LEARNERS' TRUST AND RELIANCE ON GENERATIVE AI IN ENGLISH LANGUAGE LEARNING

Elizabeth Meiske Maythy Lasut[1]
[1]Pendidikan Bahasa Inggris, Universitas Klabat, Airmadidi, Indonesia
(elizabethmmlasut@unklab.ac.id)

Abstrak

Kecerdasan artifisial generatif (GenAI) semakin digunakan dalam pembelajaran bahasa Inggris, tetapi nilai pendidikannya bergantung pada cara pembelajar menilai, mempercayai, dan mengandalkan keluaran yang dihasilkan. Tinjauan sistematis ini mensintesis bukti empiris mengenai kepercayaan dan ketergantungan pembelajar pada GenAI dalam pembelajaran bahasa Inggris. Mengikuti prinsip pelaporan PRISMA 2020, Scopus ditelusuri pada 6 Oktober 2026 untuk artikel jurnal sejawat terbitan 2022–2026. Lima studi memenuhi kriteria akhir. Kepercayaan berkaitan dengan persepsi kegunaan, penggunaan berkelanjutan, niat perilaku, dan pola ketergantungan. Keluaran yang tidak akurat atau tidak sesuai konteks melemahkan kepercayaan, sedangkan kepercayaan berlebihan dapat mendorong ketergantungan berlebih, pemantauan yang berkurang, dan pembelajaran mandiri yang lebih lemah. Temuan menunjukkan bahwa tujuan pendidikan seharusnya adalah kepercayaan yang terkalibrasi, sehingga pembelajar memanfaatkan GenAI sambil tetap mengevaluasi keluaran secara kritis dan mempertahankan penilaian manusia.

Kata Kunci: *Kecerdasan Artifisial Generatif; Pembelajaran Bahasa Inggris; Kepercayaan; Ketergantungan; Kepercayaan Terkalibrasi*

Abstract

Generative artificial intelligence (GenAI) is increasingly used in English language learning, but its educational value depends on how learners judge, trust, and rely on generated outputs. This systematic review synthesizes empirical evidence on learner trust and reliance in GenAI-supported English language learning. Following PRISMA 2020 reporting principles, Scopus was searched on 6 October 2026 for peer-reviewed journal articles published from 2022 to 2026. Five studies met the final eligibility criteria. Trust was associated with perceived usefulness, continued use, behavioral intention, and patterns of reliance. Inaccurate or contextually inappropriate outputs weakened trust, while excessive trust could encourage over-reliance, reduced monitoring, and weaker independent learning. The findings suggest that the educational goal should be calibrated trust: learners should benefit from GenAI while critically evaluating its outputs and retaining human judgment. The small, heterogeneous evidence base warrants cautious interpretation.

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

Keywords: *Generative Artificial Intelligence; English Language Learning; Trust; Reliance; Calibrated Trust*

A. Introduction

Generative artificial intelligence (GenAI) has rapidly expanded the range of digital resources available to language learners. In English language learning, GenAI can support writing, vocabulary development, grammar practice, translation, reading, feedback, and self-directed learning. Its conversational and adaptive characteristics can make language practice more immediate and individualized. However, the pedagogical value of these systems depends on more than their technical capability. Learners must decide whether generated explanations, corrections, recommendations, and feedback are sufficiently credible to use, revise, or reject. (Kasneci et al., 2023; Kohnke et al., 2023; Ouyang & Jiao, 2021)

Trust is therefore a central issue in GenAI-supported learning. When learners perceive GenAI as accurate, useful, and dependable, they may be more willing to incorporate it into routine learning. Yet trust can become problematic when it exceeds the actual trustworthiness of the system. Learners who accept generated outputs without verification may become over-reliant, whereas learners who distrust the technology may resist or abandon potentially useful support. The educational problem is thus not simply whether students trust GenAI, but whether that trust is appropriately calibrated to the system's strengths and limitations. (Lee & See, 2004; Hoff & Bashir, 2015; Glikson & Woolley, 2020)

