability (Creswell & Creswell, 2023). Taken together, these patterns add nuance to the broader point about methodological concentration: despite differences in analytic sophistication, many studies still rely on similar cross-sectional survey logics, with less attention given to mechanisms, development over time, and contextual complexity (Sundqvist, 2024).

5. Implications

This review suggests that the field still shows methodological homogeneity, with a recurring reliance on cross-sectional, correlational, self-report designs. These approaches have helped identify broad patterns, but they are less suited to explaining how IDLE works, for whom, and under what conditions. To strengthen explanatory claims, future research should better align its questions with designs that capture (a) what learners actually do in digital spaces, (b) how

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

engagement changes over time, and (c) which mechanisms plausibly connect engagement to development (Choubsaz et al., 2024).

A practical starting point is clearer construct reporting. Future studies could specify (a) boundary conditions (what counts as IDLE), (b) intentionality assumptions (incidental exposure vs. goal-directed practice), and (c) activity structure (unstructured use vs. app-based progression or feedback). A brief "IDLE eligibility checklist" aligned with Benson's (2011) dimensions could reduce definitional drift and improve comparability across studies. Because "time" is not equivalent across engagement modes, instruments could also move beyond frequency and duration by adding a small set of engagement-quality indicators, such as interaction, language density, and feedback uptake, while separating activity type from engagement mode (passive vs. active).

To make longitudinal work more feasible in informal settings, researchers could use short micro-longitudinal or burst designs that balance ecological validity with participant burden, for example, six to eight weeks of weekly check-ins or 14 days of brief daily diaries repeated at two or three points in a semester. A minimal protocol could combine a two- to three-minute end-of-day mobile diary (activity type, minutes, perceived challenge, and one "learning moment") with a weekly short outcome task, such as a voice note or brief writing prompt. This keeps data collection light while allowing researchers to examine within-person change and spacing patterns over time.

Learning analytics and, where appropriate, AI-related data sources can also be integrated into IDLE studies in transparent, low-burden ways. One option is participant-mediated analytics, in which learners periodically export and donate selected traces they control rather than installing invasive trackers. For example, participants could share weekly exports of AI-chat transcripts or revision histories, simple dashboard indicators such as screen-time breakdowns, or, in AI-mediated IDLE contexts, carefully contextualized chat transcripts, together with brief periodic performance checks such as a monthly speaking sample. Importantly, AI transcripts and usage metrics should be treated as records of AI-mediated interaction, not as direct evidence that learners have acquired new language knowledge or communicative ability. AI-based analytics may then support theory-driven process coding, such as identifying feedback moves, uptake, and strategy use, but such coding should account for the fact that AI systems actively shape interaction through topic suggestions, feedback style, simplification, and response design. These data should therefore be interpreted alongside other evidence, including learner reflections, task performance, interactional context, and longitudinal change.

From an applied perspective, more rigorous designs can support clearer educational decisions. For educators, distinguishing "time on English" from quality of engagement can inform task design by emphasizing interaction, feedback uptake, and retrieval practice rather than passive exposure. For policymakers and program leaders, evidence combining trace-informed engagement indicators with repeated outcome measures can better justify investment in digital resources, including when and how AI-supported practice may contribute to measurable gains over a term, for which learners, and under what conditions, while clarifying what data can be collected ethically and with minimal intrusion. Finally, frameworks attuned to nonlinear development, such as complex dynamic systems theory, may help interpret variability and context sensitivity in these datasets and strengthen theory building in IDLE (Larsen-Freeman, 2019).

6. Limitations

This review has several limitations. First, our inclusion criteria may have introduced sampling bias. By focusing on studies published in A&HCI/SSCI-indexed, English-language journals, we likely excluded relevant work published in regional or emerging venues, other languages, or book chapters. This may have overrepresented methodological norms in highly visible international outlets while underrepresenting localized or resource-sensitive innovations, especially in Global

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

South contexts (Guo & Lee, 2023). As a result, the methodological convergence reported here should be interpreted mainly as applying to this indexed corpus rather than to the full landscape of IDLE scholarship. Future reviews could broaden sampling frames and code contextual variables such as region, educational level, and age group. We did not code region or age because these details were often reported inconsistently or only implicitly, and doing so would have required inference that could reduce comparability and inter-coder reliability.

