

Empowerment or Illusion? Generative AI and the Promise of Inclusion for EFL Learners with Learning Difficulties: Teachers' Perspectives

Kashif Ali Sabiri^{1*}, Yehia Ahmed Al-Sohbai², Muhammad Shaharyar Sabiri³,
Adnan Mohammed Bataineh⁴, Mamoonah Kanwal⁵

¹Ph. D. candidate @ Tampere University, Tampere, Finland, <https://orcid.org/0000-0002-8663-579X>, kashif.sabiri@tuni.fi; English Language Instructor, University of Buraimi, Buraimi, Oman, kashif.s@uob.edu.om

²Arab Open University: Riyadh, Riyadh Region, Saudi Arabia

³Lecturer in English, The University of Faisalabad, Faisalabad, Punjab, Pakistan, <https://orcid.org/0000-0001-6694-562X>, shaharyarsabiri6@gmail.com

⁴University of Buraimi, Buraimi, Oman, <https://orcid.org/0000-0002-3888-4759>, adnan.m@uob.edu.om

⁵University of Balearic Island, Spain; mamoonah.kanwal0@gmail.com; <https://orcid.org/my-orcid?orcid=0000-0002-0263-256844>

Corresponding Author: kashif.sabiri@tuni.fi

HOW TO CITE:

Kashif Ali Sabiri, Yehia Ahmed Al-Sohbai, Muhammad Shaharyar Sabiri, Adnan Mohammed Bataineh, Mamoonah Kanwal (2026). Empowerment or Illusion? Generative AI and the Promise of Inclusion for EFL Learners with Learning Difficulties: Teachers' Perspectives International Journal of Special Education, 41(22s),819-831

ABSTRACT:

English as a Foreign Language (EFL) instruction is making use of generative artificial intelligence (AI) tools such as students with learning difficulties including dyslexia, attention-deficit/hyperactivity disorder (ADHD), or processing difficulties. The effectiveness of these tools as true empowerment as opposed to a superficial appearance of support has not yet been tested empirically from the teachers' point of view. The present research was conducted with 91 EFL/ESL teachers with varying levels of education and from diverse countries who taught at least one student with a learning difficulty in the last two years. A convergent parallel mixed-methods approach was used, where teachers completed a validated Empowerment Scale (with excellent internal consistency, $\alpha = .899$) and Skepticism/Illusion Scale (with excellent internal consistency, $\alpha = .955$), and open-ended questions were analyzed using thematic analysis. Empowerment and Skepticism scores were modestly, negatively correlated, $r(89) = -.258$, $p = .014$; however, there was no correlation between the training and the reduction of Skepticism. Teachers who participated in prior training reported significantly higher levels of Empowerment than those who did not, $t(89) = 2.48$, $p = .015$, $d = 0.65$. Thematic analysis showed that nearly all teachers did not label Generative AI as simply enabling or posturing, but rather as either a "real" AI support when AI output is teacher verified and used to assist students to overcome barriers to tasks, or as a "pants on fire" when students are using

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication under the
terms and conditions
of the Creative Commons Attribution (CC BY)
license (<http://creativecommons.org/licenses/by/4.0/>).

unverified AI output. Based on these findings, Critical AI Literacy in inclusive EFL contexts might be better interpreted not as a general attitude towards AI, but as a verifiable practice to teach.

Keywords: Generative AI, learning difficulties, inclusive education, EFL, Critical AI Literacy, teacher perceptions

1. Introduction

The use of generative AI tools in EFL classrooms, such as the large language model-based Chatbots and AI-powered writing assistants, has become increasingly commonplace, progressing from novelty to a staple in classrooms. A global survey recently conducted revealed that the majority of higher education students currently employ tools like ChatGPT in their studies, and that this is happening much faster than the institutions are able to keep up with (Ravselj et al., 2025). These tools can simplify text, create practice materials and give instant feedback on writing and are being used more and more with students who have diagnosed or suspected learning problems, including processing difficulties, attention-deficit/hyperactivity disorder (ADHD) and dyslexia. It was argued that Generative AI will be a natural evolution of assistive technology, and can make a difference in removing barriers that have prevented these learners from participating as active learners in mainstream language instruction (Almgren Bäck et al., 2023). The same attributes that make them so flexible and useful, the ability to simplify, summarize and create content on the spot, create a secondary issue: that they may appear to be helping, but aren't helping, and they may give the impression that the student is improving without actually improving.

