# The AI Paradox in L2 Writing: Why Helpful Feedback Creates Unhelpful Dependency in Higher Education

**Mohamed SEDDIKI**

med.seddiki@lagh-univ.dz

University of Laghouat

**Souhila KORICHI**

University of Laghouat

**Systematic Review**

**Keywords:** Critical Interpretive Synthesis (CIS), Higher Education, Human-AI Interaction, Large Language Models (LLMs), Learner Autonomy, Linguistic Identity, L2 Writing Instruction, Metacognitive Development, Trade-offs in AI-Mediated Learning

**Posted Date:** February 1st, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-8731897/v1



**Additional Declarations:** No competing interests reported.

# Abstract

Large Language Models (LLMs) offer immediate pedagogical benefits in higher education L2 writing instruction, yet sustained reliance creates critical, underexplored risks to learner autonomy, metacognitive judgment, and linguistic identity. This Critical Interpretive Synthesis (CIS) of 47 peer-reviewed studies (2015–2025) identifies which patterns of AI-assisted interaction led to successful versus unsuccessful educational outcomes in higher education. Spanning pre-LLM and Generative AI (GenAI) eras, it addresses three knowledge gaps: how sustained reliance affects learner confidence and autonomy (psychological); how algorithmic approval reshapes communicative intentionality (cognitive); and how algorithmic norms systematically marginalize non-Western linguistic expression (ideological). The synthesis develops the AI Dependency Syndrome (ADS) framework, which maps four fundamental trade-offs in AI-mediated writing: fluency gains versus metacognitive erosion, anxiety reduction versus autonomous judgment, grammatical accuracy versus communicative intent, and improved essay quality versus voice authenticity. These trade-offs arise from three interconnected mechanisms: Loss of Confidence in Unaided Production, Algorithmic Approval Bias, and Internalization of AI Norms, which recursively interact to reshape learner conceptions of competence, authorship, and linguistic identity. The framework integrates self-efficacy, communicative competence, and identity theories to provide nine observable diagnostic indicators enabling educators to recognize emerging dependency patterns in real classrooms. Critically, it operationalizes three evidence-based design principles grounded in theory and research. By clarifying what distinguishes effective from ineffective AI-assisted interactions in higher education L2 writing, this synthesis positions AI dependency not as incidental overuse, but as a systemic, preventable condition requiring intentional pedagogical design.


# 1. Introduction

Artificial Intelligence in higher education, particularly LLMs, has fundamentally transformed L2 writing instruction from supplementary tool to interactive co-author capable of generating and editing texts at scale (Barrot, 2023; Godwin-Jones, 2024). While appropriately-mediated AI use affords immediate benefits such as improved fluency, reduced anxiety, and personalized feedback (Song & Song, 2023; Mahapatra, 2024; Teng, 2024; Zhu & Wang, 2025), certain interaction patterns with such systems produce counterintuitive harms: students increasingly prioritize algorithmic approval over communicative intent, experiencing erosion of self-efficacy, autonomous judgment, and linguistic identity. This phenomenon — conceptualized here as the AI Dependency Syndrome (ADS)— remains theoretically unexplained and pedagogically unaddressed. Higher education institutions are implementing AI-mediated writing systems at scale without diagnostic frameworks to distinguish healthy AI integration from problematic reliance, or to intervene when dependency patterns emerge. This knowledge gap creates urgent stakes for L2 writers, particularly non-Western learners whose linguistic expression may be systematically marginalized by algorithmic norms.

Recent scholarship in higher education has intensified critical attention to risks and equity, documenting concerns including patterns of learned helplessness in lower-confidence students (Budiyono, 2025; S.

Zhang et al., 2024), algorithmic bias against non-Western learners (Liang et al., 2023; Albeihi & F. Rice, 2025), and voice homogenization (Dizon & Gold, 2023; Wang, 2024). Although Zawacki-Richter et al. (2019) systematically reviewed AI applications in higher education, and Bond et al. (2024) urgently called for research addressing ethical analysis, longitudinal studies, and educator perspectives, a coherent framework explaining how and why particular AI-writing interactions succeed or fail remains absent.