Second, definitional and operational variation in what counts as "IDLE" limits comparability across studies and the construct validity of instruments that assume equivalent "exposure" across activities. Although we applied an explicit content-based screening rule (Section 2.3), our coding still depended on how clearly primary studies reported boundaries, intentionality, and activity context.

Third, incomplete reporting in some studies limited our ability to resolve disagreements about intentionality (incidental vs. goal-directed) and borderline activity types such as structured app-based learning. Relatedly, because many questionnaires operationalize IDLE through frequency or duration, our review can show how often such measures are used, but not whether "time on task" is comparable across very different engagement modes, nor could we independently assess the ecological realism of all questionnaire items. This limitation is especially relevant to AI-mediated IDLE, where available logs, prompts, responses, and usage metrics may reflect both learner activity and system behavior, rather than learner development alone.

Fourth, although we used a systematic coding framework, categorizing complex or hybrid designs still required interpretive judgment (Creswell & Creswell, 2023), and our categories may not fully reflect each author's original methodological intent. Finally, although PRISMA provided a disciplined audit trail, its use required adaptation to the field's terminological diversity and boundary judgments (Kusyk et al., 2025). Because the review covers literature up to early 2025, it may also miss the most recent methodological developments, including emerging debates about how GenAI data should be collected, interpreted, and ethically used in IDLE research (Zou et al., 2026).

7. Conclusion

This methodological review has examined the research designs, instruments, and analytical strategies that characterize contemporary IDLE scholarship. Our findings suggest that the methodological trajectory of IDLE research reflects both its empirical strengths and areas for further development. The field shows a degree of *methodological homogeneity*, often placing more emphasis on correlation than causation, perception than behavior, and quantification than interpretation (Chong et al., 2022; Lee, 2022a). Although this pattern has been helpful in identifying broad trends, it may also narrow the epistemological range of inquiry and limit the field's ability to inform pedagogy in more nuanced ways (Lee, 2022b; Sundqvist, 2024). Future research may therefore benefit from greater methodological pluralism, including the use of mixed methods, longitudinal designs, and experimental approaches, to better connect self-perceptions with behavioral evidence and statistical patterns with narrative depth (Guan et al., 2025; Taherian et al., 2024). Such diversification may help IDLE research move beyond descriptive mapping toward a more explanatory and theoretically integrated understanding of informal language learning. At the same time, methodological diversification should not be reduced to the adoption of AI-generated data or easily available digital metrics. Future IDLE research should continue to recognize the diversity of informal digital practices and treat AI-related evidence as one complementary, carefully interpreted source within a broader methodological toolkit.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344026100603

Data availability statement. Data are available upon request.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Authorship contribution statement. Xian Liu: Data curation, Formal analysis, Visualization, Writing – original draft; XinWei Pan: Data curation, Formal analysis; Ju Seong Lee: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Funding disclosure statement. This research did not receive any specific funding.

Competing interests statement. The authors declare no competing interests.

Ethical statement. Ethical approval was not required.

GenAI use disclosure statement. The authors used AI (https://www.deepseek.com) to improve readability. The authors take full responsibility for the content of the publication.