This is not an abstract worry. While large language models lack the capacity to accurately depict complex phenomena like the human body, they are known to hallucinate and generate inaccurate or fabricated content even when asked to perform tightly controlled tasks like generating references for systematic reviews, their hallucination or inaccuracy rates are still sufficiently high for a critical human review to be

regarded as essential and not optional (Chelli et al., 2024). If teacher generated feedback and/or simplified content is similarly unreliable when used in a classroom, then a tool designed to help a student who has a learning difficulty may reinforce a misunderstanding, or hide or mask, without the teacher and the student realizing that the underlying learning difficulty is not being addressed.

This tension is not just a theoretical one. It is at the core of recent calls for Critical AI Literacy in education, where both teachers and learners need to be able to question AI output instead of taking it as gospel (Long & Magerko, 2020; Tour et al., 2025). In the context of EFL teaching, in which language is being taught and tested, the importance of this dichotomy is further amplified: a device that makes a passage easier to read can help students understand the passage, or it can subtly allow students to bypass the language-processing requirement that the lesson is meant to develop. For students with learning difficulties, who may already face other obstacles to full participation, this is not just a feature of generative AI systems but a crucial aspect of the impact of generative AI on inclusive practice.

Although these tools have been quickly adopted with students in the classroom, little research has been conducted directly with the teacher who acts as the mediator between the tool and the student with learning difficulties. The researchers found that previous studies examining Generative AI in language education have primarily focused on general language learners and general outcome measures (Hwang et al., 2024; Wang & Xue, 2024), while previous studies that investigated assistive technology for learning disabilities were largely conducted prior to the advent of the current generation of Generative

AI tools (Almgren Bäck et al., 2023). Research on teachers' general acceptance of AI tools have been conducted via technology acceptance frameworks (Zhang et al., 2023, Zou & Huang, 2023), but such studies have not often been linked to the specific context of teaching students with learning difficulties, which is also of high educational significance. The effects of Generative AI on teachers' perception of empowerment versus illusion in this population has not been directly explored.

This study attempts to fill in that gap by adopting a mixed methods survey method to investigate the experiences of EFL/ESL teachers who have taught students with learning difficulties. It explores not only teachers' general attitudes towards GenAI but the specific concrete situations in which they view GenAI as an actual helper and not as a quick-and-easy trick, and what this means for conceptualizing and teaching Critical AI Literacy in inclusive EFL settings.

1.1 Research Questions

How can Generative AI tools empower or mislead EFL teachers in helping students who have learning challenges?

Sub-Question 1: In what ways do EFL teachers perceive that the use of Generative AI tools with LLLD students has pedagogical advantages and disadvantages?

Sub-Question 2: What are the differences in EFL teachers' perceptions of empowerment and skepticism with regard to years of teaching and prior training with AI?

2. Literature Review

2.1 Generative AI and Automated Feedback in Language Education

When used regularly over time, mobile- and AI-supported language learning tools are found to support sustained learner engagement, with mixed results regarding their effects on deeper learning outcomes rather than surface engagement, depending on the measure used for the behavioral outcomes (Hwang et al., 2024). Automated writing evaluation (AWE) systems have been the subject of study in a more relevant

and directly related body of literature. The results from multiple meta-analyses show that AWE feedback has a medium-to-large positive impact on writing quality, with an average *g* value from 0.55 to 0.86 in different analyses (Fleckenstein et al., 2023; Zhai & Ma, 2023), especially for EFL and ESL students rather than native English speakers (Zhai & Ma, 2023).

Experimental work with specific AWE tools has similarly found improvements in both writing performance and writing-related anxiety among EFL university students (Du & Mohd Nordin, 2026), and AI-driven chatbot interventions have been linked to improved academic engagement among EFL learners more broadly (Wang & Xue, 2024). Much of this literature, however, has examined general EFL learner populations, evaluating engagement, motivation, or language gains, rather than the specific population of learners with diagnosed or suspected learning difficulties, for whom the underlying cognitive and processing demands of a task, and therefore the potential benefit or risk of AI-mediated simplification, may differ substantially from the general learner population.

2.2 Assistive Technology and Learning Difficulties

Prior to the emergence of generative AI, assistive technologies such as text-to-speech and speech-to-text tools were already established as valuable supports for students with dyslexia and related learning difficulties. A five-year longitudinal study of dyslexic students' experiences with assistive technology found that benefits were highly dependent on contextual factors, including the degree of teacher-supported, systematic training in how to use the technology, and on students' emotional relationship to the tool itself (Almgren Bäck et al., 2023). Specifically, this literature identified that assistive technology benefits were not a given or universal, but were dependent on the conditions of its use; a clear application of this research is related to the central question of this study: Does using Generative AI constitute empowerment or illusion based on the context in which it is used?