Recent empirical work in language-learning contexts illustrates this tension. A large mixed-methods study of university students found that trust shaped reliance, resistance, behavioral intention, and actual GenAI use, while also identifying over-reliance, appropriate reliance, and under-reliance as distinct patterns (Wang et al., 2025). Other studies have linked trust with continued use through perceived usefulness (Tram et al., 2024), documented learner concerns about over-reliance and lack of trust (Alhamam, 2025), and reported that inconsistent or inaccurate outputs may weaken trust and encourage learners to seek external verification (Abdurasul, 2025). Learner-perception research has likewise shown that perceived trustworthiness can coexist with concerns about accuracy, misinformation, and over-dependence (Esteban et al., 2025). (Jacovi et al., 2021; Schemmer et al., 2023; Shin, 2021)

Despite these emerging findings, the evidence is fragmented across different theoretical perspectives, methods, and language-learning settings. Trust may be treated as a predictor of adoption, a perception of output quality, a dimension of learner experience, or a condition that

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

shapes reliance. Likewise, reliance may be described positively as productive use or negatively as dependence and loss of critical engagement. A focused synthesis is needed to clarify how trust is conceptualized, what factors shape it, and what consequences follow from different levels of trust and reliance. (Yusuf et al., 2024; Zhang et al., 2024; Lo, 2023)

Accordingly, this systematic review synthesizes empirical research on learners' trust and reliance in GenAI-supported English language learning. Given the deliberately narrow scope and the emerging nature of the evidence base, the review emphasizes depth of thematic synthesis rather than breadth. It addresses three research questions: (1) How is learner trust in GenAI conceptualized and manifested in English language learning? (2) What factors shape learners' trust and reliance on GenAI for English language learning? and (3) What learning and behavioral outcomes are associated with trust, distrust, and over-reliance on GenAI? (Page et al., 2021; Yusuf et al., 2024; Zhang et al., 2024)

B. Metodologi

Review Design

The review followed PRISMA 2020 principles for transparent identification, screening, eligibility assessment, and reporting. The scope was intentionally narrow: empirical studies had to examine learner trust, trustworthiness, reliance, credibility, or closely related judgments in GenAI-supported English or language-learning contexts. This narrow eligibility framework was selected to support an in-depth synthesis of a small, emerging literature rather than a broad review of all GenAI use in language education. (Page et al., 2021; Hong et al., 2018; Yusuf et al., 2024)

Information Source and Search Strategy

Scopus was used as the sole bibliographic database because it provides broad interdisciplinary coverage across education, social sciences, language studies, psychology, and technology-related research. The database search was conducted on 6 October 2026. No subject-area restriction was applied because research on GenAI and language learning is interdisciplinary and may be indexed across multiple fields. (Zawacki-Richter et al., 2019; Yusuf et al., 2024; Zhang et al., 2024)

The search covered publications from 2022 through 2026 and used title, abstract, and keyword fields. The search combined three concept blocks: GenAI technologies, English/language learning, and trust-related constructs. The exact Scopus advanced-search string is reported below to support reproducibility.

TITLE-ABS-KEY(("generative artificial intelligence" OR "generative AI" OR GenAI OR ChatGPT OR "large language model" OR "large language models" OR LLM OR LLMs OR Gemini OR Claude

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

OR Copilot) AND ("English language learning" OR "English learning" OR EFL OR ESL OR "foreign language learning" OR "second language learning" OR "language learning" OR "language education") AND (trust OR distrust OR trustworthy OR trustworthiness OR reliability OR credibility OR reliance OR confidence)) AND PUBYEAR > 2021 AND PUBYEAR < 2027