References

Arndt, H. L., & Kusyk, M. (2026). Informal second language learning. *Language Teaching*. Advance online publication. https://doi.org/10.1017/S0261444826101219

Arndt, H., & Rose, H. (2023). Capturing life as it is truly lived? Improving diary data in educational research. *International Journal of Research & Method in Education*, 46(2), 175–186. https://doi.org/10.1080/1743727x.2022.2094360

Ballance, O. J. (2024). Sampling and randomisation in experimental and quasi-experimental CALL studies: Issues and recommendations for design, reporting, review, and interpretation. *ReCALL*, 36(1), 58–71. https://doi.org/10.1017/s0958344023000162

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–16). Palgrave Macmillan. https://doi.org/10.1057/9780230306790_2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brevik, L. M., & Holm, T. (2023). Affinity and the classroom: Informal and formal L2 learning. *ELT Journal*, 77(1), 83–93. https://doi.org/10.1093/elt/ccac012

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press. https://doi.org/10.2307/j.ctv26071r6

Burston, J., Athanasiou, A., & Giannakou, K. (2024). Quantitative experimental L2 acquisition MALL studies: A critical evaluation of research quality. *ReCALL*, 36(1), 22–39. https://doi.org/10.1017/s0958344023000149

Cadierno, T., Hansen, M. B., Lauridsen, J. T., Eskildsen, S. W., Fenyvesi, K., & Hannibal Jensen, S. (2022). A multi-factor study of the development of English receptive skills by young Danish children. *Bilingualism: Language and Cognition*, 25(4), 566–578. https://doi.org/10.1017/s1366728921001085

Calafato, R., & Clausen, T. (2025). Vocabulary learning strategies in extramural English gaming and their relationship with vocabulary knowledge. *Computer Assisted Language Learning*, 38(8), 1825–1843. https://doi.org/10.1080/09588221.2024.2328023

Chong, S. W., Lin, T. J., & Chen, Y. (2022). A methodological review of systematic literature reviews in higher education: Heterogeneity and homogeneity. *Educational Research Review*, 35, Article 100426. https://doi.org/10.1016/j.edurev.2021.100426

Chong, S. W., & Reinders, H. (2021). A methodological review of qualitative research syntheses in CALL: The state-of-the-art. *System*, 103, Article 102646. https://doi.org/10.1016/j.system.2021.102646

Choubsaz, Y., Jalilifar, A., & Boulton, A. (2024). A longitudinal analysis of highly cited papers in four CALL journals. *ReCALL*, 36(1), 40–57. https://doi.org/10.1017/s0958344023000137

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.

Dizon, G. (2026). ChatGPT as a tool for self-directed foreign language learning. *Innovation in Language Learning and Teaching*, 20(2), 334–350. https://doi.org/10.1080/17501229.2024.2413406

Fang, F., Meng, Y., Tang, L., & Cui, Y. (2025). The impact of informal digital learning of English (IDLE) on EFL learners' engagement: Mediating roles of flow, online self-efficacy, and behavioral intention. *Behavioral Sciences*, 15(7), Article 851. https://doi.org/10.3390/bs15070851

Fu, R. (2025). Enjoyment and grit as mediators between informal digital learning English, basic psychology needs, and willingness to communicate among English as a foreign language university students. *Current Psychology*, 44(3), 1660–1674. https://doi.org/10.1007/s12144-024-07220-w

Gao, Y., Wang, X., & Reynolds, B. L. (2025). The mediating roles of resilience and flow in linking basic psychological needs to tertiary EFL learners' engagement in the informal digital learning of English: A mixed-methods study. *Behavioral Sciences*, 15(1), Article 85. https://doi.org/10.3390/bs15010085

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Lawrence Erlbaum Associates.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Ghorbani, M. R., & Golparvar, S. E. (2020). Modeling the relationship between socioeconomic status, self-initiated, technology-enhanced language learning, and language outcome. *Computer Assisted Language Learning*, 33(5–6), 607–627. https://doi.org/10.1080/09588221.2019.1585374

Guan, L., Zhang, E. Y., & Gu, M. M. (2025). Examining generative AI–mediated informal digital learning of English practices with social cognitive theory: A mixed-methods study. *ReCALL*, 37(3), 315–331. https://doi.org/10.1017/s0958344024000259

Guo, X., & Lee, J. S. (2023). A systematic review of informal digital learning of English: An ecological systems theory perspective. *System*, 117, Article 103097. https://doi.org/10.1016/j.system.2023.103097