This literature also provides a larger theoretical context that aligns with the concepts of Universal Design for Learning (UDL) that states that universal design for learning is the design of flexible instruction environments and materials that support learners with differences from the beginning, instead of modifying it for learners with differences after the fact (CAST 2018). While conceptually the ability of generative AI to create multiple representations and simplify material to the reader's level of ability aligns with UDL's focus on multiple means of representation, the question of whether this ability will truly benefit the user pedagogically, or simply provide a superficial approximation of UDL principles, is the empirical question that this study addresses.

2.3 Reliability, Hallucination, and the Verification Problem

Some of the evidence is growing that generative AI systems generate fluent, plausible, but often factually incorrect or even fabricated text, which is known as hallucination. Hallucination and citation-inaccuracy rates have been found to be significant even when used in highly controlled settings like generating literature references for systematic reviews (Chelli et al., 2024), which is why researchers recommend that these tools should not be used without being checked manually. This is found to be the case by applying survey research among students in higher education institutions, where awareness of the problem of reliability is found alongside continued use of these tools, and sometimes in contrast, indicating that awareness of a reliability problem does not necessarily lead to regular verification behavior (Ravselj et al., 2025). This literature has rarely been connected to the specific context of students with learning difficulties, for whom an unverified simplification or an inaccurate piece of AI-generated feedback may be considerably harder to detect and correct independently than for a general learner population, raising the stakes of the verification problem in exactly the population this study examines.

2.4 Critical AI Literacy

AI literacy has been defined as the set of competencies that enable individuals to understand, effectively use, and critically evaluate AI systems (Long & Magerko, 2020). Building on this foundation, recent scholarship has argued for a specifically critical orientation to AI literacy in education, one that goes beyond functional competence to include the capacity to interrogate how AI systems produce their output and whose interests that output serves (UNESCO, 2024). Within language education specifically, this critical orientation has been proposed as essential for both teachers and learners, given that large language models are trained predominantly on data reflecting dominant linguistic and cultural norms, raising equity concerns that are amplified, rather than resolved, when these tools are deployed with already-marginalized learner populations such as students with learning difficulties (Tour et al., 2025).

This body of literature, however, focuses on Critical AI Literacy more as a general disposition or attitude to developing the right kind of skepticism rather than as a specific, observable teaching practice. The present study adds to this literature by analyzing the empirical evidence for whether teachers themselves make a distinction between a concrete behavioral condition and what is often termed 'verification of AI output before reaching the student' as the factor that differentiates genuine empowerment from 'surface level illusion' in EFL classrooms, providing an operational account of what Critical AI Literacy looks like in inclusive EFL practice.

2.5 Teacher Acceptance and Training for AI Integration

A large and growing body of research investigates teachers and pre-service teachers' receptivity of AI tools, which are generally based on technology acceptance frameworks such as the one that assumes that perceived usefulness and perceived ease of use predict technology adoption intention (Zou & Huang, 2023). The desire to use AI has been found to be consistently related to perceived usefulness, ease of use, and trust in AI by large-scale surveys of pre-service teachers (Zhang et al., 2023), and to be related to

perceived risk and AI-related trust in more complex structural pathways (Zhang et al., 2024). Beyond general acceptance, dedicated AI literacy training programs have been shown to meaningfully improve pre-service teachers' AI literacy skills (Younis, 2024), and pre-service teachers themselves report that structured AI integration training shapes how they conceptualize AI's role in their future teaching (Zakaria & Hashim, 2024). Foundational work on AI literacy among pre-service teachers further suggests that literacy in this domain is multidimensional, encompassing not only technical familiarity but also ethical and critical evaluation competencies (Ayanwale et al., 2024).

This literature consistently treats training and acceptance as related but distinct constructs, a distinction with direct relevance to the present study's finding that AI training was associated with greater perceived empowerment but not with reduced skepticism (Section 4.5): training may plausibly build the confidence and perceived usefulness that this literature associates with acceptance, without necessarily building the specific critical evaluation competency that Critical AI Literacy scholarship treats as a separate skill.

2.6 The Present Study

Taken together, this literature suggests four things relevant to the present study. First, Generative AI's impact on language learners is not uniform and depends substantially on how it is used, a pattern consistent with earlier assistive technology research on students with learning difficulties specifically, and with meta-analytic evidence that automated feedback effects vary considerably by tool and context. Second, the well-documented tendency of generative AI systems to produce unreliable or fabricated output means that any benefit these tools offer is conditional on some form of verification, a concern that has not been directly examined for students with learning difficulties. Third, Critical AI Literacy has been conceptualized mostly as a disposition rather than based on empirical research of teachers' own experiences of what distinguishes effective from AI-infused

superficial uses. Fourth, though a large body of texts has studied teachers' general adaption of AI tools, these studies have not been linked to the specific and pivotal setting of students with learning challenges. The gaps are addressed in this study by employing a mixed-methods survey to empirically test and verify three distinct teachers' perceptions: Generative AI is empowering, Generative AI is illusory, or Generative AI is empowering conditionally.