Eligibility Criteria and Study Selection

Studies were included when they were peer-reviewed journal articles published in English, reported empirical data, involved learners or students, examined GenAI or an LLM-based system in an English/foreign/second-language learning context, and contained trust, distrust, trustworthiness, credibility, reliance, resistance, or a closely related trust construct as a substantive variable, theme, or finding. Studies were excluded if they were reviews, conceptual papers, editorials, conference papers, focused exclusively on teachers or technical system performance, addressed contexts unrelated to language learning, did not contain a learner-centered trust-related phenomenon, were published in a language outside the review criterion, or could not be retrieved in full text for eligibility assessment. (Page et al., 2021; Hong et al., 2018; Zawacki-Richter et al., 2019)

The Scopus search yielded 257 records. After applying the intended limits for publication year, English language, and journal-article document type, 190 records remained for title-and-abstract screening. Following screening, 11 reports were sought for full-text retrieval. Four reports could not be retrieved, leaving seven reports for full-text eligibility assessment. One non-English report and one conference paper were excluded at this stage, resulting in five studies in the final synthesis. Screening decisions were based on predefined eligibility criteria, and uncertain records were retained until sufficient information was available for a defensible decision. The selection process is summarized in Figure 1.

Gambar 1. PRISMA 2020 Flow Diagram Of The Study Selection Process.