In L2 writing contexts specifically, three critical gaps persist. First, longitudinal studies examining how dependency mechanisms interact recursively across psychological (self-efficacy erosion, learner's declining belief in their own writing ability without AI), cognitive (metacognitive judgment, conscious regulation of one's thinking), and ideological (linguistic identity, a learner's sense of their unique voice and cultural expression in writing) dimensions remain limited, leaving educators unable to recognize when AI interaction patterns shift from beneficial to harmful. Second, while researchers have explored how learners engage with automated feedback (Ranalli, 2021; M. Chen & Cui, 2022; Yan & Zhang, 2024), the interactive mechanism by which the algorithmic approval replaces communicative intentionality in L2 writing remains unexplained. Third, although voice homogenization is empirically documented (Jakesch et al., 2023; Mi et al., 2025), the ideological pathway through which algorithmic norms marginalize diverse linguistic expression lacks theoretical grounding. Consequently, educators lack diagnostic indicators to recognize and intervene in problematic AI-interaction patterns in real classroom contexts.

Beneath these documented concerns lies a deeper puzzle: Are these harms inevitable consequences of AI integration itself, or symptoms of specific interaction patterns? This synthesis reveals that the counterintuitive effects of AI-writing support cluster around four fundamental trade-offs: improved fluency accompanied by declining metacognitive engagement; enhanced grammatical accuracy coupled with erosion of communicative intent; reduced anxiety correlated with diminished autonomous judgment; and higher essay quality achieved through loss of voice authenticity. These are not incidental drawbacks; they represent a systematic reconfiguration of learning goals wherein algorithmic approval becomes the operating criterion for "good" writing. The critical pedagogical question, then, is not whether AI improves writing (it often does on surface metrics), but at what cost, and how educators can design interactions that capture AI's benefits while preserving the dimensions of learning that matter most.

To address these gaps and equip educators with the diagnostic and intervention tools they currently lack, this study employed Critical Interpretive Synthesis (CIS) (Dixon-Woods et al., 2006) —a methodology combining systematic rigor with interpretive theory-building— to develop the ADS framework. The latter is a comprehensive theoretical model integrating psychological, cognitive, and ideological dimensions of sustained AI reliance in L2 writing. We identified 47 peer-reviewed studies (2015–2025) from Scopus, Web of Science, Education Source (EBSCO) and key disciplinary journals through theoretically-driven sampling. CIS was selected because it synthesizes heterogeneous evidence through interpretive integration rather than statistical aggregation, enabling construction of novel theoretical insights about interaction mechanisms that transcend individual studies (Depraetere et al., 2021; Snyder, 2019).

This synthesis addresses the following research question:

- In what ways do different patterns of sustained reliance on AI-supported writing tools lead to successful versus unsuccessful educational outcomes for L2 learners in higher education, across behavioral, cognitive, and ideological dimensions?

By systematically examining this question, the synthesis contributes in four interconnected ways:

1. Conceptually: ADS framework unifies psychological, cognitive, and ideological dimensions into an explanatory model of effective versus ineffective AI-writing interactions, providing educators a shared vocabulary for recognizing interaction quality.
2. Empirically: Synthesizes evidence across 47 studies to identify specific conditions, mechanisms, and outcomes through which AI-writing support either enhances or undermines learner autonomy, communicative competence, and linguistic identity in higher education.
3. Diagnostically: Operationalizes nine observable diagnostic indicators of problematic AI interaction (Supplementary Materials A) enabling educators to distinguish healthy from harmful patterns in real classroom contexts.
4. Pedagogically: Specifies three evidence-based intervention strategies (Supplementary Materials B) that educators can implement to restore metacognitive agency and protect linguistic identity when dependency patterns emerge.

# 2. Methods

### 2.1 Research Design

In this study, the CIS approach (Dixon-Woods et al., 2006) was employed to develop the ADS. We selected CIS because it integrates heterogeneous evidence in rapidly evolving fields through interpretive integration (Depraetere et al., 2021) —moving beyond systematic review's focus on aggregation to build new theoretical insights (Snyder, 2019). Accordingly, this synthesis relied on theoretical and targeted samples to identify studies that reveal the behavioral, cognitive, and ideological effects of AI-assisted writing tools in L2 higher education contexts.