3. Method

3.1 Research Design

The type of design used in this study was convergent parallel mixed methods design in which quantitative and qualitative data were collected at the same time, using one instrument (the online questionnaire), analyzed separately using methods that are suitable for each type of data, and then integrated at the time of interpretation (Creswell & Plano Clark 2018). This design was chosen because the two conceptually different constructs perceived empowerment and perceived skepticism needed to be measured as separate quantitative scales, as well as the qualitative reasoning participants gave for their responses on the open-ended items within the same instrument.

3.2 Participants

The target population included English as foreign language/English as second language (EFL/ESL) teachers who had at least one student in their classrooms who had a diagnosed or suspected learning difficulty during the previous two years. The recruitment method was a convenience sampling with the assistance of the researcher's professional networks, EFL teacher associations and social media networks. A total of 98 teachers completed the survey, seven of whom were excluded based on the eligibility criterion and were not included in the analytical sample of 91 teachers across various countries and education levels (primary, secondary, university and adult education).

3.3 Instrumentation

The data were gathered using a single anonymous self-administered online questionnaire that

consisted of four parts: (a) demographic and background information, such as years of teaching experience, prior training on Generative AI, and the type(s) of learning difficulty that were most common; (b) a six-item Empowerment Scale to assess teachers' perceived value when Generative AI truly helps students with learning difficulties; (c) a six-item Skepticism/Illusion Scale to determine the teachers' doubts that the technology that appeared helpful actually was; and (d) three open-ended questions asking the teachers to describe specific examples in which they observed Generative AI helped, or hindered, a student with a learning difficulty, and their general perception of Generative AI as a support or surface-level fix for students with learning difficulties. Both scales used a five-point Likert format (1 = Strongly Disagree to 5 = Strongly Agree).

3.4 Procedure and Ethical Considerations

Following the presentation of an informed consent statement, prospective participants completed a screening question confirming their eligibility before proceeding to the substantive questionnaire. Participation was voluntary and anonymous; no personally identifying information was collected, and participants could withdraw at any point prior to submission. No data were collected directly from students.

3.5 Data Analysis

Prior to substantive analysis, the dataset was screened for missing data, straight-lining, and duplicate response patterns. Internal consistency of both scales was assessed using Cronbach's alpha. Descriptive statistics summarized sample characteristics and scale scores. A Pearson correlation examined the relationship between Empowerment and Skepticism scores (Main Research Question). Independent-samples t-tests and one-way ANOVA examined differences in scale scores by AI training and teaching experience (Sub-Question 2). Open-ended responses were analyzed using Braun and Clarke's (2006) six-phase thematic analysis framework, coded inductively at the semantic

level (Sub-Question 1). Quantitative and qualitative findings were integrated through a joint display comparing statistical patterns against corresponding qualitative themes (Creswell & Plano Clark, 2018).

4. Results

4.1 Sample Characteristics

A total of 98 EFL/ESL teachers accessed the survey. Seven respondents (7.1%) were screened out based on the eligibility criterion (having taught at least one student with a diagnosed or suspected learning difficulty within the past two years), yielding a final analytic sample of 91 teachers. Respondents represented a range of countries, most frequently Pakistan, Oman, the United Arab Emirates, India, Saudi Arabia, and the United Kingdom, and taught across primary, secondary, university, and adult education contexts. The sample was distributed across four experience bands: 0–3 years ($n = 23$), 4–7 years ($n = 20$), 8–15 years ($n = 23$), and 16+ years ($n = 25$). The majority of respondents ($n = 73$, 80.2%) reported having received some formal training in using Generative AI tools for teaching, while 18 (19.8%) reported no such training.

4.2 Data Screening and Reliability of Measures

There were no missing values in any required items, and no identical response patterns for the entire set of questions were found. 12.1% of the respondents ($n = 11$) gave responses that were limited to two adjacent scale points on the 12 Likert scale items; examination of those cases revealed consistent and non-random patterns, with some respondents having Empowerment and Skepticism scale means that were very close together, suggesting that they did not fail to include the items but rather were in fact responding to both. No cases were excluded for this reason.

The internal consistency was high for both scales, with Cronbach's α for the Empowerment Scale being .899 and for the Skepticism/Illusion Scale being .955 (Table 1).

Table 1 Descriptive Statistics and Reliability of the Empowerment and Skepticism Scales

Scale	M	SD	Cronbach's α	n items
Empowerment Scale	4.36	0.70	.899	6
Skepticism/Illusion Scale	1.90	1.02	.955	6

4.3 Descriptive Statistics

The teachers in the full sample ($n = 91$) agreed relatively well with the items related to Empowerment ($M = 4.36$, $SD = 0.70$), and relatively less with the items related to Skepticism ($M = 1.90$, $SD = 1.02$), indicating that this sample of teachers was more likely to consider Generative AI tools empowering than skeptical of them for students with learning disabilities.