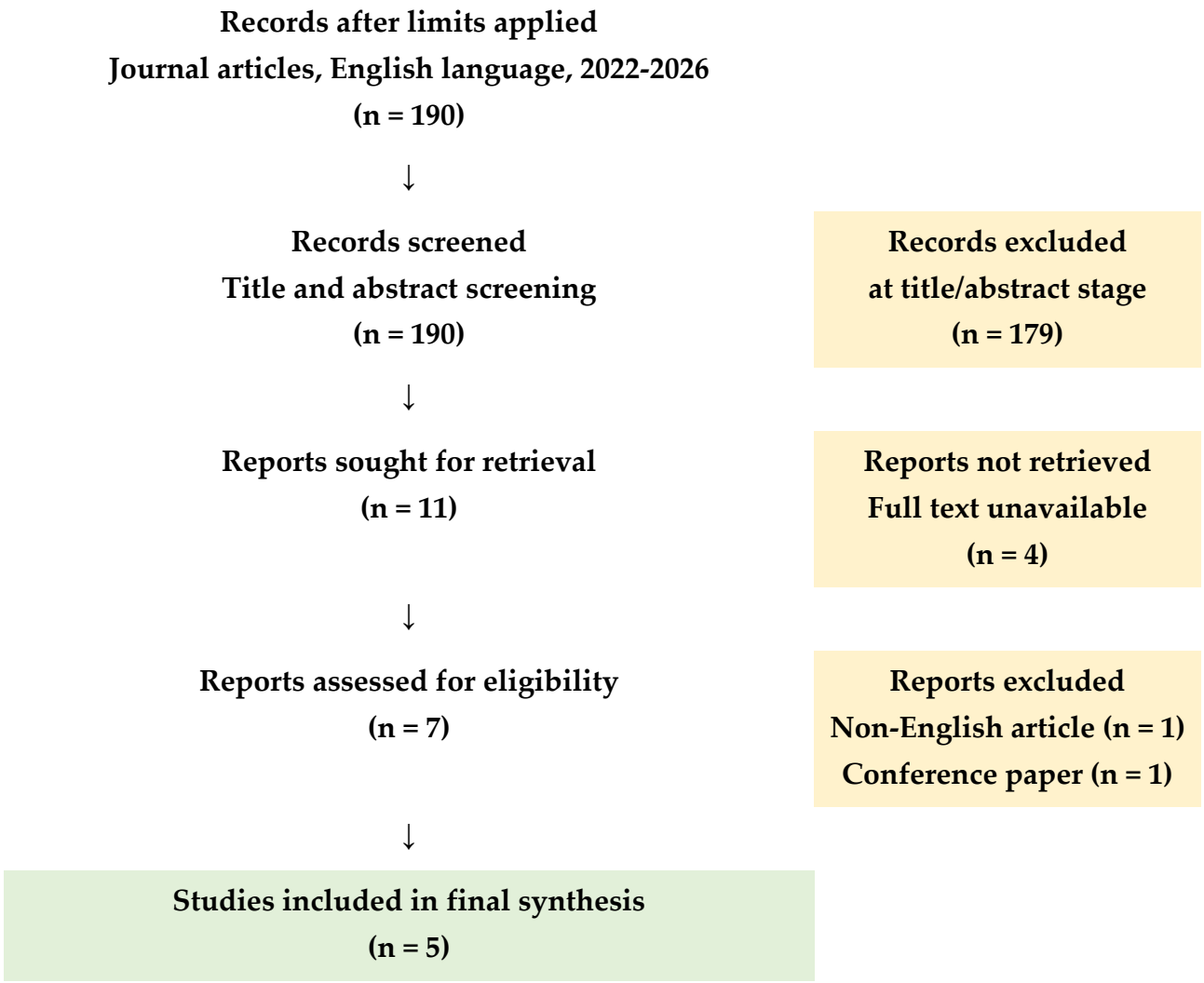
Identification and screening

Records identified from Scopus
(n = 257)

↓

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522



The Scopus search yielded 257 records. After limiting the results to journal articles within the selected publication period and language criteria, 190 records remained for title and abstract screening. Screening identified 11 reports for full-text consideration. Of these, one was excluded because it was not published in English, four were excluded because full text could not be retrieved, and one was excluded because it was a conference paper rather than a journal article. Five studies met the final eligibility criteria and were included in the qualitative synthesis. (Page et al., 2021; Hong et al., 2018; Zhang et al., 2024)

Data Extraction and Synthesis

Data were extracted on publication year, country or setting, participant characteristics, study design, GenAI context, conceptualization of trust or reliance, and findings relevant to the review questions. Descriptive study characteristics were summarized before thematic synthesis. Findings were coded iteratively into recurring patterns concerning antecedents of trust, forms of reliance, behavioral outcomes, and

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

learning consequences. Because the included studies differed substantially in design, samples, constructs, and outcomes, meta-analysis was not appropriate. (Hong et al., 2018; Page et al., 2021; Yusuf et al., 2024)

Methodological Quality Appraisal

Methodological quality was considered using an MMAT-informed, design-sensitive appraisal rather than an aggregate numerical score. Appraisal focused on the appropriateness of the design for the research question, adequacy of sampling and data collection, coherence of analysis, and the main limitations reported or evident from the study design. This approach is consistent with MMAT guidance, which cautions against reducing heterogeneous designs to a single summary score. (Hong et al., 2018; Page et al., 2021; Zawacki-Richter et al., 2019)

C. Result and Discussion

Characteristics of Included Studies

The five included studies were published in 2024 or 2025 and examined higher-education learners in varied contexts, including Saudi Arabia, Vietnam, South Korea, the Philippines, and a multi-university Chinese sample. Designs included qualitative, mixed-methods, and multi-method approaches. Sample sizes ranged from 10 interview participants to 682 survey respondents, reflecting substantial methodological heterogeneity. The evidence base is therefore small but diverse in both design and context.

Tablel 1. Characteristics of included studies.

Study	Context / participants	Design	Trust / reliance focus	Key finding
Alhamam (2025)	Saudi college students learning English (n=10)	Qualitative interviews; thematic analysis	Lack of trust; over-reliance; critical use	Learners valued GenAI support but reported over-reliance and lack of trust, alongside limited AI literacy and institutional support.