Jensen, S. H. (2019). Language learning in the wild: A young user perspective. *Language Learning & Technology*, 23(1), 72–86. https://doi.org/10.64152/10125/44673

Jurkovič, V. (2019). Online informal learning of English through smartphones in Slovenia. *System*, 80, 27–37. https://doi.org/10.1016/j.system.2018.10.007

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications. https://doi.org/10.4135/9781071878781

Kusyk, M., Arndt, H. L., Schwarz, M., Yibokou, K. S., Dressman, M., Sockett, G., & Toffoli, D. (2025). A scoping review of studies in informal second language learning: Trends in research published between 2000 and 2020. *System*, 130, Article 103541. https://doi.org/10.1016/j.system.2024.103541

Lai, C., Liu, Y., Hu, J., Benson, P., & Lyu, B. N. (2022). Association between the characteristics of out-of-class technology-mediated language experience and L2 vocabulary knowledge. *Language Learning & Technology*, 26(1), 1–24. https://doi.org/10.64152/10125/73485

Lai, C., & Sundqvist, P. (2026). Research agenda: Synergizing in-class and out-of-class language learning with technology. *Language Teaching*, 59(1), 60–81. https://doi.org/10.1017/s026144482500014x

Larsen-Freeman, D. (2019). Foreword: Taking the next step. In P. Hiver & A. H. Al-Hoorie (Eds.), *Research methods for complexity theory in applied linguistics* (pp. vii–x). Multilingual Matters. https://doi.org/10.21832/9781788925754-001

Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System*, 81, 26–38. https://doi.org/10.1016/j.system.2018.12.013

Lee, J. S. (2022a). Evaluation of instruments for researching learners' LBC. In H. Reinders, C. Lai, & P. Sundqvist (Eds.), *The Routledge handbook of language learning and teaching beyond the classroom* (pp. 312–326). Routledge. https://doi.org/10.4324/9781003048169-26

Lee, J. S. (2022b). *Informal digital learning of English: Research to practice*. Routledge. https://doi.org/10.4324/9781003043454

Lee, J. S. (2026). *Informal digital learning of English*. Cambridge University Press. https://doi.org/10.1017/9781009641647

Lee, J. S., & Chiu, M. M. (2024). Modelling trait and state willingness to communicate in a second language: An experience sampling approach. *Studies in Second Language Learning and Teaching*, 14(3), 483–514. https://doi.org/10.14746/ssllt.37541

Lee, J. S., & Xie, Q. (2023). Profiling the affective characteristics of EFL learners' digital informal learning: A person-centered approach. *Innovation in Language Learning and Teaching*, 17(3), 552–566. https://doi.org/10.1080/17501229.2022.2085713

Lee, Y.-J. (2023). Language learning affordances of Instagram and TikTok. *Innovation in Language Learning and Teaching*, 17(2), 408–423. https://doi.org/10.1080/17501229.2022.2051517

Lee, Y.-J., & Roger, P. (2023). Cross-platform language learning: A spatial perspective on narratives of language learning across digital platforms. *System*, 118, Article 103145. https://doi.org/10.1016/j.system.2023.103145

Leona, N. L., van Koert, M. J. H., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of extramural English exposure and motivation. *System*, 96, Article 102402. https://doi.org/10.1016/j.system.2020.102402

Li, B., Bonk, C. J., Wang, C., & Kou, X. (2024). Reconceptualizing self-directed learning in the era of generative AI: An exploratory analysis of language learning. *IEEE Transactions on Learning Technologies*, 17, 1489–1503. https://doi.org/10.1109/tlt.2024.3386098

Liu, G. L., Darvin, R., & Ma, C. (2024). Unpacking the role of motivation and enjoyment in AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation in the Chinese context. *Computers in Human Behavior*, 160, Article 108362. https://doi.org/10.1016/j.chb.2024.108362

Liu, G. L., Darvin, R., & Ma, C. (2025). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 38(7), 1632–1660. https://doi.org/10.1080/09588221.2024.2310288