4.4 Relationship Between Empowerment and Skepticism (Main Research Question)

To assess the relationship between the scores for Empowerment and Skepticism, a Pearson

correlation was calculated, which directly answered the key research question of the study. A statistically significant, negative correlation was found, $r(89) = -.258$, $p = .014$. While statistically significant, the size of this correlation is small, suggesting that Empowerment and Skepticism are two separate and distinct constructs that are related, not two extremes of a single underlying attitude. Many teachers, however, endorsed both scores at a similar level, indicating that a large proportion of teachers endorse both views of empowerment and skepticism, rather than choosing one view over another.

Figure 2. Relationship Between Empowerment and Skepticism Scores ($n = 91$)

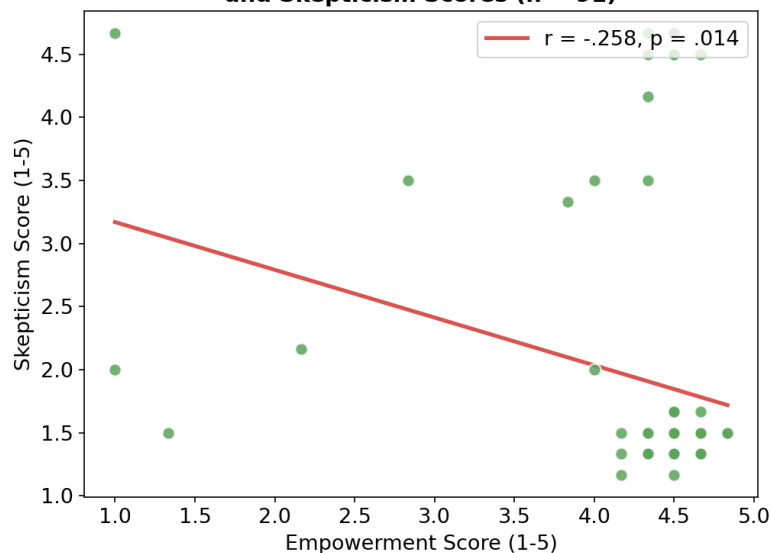


Figure 2. Relationship between Empowerment and Skepticism scores ($n = 91$).

4.5 Differences by Teaching Experience and AI Training (Sub-Question 2)

The difference in Empowerment and Skepticism scores were evaluated via independent-samples t -tests based on the participants' prior training in Generative AI. AI trained teachers also scored

significantly higher on the Empowerment scale than AI non-trained teachers ($M = 4.45$, $SD = 0.52$; $M = 4.01$, $SD = 1.13$, respectively; $t(89) = 2.48$, $p = .015$, $d = 0.65$, a medium-sized effect). Skepticism scores did not differ significantly by training status, $t(89) = -1.25$, $p = .214$, $d = -0.33$.

One-way ANOVAs were used to compare the four teaching experience bands (0-3 years, 4-7 years, 8-15 years and 16+ years). Neither

Empowerment, $F(3, 87) = 1.60, p = .194$, nor Skepticism, $F(3, 87) = 1.87, p = .141$, differed significantly across experience groups (Table 2).

Table 2 Empowerment and Skepticism Scores by AI Training and Teaching Experience

Comparison	Group	n	M (SD)	t / F	p	d
Empowerment by AI training	Trained	73	4.45 (0.52)	$t(89)=2.48$	.015	0.65
	Untrained	18	4.01 (1.13)			
Skepticism by AI training	Trained	73	1.83 (0.99)	$t(89)=-1.25$	.214	-0.33
	Untrained	18	2.17 (1.13)			
Empowerment by experience	0–3 yrs	23	4.09 (1.19)	$F(3,87)=1.60$	.194	—
	4–7 yrs	20	4.49 (0.15)			
	8–15 yrs	23	4.43 (0.53)			
	16+ yrs	25	4.45 (0.40)			
Skepticism by experience	0–3 yrs	23	2.29 (1.35)	$F(3,87)=1.87$	.141	—
	4–7 yrs	20	1.73 (0.95)			
	8–15 yrs	23	1.64 (0.66)			
	16+ yrs	25	1.91 (0.94)			