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

Tram et al. (2024)	EFL learners; survey n=344 and interviews n=19	Multi-methods	Trust; perceived usefulness; continued use	Trust, interactivity, enjoyment, and subjective norms influenced continued use through perceived usefulness.
Esteban et al. (2025)	Korean and international undergraduates (n=90)	Mixed methods	Trustworthiness; accuracy; over-reliance	Learners perceived GenAI as useful and trustworthy while also reporting accuracy concerns and risks of over-reliance.
Wang et al. (2025)	University students using GenAI for language learning (n=682) plus interviews (n=40)	Explanatory sequential mixed methods; SEM + interviews	Trust; reliance; resistance; perceived risk; use	Trust shaped reliance and resistance and, through them, behavioral intention and actual use; reliance ranged from under- to appropriate to over-reliance.
Abdurasul (2025)	College students using GenAI for English learning (n=15)	Qualitative interviews	Erosion of trust; reliance; overdependence	Inaccurate, vague, or contextually inappropriate outputs

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

			undermined trust; overdependence was linked to passive learning and reduced independent thinking.

Methodological Quality Appraisal

The MMAT-informed appraisal identified generally appropriate alignment between research questions and study designs, while also highlighting recurring constraints related to purposive or non-probability sampling, self-report data, cross-sectional designs, and limited transferability from single settings. Table 2 summarizes the main design strengths and concerns. No aggregate quality score was calculated.

Table 2. MMAT-Informed Design-Sensitive Appraisal Of Included Studies.

Study	Design appraisal	Key strengths	Main methodological concerns
Alhamam (2025)	Qualitative	Semi-structured interviews and thematic analysis matched the exploratory aim.	Small purposive sample (n=10) and single contextual setting limit transferability.
Tram et al. (2024)	Multi-methods	Validated measurement model, large survey sample, follow-up interviews, and triangulation across methods.	Self-selected/social-media recruitment and self-report data may introduce selection and common-source limitations despite reported checks.
Esteban et al. (2025)	Mixed methods	Combines quantitative	Single-university sample (n=90),

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

		perceptions with qualitative learner perspectives in an English-learning context.	uneven subgroup sizes, and perception-based outcomes limit generalizability.
Wang et al. (2025)	Explanatory sequential mixed methods	Large survey (n=682), SEM with reliability/validity testing, 40 follow-up interviews, and explicit integration of phases.	Purposive, cross-sectional, self-report recruitment limits causal inference and population representativeness.
Abdurasul (2025)	Qualitative	Semi-structured interviews provide direct accounts of trust erosion, dependency, and learning behavior.	Small purposive sample (n=15) and single setting constrain transferability and breadth of inference.

Theme 1: Accuracy and reliability as foundations of trust

Across the studies, trust was strongly tied to learners' judgments about the quality of GenAI outputs. Inaccurate, vague, irrelevant, or contextually inappropriate responses weakened confidence and encouraged learners to verify information elsewhere (Abdurasul, 2025). Similarly, participants in the South Korean study simultaneously recognized the usefulness of GenAI and expressed concern about inconsistent accuracy and misinformation (Esteban et al., 2025). These findings indicate that trust in educational GenAI is not merely a general attitude toward technology; it is continually negotiated through learners' experiences of output quality.

Perceived usefulness also appeared to reinforce trust-related adoption. Tram et al. (2024) found that trust contributed to continued use through perceived usefulness, suggesting that learners are more willing to sustain engagement when GenAI is both credible enough to rely on and useful enough to support concrete language-learning goals. The evidence therefore points to an interaction between functional value and epistemic confidence: usefulness can encourage adoption, but usefulness alone may not sustain learning

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

if learners repeatedly encounter outputs they perceive as unreliable.

Theme 2: Trust as a driver of continued use and behavioral intention

Trust was consistently associated with willingness to continue using GenAI. In the multi-method study by Tram et al. (2024), trust was among the significant antecedents of perceived usefulness, which in turn contributed to continued use. Wang et al. (2025) provided the most developed behavioral explanation: trust influenced reliance and resistance, which then shaped behavioral intention, while self-directed learning helped explain whether intention translated into actual use. These findings position trust as a central mechanism within technology adoption rather than a peripheral perception.