Liu, G., Ma, C., Bao, J., & Liu, Z. (2025). Toward a model of informal digital learning of English and intercultural competence: A large-scale structural equation modeling approach. *Computer Assisted Language Learning*, 38(3), 342–368. https://doi.org/10.1080/09588221.2023.2191652

Liu, G. L., Soyoof, A., Lee, J. S., & Zhang, L. J. (2025). Informal digital learning of English in Asian English as a foreign language contexts: A thematic review. *RELC Journal*. Advance online publication. https://doi.org/10.1177/00336882251332309

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Liu, G. L., Zhang, Y., & Zhang, R. (2024). Examining the relationships among motivation, informal digital learning of English, and foreign language enjoyment: An explanatory mixed-method study. *ReCALL*, 36(1), 72–88. https://doi.org/10.1017/s0958344023000204

Lyu, B., & Lai, C. (2024). Analysing learner engagement with native speaker feedback on an educational social networking site: An ecological perspective. *Computer Assisted Language Learning*, 37(1–2), 114–148. https://doi.org/10.1080/09588221.2022.2030364

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. https://doi.org/10.11613/bm.2012.031

O'Grady, S., Gokturk, N., & Rossi, O. (2026). Artificial intelligence in language assessment research: A scoping review. *Research Synthesis in Applied Linguistics*, 2(1), 103–139. https://doi.org/10.1080/29984475.2026.2633389

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

Rezai, A., Ashkani, P., & Moradian, M. R. (2025). Is informal digital learning of English significantly correlated with online motivation for learning, online foreign language enjoyment, and telecollaborative foreign language anxiety? Evidence from higher education. *International Journal of Human–Computer Interaction*, 41(18), 11775–11789. https://doi.org/10.1080/10447318.2024.2445102

Rezai, A., Soyoof, A., & Reynolds, B. L. (2024a). Ecological factors affecting students' use of informal digital learning of English: EFL teachers' perceptions. *Teaching and Teacher Education*, 145, Article 104629. https://doi.org/10.1016/j.tate.2024.104629

Rezai, A., Soyoof, A., & Reynolds, B. L. (2024b). Informal digital learning of English and EFL teachers' job engagement: Exploring the mediating role of technological pedagogical content knowledge and digital competence. *System*, 122, Article 103276. https://doi.org/10.1016/j.system.2024.103276

Rezai, A., Soyoof, A., & Reynolds, B. L. (2026). Effectiveness of informal digital learning of English on EFL learners' vocabulary knowledge: A mixed-methods investigation. *Computer Assisted Language Learning*, 39(1–2), 1–26. https://doi.org/10.1080/09588221.2024.2350419

Rød, A. J., & Calafato, R. (2025). Bridging extramural English and formal instruction. *ELT Journal*, 79(3), 363–371. https://doi.org/10.1093/elt/ccaf026

Rosell-Aguilar, F. (2018). Autonomous language learning through a mobile application: A user evaluation of the *busuu* app. *Computer Assisted Language Learning*, 31(8), 854–881. https://doi.org/10.1080/09588221.2018.1456465

Sockett, G. (2013). Understanding the online informal learning of English as a complex dynamic system: An emic approach. *ReCALL*, 25(1), 48–62. https://doi.org/10.1017/s095834401200033x

Soyoof, A., Reynolds, B. L., Neumann, M. M., & Vazquez-Calvo, B. (2025). Maternal scaffolding of Iranian children's extramural informal digital learning of English (IDLE): A qualitative study. *Early Childhood Education Journal*, 53(5), 1865–1881. https://doi.org/10.1007/s10643-024-01675-z

Soyoof, A., Reynolds, B. L., Vazquez-Calvo, B., & McLay, K. (2023). Informal digital learning of English (IDLE): A scoping review of what has been done and a look towards what is to come. *Computer Assisted Language Learning*, 36(4), 608–640. https://doi.org/10.1080/09588221.2021.1936562