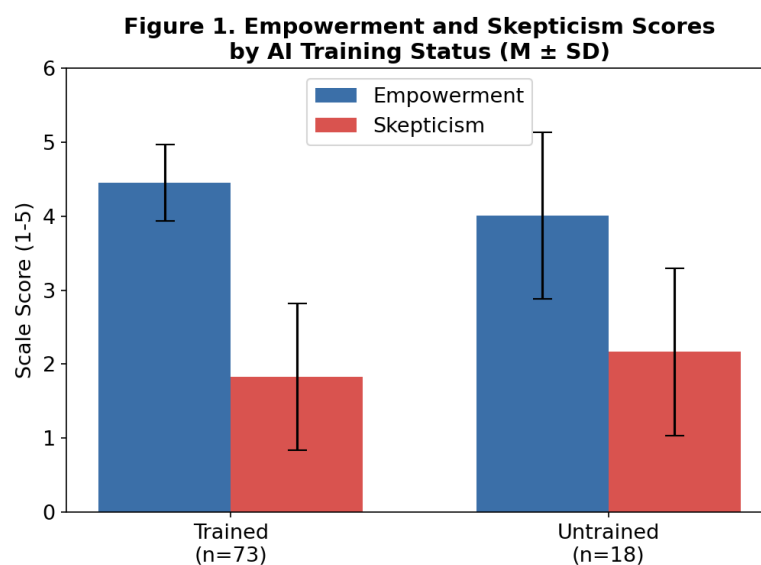


Figure 1. Empowerment and Skepticism scores by AI training status (M ± SD).

4.6 Qualitative Findings (Sub-Question 1)

The open-ended responses ($n = 91$ in Questions 1 and 3; $n = 90$ in Question 2) were thematically analyzed using the six phases of Braun and Clarke (2006). The coding was done inductively at semantic level. Six themes were identified that

ranged from reported benefits, reported limitations, and overall characterization by teachers of Generative AI as support or “surface-level fix” (Table 3).

Reported Benefits

Task decomposition was the most common reported benefit and referred to how Generative AI tools help teachers divide lengthy or complex tasks into smaller, sequential steps that mitigated the feeling of being overwhelmed for students for whom attention or processing is a challenge. Another significant trend was the use of AI to simplify language, specifically for students with dyslexia or processing challenges, enabling them to comprehend complex sentences or restate challenging passages in simpler terms. Another theme was the private, immediate nature of AI feedback, especially in the context of spelling and grammar corrections, which several teachers explained could lessen the embarrassment of having to go through the process of being corrected repeatedly by others.

Reported Limitations

The most frequent problem reported was a mismatch between the complexity of the created content and the student's ability to comprehend it: on one hand, the AI generated content was too advanced for the student, and on the other hand, the AI content was too simple to include what was needed in the lesson. The second theme was about over-reliance, where students were said to plagiarize AI-generated texts without comprehending them, which was counterproductive to the process of learning. A

third theme was accuracy and reliability issues, with some teachers noting that AI generated content needed to be verified or edited before being used in the classroom.

Overall Characterization: Conditional Support

When asked directly about the role of Generative AI in the classroom, the vast majority of teachers (around 90%) did not choose either extreme: "is it really useful as an actual support" or "is it just a quick hack". Rather, we saw a form of conditional support logic emerge: Generative AI was seen as a true form of support when it lowered the student's barriers to a task without taking away the student's thinking and effort, and as a "surface" kind of support when the student was able to copy or use AI-generated material without thinking or even teacher verification. The logic of the vast majority of responses was not a general optimism or skepticism about the technology per se, but rather the distinction.

"I think it can be a true help in that sense, encouraging low level pupils to be involved in the class. I think it can be a real help and not just a patch on the surface" (Respondent 8)

"It turns into a shallow tickle due to the difficulty not being solved." (Respondent 9)

Table 3 Summary of Qualitative Themes with Illustrative Responses

Theme	Description	Illustrative Response
Task decomposition	Breaking material into smaller, sequential steps or sections	"Turned a complicated set of instructions into a numbered checklist."
Language simplification	Simplifying vocabulary or sentence complexity	"Explained unfamiliar terminology...using simpler examples."
Private, immediate feedback	Reducing embarrassment associated with repeated correction	"Could ask questions repeatedly without feeling embarrassed."

Complexity mismatch	AI output too advanced or, conversely, oversimplified	"The shortened explanation removed too much of the original meaning."
Over-reliance risk	Students copying output without understanding it	"The student copied...without fully understanding their meaning."
Conditional support logic	Genuine support when teacher-verified; surface-level fix when unsupervised	"It becomes a surface-level fix because the underlying difficulty has not been addressed."