However, the relationship was not simply linear. High trust did not automatically indicate productive learning, and low trust did not necessarily indicate poor judgment. Instead, trust influenced how learners positioned GenAI within their decision-making process. Students could accept suggestions uncritically, use them selectively after evaluation, or reject them even when useful. This variation makes reliance a critical bridge between trust and educational outcomes.

Theme 3: From trust to reliance—under-reliance, appropriate reliance, and over-reliance

The most important cross-study pattern concerned the quality of reliance. Wang et al. (2025) distinguished three forms: over-reliance, appropriate reliance, and under-reliance. Over-reliance occurred when learners accepted generated suggestions without sufficient evaluation; appropriate reliance occurred when learners critically examined outputs and retained control over final decisions; under-reliance occurred when learners rejected or ignored useful output despite potential benefit. This typology clarifies why simply increasing trust is not an adequate pedagogical objective.

Other studies echoed the risk of excessive reliance. Alhamam (2025) reported that learners often used GenAI for writing, brainstorming, and translation without consistent critical reflection or revision, and participants expressed concerns about over-reliance and lack of trust. Esteban et al. (2025) also identified over-reliance as a potential drawback associated with reduced critical thinking. Abdurasul (2025) described overdependence on immediate correction and translation support as contributing to passive learning. Across the studies, therefore, productive GenAI use required a balance between openness to assistance and continued learner judgment.

Theme 4: Critical evaluation and learner capability as conditions for calibrated trust

The synthesis suggests that trust

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

becomes educationally productive when learners possess the capability to evaluate GenAI outputs. Critical evaluation appeared repeatedly as the mechanism separating appropriate reliance from over-reliance. Learners who questioned, revised, compared, or verified outputs were better positioned to use GenAI as support rather than as a substitute for reasoning. Conversely, limited AI literacy, limited experience, or strong convenience-oriented use increased the risk that learners would accept output without adequate reflection.

Institutional and pedagogical guidance also mattered. Alhamam (2025) identified limited institutional support and AI literacy as barriers to effective use. Wang et al. (2025) further showed that external cues—including teachers, institutional guidance, evidence, and perceived risk—could influence trust and reliance. These findings suggest that trust calibration is not only an individual cognitive process; it is also shaped by educational environments that communicate when, why, and how GenAI should be used.

Theme 5: Learning consequences of miscalibrated trust

Miscalibrated trust was associated with two opposite risks. Excessive trust could lead to over-reliance, misuse, reduced monitoring, and weakened independent learning. In Abdurasul (2025), reliance on immediate GenAI assistance was linked to passive learning and reduced motivation for deeper engagement. Esteban et al. (2025) reported concern that over-reliance could reduce critical thinking. Wang et al. (2025) similarly identified misuse and abuse when students relied on GenAI without sufficient critical evaluation.

At the other extreme, distrust or strong resistance could result in disuse, preventing learners from benefiting from potentially useful feedback, explanations, or personalized support. This dual risk reinforces the importance of calibrated trust: learners need enough confidence to engage with useful GenAI support, but enough skepticism and capability to detect limitations, verify uncertain outputs, and preserve agency over learning decisions.

Discussion

This systematic review shows that learner trust in GenAI-supported English language learning is best understood as a dynamic judgment rather than a fixed positive attitude. Across the five studies, trust was shaped by perceived accuracy, usefulness, reliability, learner experience, and contextual guidance. Trust then influenced whether learners continued using GenAI, resisted it, relied on it appropriately, or became over-dependent. The evidence therefore suggests that the most useful conceptual shift is from increasing trust to calibrating trust. (Lee & See, 2004; Hoff & Bashir, 2015; Glikson & Woolley, 2020)