Sundqvist, P. (2024). Extramural English as an individual difference variable in L2 research: Methodology matters. *Annual Review of Applied Linguistics*, 44, 79–91. https://doi.org/10.1017/s0267190524000072

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-46048-6

Sundqvist, P., & Uztosun, M. S. (2024). Extramural English in Scandinavia and Asia: Scale development, learner engagement, and perceived speaking ability. *TESOL Quarterly*, 58(4), 1638–1665. https://doi.org/10.1002/tesq.3296

Taherian, T., Shirvan, M. E., Yazdanmehr, E., Kruk, M., & Pawlak, M. (2024). A longitudinal analysis of informal digital learning of English, willingness to communicate and foreign language boredom: A latent change score mediation model. *The Asia-Pacific Education Researcher*, 33(4), 997–1010. https://doi.org/10.1007/s40299-023-00751-z

Tajeddin, Z., Hassani, A., & Bakhshandeh, N. (2024). Successful language learning out of the classroom: A case study of Iranian learners' activities. *Innovation in Language Learning and Teaching*, 18(4), 287–300. https://doi.org/10.1080/17501229.2023.2296463

Tam, H. I., & Reynolds, B. L. (2023). The relationship between extramural English engagement and the vocabulary size of L1 Cantonese speakers in Macau. *ITL - International Journal of Applied Linguistics*, 174(1), 49–82. https://doi.org/10.1075/itl.21003.tam

Tsang, A. (2026). The relationships between spoken and written input beyond the classroom and young EFL learners' proficiency. *Language Teaching Research*, 30(6), 3291–3313. https://doi.org/10.1177/13621688231211367

Tsang, A., & Lam, W.-K. (2025). De-foreignizing English outside class: Junior-secondary-level English-as-a-foreign-language learners' extramural English engagements and listening and reading proficiency. *RELC Journal*, 56(2), 377–396. https://doi.org/10.1177/00336882241229459

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press

Vazquez-Calvo, B. (2021). Guerrilla fan translation, language learning, and metalinguistic discussion in a Catalan-speaking community of gamers. *ReCALL*, 33(3), 296–313. https://doi.org/10.1017/s095834402000021x

Zhang, R., Zou, D., Cheng, G., Xie, H., Wang, F. L., & Au, O. T. S. (2021). Target languages, types of activities, engagement, and effectiveness of extramural language learning. *PloS ONE*, 16(6), Article e0253431. https://doi.org/10.1371/journal.pone.0253431

Zhang, Y., & Liu, G. L. (2025). Examining the impacts of learner backgrounds, proficiency level, and the use of digital devices on informal digital learning of English: An explanatory mixed-method study. *Computer Assisted Language Learning*, 38(5–6), 1113–1140. https://doi.org/10.1080/09588221.2023.2267627

Zhao, R., Zhuang, Y., Xie, Z., & Yu, P. L. H. (2024). Facilitating self-directed language learning in real-life scene description tasks with automated evaluation. *Computers & Education*, 219, Article 105106. https://doi.org/10.1016/j.compedu.2024.105106

Zou, M., Reinders, H., & Amjad, F. (2026). Understanding the potential role of GenAI-mediated informal digital learning of English (GenAI-IDLE) in the Global South: AI literacy, emotions, and willingness to communicate as outcomes. *ReCALL*, 38(1), 131–149. https://doi.org/10.1017/s0958344025100360

About the authors

Xian Liu is a doctoral student in the Department of English Language Education at the Education University of Hong Kong. Her work focuses on informal digital learning of English (IDLE) in less-resourced regions.

XinWei Pan is a doctoral student in the Department of English Language Education at the Education University of Hong Kong. Her work focuses on informal digital learning of English (IDLE) and informal digital learning of Chinese.

Ju Seong Lee is an associate professor in the Department of English Language Education at the Education University of Hong Kong. His work focuses on informal digital learning of English (IDLE) and its social impact.

https://doi.org/10.1017/S0958344026100603 Published online by Cambridge University Press