4.7 Integration of Quantitative and Qualitative Findings

The qualitative and quantitative strands are integrated on a uniform narrative. The moderate (but not strong) negative correlation between Empowerment and Skepticism (Section 4.4) can be directly explained by the qualitative finding that most teachers' responses to either construct are conditioned when using the technology, whether they endorse it or not is dependent on a condition rather than a disposition toward the technology. Likewise, the strong advantage for Empowerment that was found with prior AI training (Section 4.5) aligns with the qualitative focus of accuracy and reliability concerns, as teachers trained with AI are likely to have the ability to check and adjust AI results as needed, a quality their qualitative responses recognized as what makes a truly helpful intervention.

5. Discussion

5.1 Interpretation of the Main Finding

This study explored the views of EFL/ESL teachers on the use of Generative AI tools as tools that enhance or mislead the learning of learners with learning difficulties. The findings do not provide an easy answer to this question and the qualitative data indicate this is not just a limitation of the study but a substantive finding. Contrary to the impression of a technology optimist versus a technology skeptic division within the sample of teachers, the overwhelming majority expressed a conditional attitude: that the same technology can be empowering or an illusion, depending on the use made of it. The central question of the study is rephrased as follows: The question is not whether Generative

AI empowers or deceives, but when does which occur.

5.2 The Role of Training

The strong correlation between previous AI training and increases in Empowerment scores, but not in Skepticism scores, suggests that AI training can increase teacher confidence in effectively using these tools, potentially without addressing concerns about reliability. This aligns with the qualitative theme of accuracy and reliability concerns, which remained consistent across all teachers regardless of their overall perspective, indicating that ensuring AI-generated content is accurate and reliable might require a specific skill set that is different from simply knowing how to use an AI tool, and that it needs to be explicitly included in training programs rather than assumed through exposure. This is consistent with the general literature on teacher acceptance, which reveals that teacher acceptance training is related to perceived usefulness and ease of use (e.g., Zhang et al., 2023; Zou & Huang, 2023) but does not overlap with the critical evaluation competency that is evidenced in the qualitative data of this study and suggests that genuine support is lacking in the classroom. The above calls for a concrete direction in the development of AI literacy training; that is, dedicated training programs rather than general programs of AI-exposure have proved more directly to build these competencies (Younis, 2024).

5.3 Theoretical Implications

These findings expand on the idea of Critical AI Literacy, a general disposition towards questioning AI-produced content, to one specific

and identifiable teacher practice: verifying and adapting AI-generated content before presenting it to a student with a learning difficulty. The conditional support logic identified in this sample suggests that Critical AI Literacy, in this context, is not so much about having a negative stance toward AI in general; it's more about applying a verification step when using AI in practice. The operationalization of Critical AI Literacy for inclusive education around this concrete behavior could benefit future conceptualizations of the construct as an attitude or awareness. This is consistent with documented reliability limitations of generative AI systems more broadly: even in tightly controlled academic tasks, hallucination and inaccuracy rates remain high enough to necessitate thorough human verification (Chelli et al., 2024), and awareness of this limitation does not appear to automatically translate into consistent verification behavior among users (Ravselj et al., 2025). The present findings suggest that EFL teachers working with students with learning difficulties have, largely independently, arrived at the same conclusion through their own classroom practice.

5.4 Limitations

Several limitations qualify these findings. The sample was recruited through convenience and snowball sampling, predominantly through the researcher's professional networks, and while it spans multiple countries, it is not representative of EFL/ESL teachers globally. Data on learning difficulties and AI tool use were self-reported by teachers rather than independently verified or reported by students themselves. The qualitative data were collected as open-ended survey responses rather than through dedicated interviews, which likely constrained the depth and elaboration possible compared to a spoken interview format; open-ended responses in this study averaged between 16 and 29 words, considerably shorter than transcript-based qualitative data typically permits. Finally, the cross-sectional design captures teachers' perceptions at a single point in time and cannot establish whether the conditional support logic identified here reflects a stable belief or one that shifts with further experience or training.

5.5 Implications for Practice

For teacher training programs, these findings suggest that building Empowerment alone, through general exposure to AI tools, is insufficient without also addressing the verification practices that teachers themselves identify as the condition separating genuine support from a surface-level fix. Courses specifically designed to help students assess the complexity and accuracy of AI-generated content could better address teachers' concerns than courses focused on learning how to use the tools. The opacity-to-opacity theme is one that seems to repeat in the literature for instructional designers developing AI-based education instruments, and a concrete goal seems to be to have their tools produce more appropriate output for a student from their documented difficulty profile, rather than having to be tweaked by the teacher.

5.6 Directions for Future Research

Future studies should examine the types of support logic found in this study to determine if they are predictive of classroom results, not just teacher-reported perception, by measuring results of students in classrooms with conditional support, as opposed to those lacking conditional support. Longitudinal designs might also be used to assess if a teacher's scores on the Empowerment and Skepticism scale change over time as the teacher becomes more familiar with the tool. Finally, as the qualitative information of this study was limited by the nature of the open-ended responses in a survey format, future studies on the topic would gain by conducting in-depth semi-structured interviews with a purposefully selected subsample to enable the conditional support logic identified in this study to be explored further.