A calibrated-trust perspective helps reconcile findings that might otherwise

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

appear contradictory. Learners may report that GenAI is trustworthy while simultaneously expressing concern about inaccurate information. They may value rapid feedback yet worry about dependency. These positions are not inconsistent. They reflect an emerging form of human–GenAI learning in which students must repeatedly decide when to accept, question, revise, or ignore generated content. Trust is therefore relational and task-dependent: it can vary according to the type of language task, the perceived stakes of error, prior experience, and learners' ability to verify output. (Dietvorst et al., 2015; Logg et al., 2019; Schemmer et al., 2023)

The review proposes a concise conceptual pathway. First, technological and informational cues—especially perceived accuracy, reliability, usefulness, and contextual appropriateness—combine with learner and environmental factors such as AI literacy, prior experience, self-directed learning, and institutional guidance. These factors shape learner trust. Second, trust influences reliance and resistance. Reliance may be under-, appropriately, or over-calibrated. Third, these patterns influence behavioral outcomes such as continued use, adoption, and resistance, as well as learning outcomes related to engagement, critical thinking, and learner autonomy. (Jacovi et al., 2021; Shin, 2021; Schemmer et al., 2023)

This framework extends a simple technology-adoption interpretation by emphasizing that adoption is not necessarily equivalent to successful learning. High behavioral intention may coexist with uncritical dependence, while cautious or conditional use may represent more sophisticated judgment. For language education, the quality of the learner–GenAI relationship should therefore be assessed not only by frequency of use, but also by whether learners retain evaluative control over the learning process. (Davis, 1989; Venkatesh et al., 2003; Ouyang & Jiao, 2021)

Implications for English Language Education

The findings have several implications for educators. First, GenAI literacy should include verification practices, not only prompting skills. Learners should be taught to compare generated outputs with credible sources, identify uncertainty, question explanations, recognize potential bias, and distinguish fluent language from evidential accuracy. Such competencies are consistent with contemporary definitions of AI literacy that emphasize understanding, use, critical evaluation, and ethical judgment rather than operational skill alone. (Long & Magerko, 2020; Ng et al., 2021a; Kasneci et al., 2023)

A useful pedagogical objective is therefore appropriate reliance rather than maximal reliance. Students should learn when GenAI is suitable for low-stakes support, when external verification is

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

P-ISSN : 2715-1638
E-ISSN : 2828-5522

necessary, and when human judgment should remain primary. Instructional activities that require comparison, justification, source checking, and revision can make learners active evaluators of generated content instead of passive recipients. This orientation is consistent with human-automation research showing that effective performance depends on aligning trust with actual system capability. (Lee & See, 2004; Schemmer et al., 2023; Farrokhnia et al., 2024)

Implications for Future Research

The evidence base remains small and methodologically heterogeneous. Future research should use longitudinal designs to examine whether trust changes as learners gain experience and encounter both successful and unsuccessful outputs. Intervention studies should test whether AI-literacy instruction, verification routines, transparency cues, or teacher guidance improve calibration between trust and system performance. Cross-cultural and proficiency-level comparisons are also needed because trust may differ according to prior experience, perceived risk, disciplinary expectations, and educational context. (Chan & Zhou, 2023; Chan & Hu, 2023; Glikson & Woolley, 2020)

Limitations

This review has several limitations. First, Scopus was the sole bibliographic database, so relevant studies indexed exclusively elsewhere may have been missed. Second, the final corpus was small, reflecting both the narrow scope and the emerging state of empirical research on learner trust in GenAI-supported English language learning. Third, four of the 11 reports sought for full-text retrieval could not be obtained. Because these reports could not undergo eligibility assessment or contribute data to the synthesis, availability bias cannot be ruled out. Fourth, the included studies were methodologically heterogeneous and largely relied on purposive or non-probability samples and self-reported perceptions or behavior. Finally, most evidence came from higher-education settings, which limits transferability to younger learners, informal learning contexts, and other educational systems. These limitations mean that the review should be interpreted as an emerging synthesis rather than a definitive estimate of effects. (Page et al., 2021; Zawacki-Richter et al., 2019; Yusuf et al., 2024)