6. Conclusion

The present study focused on how EFL/ESL teachers view the use of Generative AI tools in assisting learners with learning difficulties as empowering or illusory. The findings do not support either extreme but instead point to a more specific understanding: The vast majority of the teachers in this sample expressed a

conditional support logic—that is, a sense of generative AI as a true source of empowerment when used by students with teacher verification, but a hallucinatory one when used without such verification. The quantitative finding of this conditional logic was that there were only weak relationships between Empowerment and Skepticism scores, whereas the qualitative finding was that teachers' descriptions of when and why AI use was successful or unsuccessful was explained by this conditional logic. Perceived empowerment was tied to prior training with

Generative AI tools, whereas a decrease in skepticism was not. This indicates that confidence in using the Generative AI tools and confidence in verifying their output may be somewhat independent. Such results indicate that Critical AI Literacy within the context of inclusive EFL teaching cannot be viewed merely as an attitude toward AI but rather should be seen as a concretized and teachable verification practice, which should be developed and evaluated within EFL in the state of AI tool use rather

than with the expectation that it will occur naturally when students are exposed to AI tools.

References

- Almgren Bäck, G., Lindeblad, E., Elmqvist, C., & Svensson, I. (2023). Dyslexic students' experiences in using assistive technology to support written language skills: A five-year follow-up. *Disability and Rehabilitation: Assistive Technology*, 19(4), 1217–1227. <https://doi.org/10.1080/17483107.2022.2161647>
- Ayanwale, M. A., Adelana, O. P., Molefi, R. R., Adeeko, O., & Ishola, A. M. (2024). Examining artificial intelligence literacy among pre-service teachers for future classrooms. *Computers and Education Open*, 6, 100179. <https://doi.org/10.1016/j.caeo.2024.100179>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- CAST. (2018). Universal design for learning guidelines (Version 2.2). Author. <https://udlguidelines.cast.org>
- Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J. L., Clowez, G., Boileau, P., & Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research*, 26, e53164. <https://doi.org/10.2196/53164>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Du, J., & Mohd Nordin, N. R. (2026). The impact of the iWrite automated writing evaluation system on university EFL students' writing performance and writing anxiety. *Education Sciences*, 16(3), 411. <https://doi.org/10.3390/educsci16030411>
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. <https://doi.org/10.3389/frai.2023.1162454>
- Hwang, H.-B., Coss, M. D., Loewen, S., & Tagarelli, K. M. (2024). Acceptance and engagement patterns of mobile-assisted language learning among non-conventional adult L2 learners: A survival analysis. *Studies in Second Language Acquisition*, 46(4), 969–995. <https://doi.org/10.1017/S0272263124000354>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Ravselj, D., Keržič, D., Tomažević, N., Umek, L., Brezovar, N., Iahad, N. A., et al. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLOS ONE*, 20(2), e0315011. <https://doi.org/10.1371/journal.pone.0315011>
- Tour, E., Pegrum, M., & Macdonald, S. (2025). Engaging English language learners in AI literacy practices: A conceptual framework and practical strategies for educators. *English Australia Journal*, 41(1), 27–46. <https://doi.org/10.61504/YBUY9086>

- UNESCO. (2024). AI competency framework for teachers. United Nations Educational, Scientific and Cultural Organization. <https://www.unesco.org/en/articles/ai-competency-framework-teachers>
- Wang, Y., & Xue, L. (2024). Using AI-driven chatbots to foster Chinese EFL students' academic engagement: An intervention study. *Computers in Human Behavior*, 159, 108353. <https://doi.org/10.1016/j.chb.2024.108353>
- Younis, B. (2024). Effectiveness of a professional development program based on the instructional design framework for AI literacy in developing AI literacy skills among pre-service teachers. *Journal of Digital Learning in Teacher Education*, 40(3), 142–158. <https://doi.org/10.1080/21532974.2024.2365663>
- Zakaria, N. Y. K., & Hashim, H. (2024). Shaping the future of education: Conceptualising pre-service teachers' perspectives on artificial intelligence (AI) integration. *International Journal of Academic Research in Business and Social Sciences*, 14(5), 643–653. <https://doi.org/10.6007/ijarbss/v14-i5/21584>
- Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research*, 61(4), 875–900. <https://doi.org/10.1177/07356331221127300>
- Zhang, C., Hu, M., Wu, W., Kamran, F., & Wang, X. (2024). Unpacking perceived risks and AI trust influences pre-service teachers' AI acceptance: A structural equation modeling-based multi-group analysis. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12905-7>