D. Conclusion and Recommendations

The emerging evidence indicates that learner trust is a central but double-edged component of GenAI-supported English language learning. Across the five included studies, trust was commonly linked to perceived usefulness, continued use, behavioral intention, and reliance, whereas inaccurate or contextually inappropriate outputs could weaken trust. Excessive trust, however, may encourage

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

over-reliance, reduced monitoring, and weaker learner autonomy. The most defensible educational goal is therefore calibrated trust: learners should be willing to use GenAI when it is useful while retaining the capacity to verify, question, revise, or reject generated outputs. Given the small and heterogeneous evidence base, future longitudinal and intervention research is needed to test how trust develops over time and whether AI-literacy and verification practices improve appropriate reliance. (Lee & See, 2004; Schemmer et al., 2023; Kasneci et al., 2023)

E. Bibliography

Abdurasul, S.-A. M. (2025). The end of English language learning: GenAI as a tool for counter productivity. Environment and Social Psychology, 10(2), 3163. https://doi.org/10.59429/esp.v10i2.3163

Alhamam, A. A. (2025). Investigating the role of AI tool interactions in enhancing English language acquisition among Saudi college students. Arab World English Journal, 16(3), 426–440. https://doi.org/10.24093/awej/vol16no3.26

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8

Chan, C. K. Y., & Zhou, W. (2023). An expectancy value theory (EVT) based instrument for measuring student perceptions of generative AI. Smart Learning Environments, 10, 64. https://doi.org/10.1186/s40561-023-00284-4

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Dwivedi, Y. K., Kshetri, N., Hughes, L., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Esteban, A. J., Park, I., Nguyen, T. N., Perunovic, S., Park, S. E., Shehzadi, & Yi, J.-I. (2025). Undergraduate students' perceptions on the use of ChatGPT for English learning at a Korean university. rEFLections, 32(2), 994–1012.

Farrokhnia, M., Banihashemi, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570

Hong, Q. N., Fàbregues, S., Bartlett, G., et al. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285–291. https://doi.org/10.3233/EFI-180221

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 624–635. https://doi.org/10.1145/3442188.3445923

Kasneci, E., Sessler, K., Küchemann, S., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. The

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

International Journal of Management Education, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3313831.3376727

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021a). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Schemmer, M., Kühl, N., Benz, C., Bartos, A., & Satzger, G. (2023). Appropriate reliance on AI advice: Conceptualization and the effect of explanations. Proceedings of the 28th International Conference on Intelligent User Interfaces, 410–422. https://doi.org/10.1145/3581641.3584066

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human–Computer Studies, 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Tlili, A., Shehata, B., Adarkwah, M. A., et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. System, 127,

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

103528. https://doi.org/10.1016/j.system.2024.103528

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478. https://doi.org/10.2307/30036540

Wang, X., Chin, H. L., & Zainuddin, Z. (2024). The effectiveness of generative AI in education: A systematic review of empirical study. JuKu: Jurnal Kurikulum & Pengajaran Asia Pasifik, 12(4), 48–66. https://doi.org/10.22452/juku.vol12no4.5

Wang, F., Li, N., Cheung, A. C. K., & Wong, G. K. W. (2025). In GenAI we trust: An investigation of university students' reliance on and resistance to generative AI in language learning. International Journal of Educational Technology in Higher Education, 22, 59. https://doi.org/10.1186/s41239-025-00547-9

Yusuf, A., Pervin, N., Román-González, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. Review of Education, 12(2), e3489. https://doi.org/10.1002/rev3.3489

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A systematic literature review of empirical research on applying generative artificial intelligence in education. Frontiers of Digital Education, 1(3), 223–245. https://doi.org/10.1007/s44366-024-0028-5

Copyright (c) 2026. Elizabeth Meiske Maythy Lasut. This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.