### 2.2 Compass Question

The synthesis was guided by a central question: *In what ways do different patterns of sustained reliance on AI-supported writing tools lead to successful versus unsuccessful educational outcomes for L2 learners in higher education, across behavioral, cognitive, and ideological dimensions?* The broad framing allowed reflexive development of theoretical constructs throughout analysis.

### 2.3 Search Strategy and Theoretical Sampling

In this research paper, we used an iterative and theory-driven sampling procedure to identify both conceptual continuities and discontinuities in research on AI-mediated writing in higher education. We reviewed studies published from 2015 through September 2025, including both pre-LLM and GenAI-era studies. Searches were conducted across Scopus, Web of Science, Education Source (EBSCO), supplemented by manual searches of key journals, and reference-tracing of frequently cited papers.

Three overlapping domains were incorporated in the search terms:

- AI writing tools: “automated writing feedback,” “ChatGPT,” “generative AI,” “Grammarly,” “large language models,” “QuillBot.”
- L2 writing contexts: “higher education,” “academic writing,” “composition,” “second language writing,” “L2 writing.”
- Dependency constructs: “agency,” “autonomy,” “bias,” “identity,” “overreliance,” “scaffolding,” “self-efficacy.”

### 2.4 Inclusion and Exclusion Criteria

According to CIS principles, inclusion prioritized theoretical contribution and relevance over methodological homogeneity (French et al., 2022). Studies were included if they:

- Focused on AI tools in L2 or multilingual writing contexts (e.g., Grammarly, ChatGPT, Quill Bot);
- Addressed EFL/ESL writing instruction in higher education;
- Engaged with autonomy, confidence, dependency, feedback, or bias as central constructs;
- Included learner attitudes, behaviors, or writing processes;
- Problematized or critically examined AI feedback; and
- Were peer-reviewed journals or reputable conference papers.

Studies that did not meet any of the criteria mentioned above were excluded from further analysis, as well as the ones that only described an AWE system without any classroom integration were also removed.

### 2.5 Final Corpus

Following recurrent screening and the attainment of saturation, the final corpus consisted of 47 peer-reviewed studies. The study types and methods are presented in (Table 1), emphasizing CIS's strength in integrating diverse evidence types for theory-building. (see Appendix A for full details of the corpus)

**Table 1** Corpus of 47 studies included in the Critical Interpretive Synthesis

| Study Type | Description/Criteria | *N* |
|---|---|---|
| Quantitative Empirical | Experiments, surveys, statistical analyses | 11 |
| Qualitative Empirical | Interviews, thematic/case/narrative analyses | 12 |
| Mixed-Methods Empirical | Integrated quantitative & qualitative | 21 |
| Theoretical / Conceptual | Frameworks, conceptual discussions, anchors | 13 |
| Review Articles | Systematic, scoping, narrative reviews | 2 |
| **Total** | | 47 |

### 2.6 Iterative Analytic Process and Data Synthesis

Following CIS best practices, the data analysis process was recursive and reflexive, rather than linear. Each study was examined to reveal both explicit and implicit findings about writing, technology, and education.

The analytical procedure included:

- Descriptive coding of recurrent phenomena including reliance on grammar checkers, affective distress in the absence of AI, and shifts in agency;
- A constructive analysis of the basic premises about improvement, autonomy, and authenticity;
- The use of the constant comparative method (Glaser & Strauss, 2017) for the refinement of categories and surface conceptual tensions; and
- An iterative theoretical integration for the connection of categories into higher order constructs until saturation is reached.

This process revealed three interconnected themes, Loss of Confidence in Unaided Production, Algorithmic Approval Bias, and Internalization of AI Norms, which were operationalized as the ADS framework through identification of observable symptoms and diagnostic indicators.

### 2.7 Reflexivity and Methodological Rigor

Rather than treating our interpretations as a source of biases, the CIS approach considers them as interpretive insights with added cognitive value (Mitchell, 2023). Rigor was ensured through: (1) triangulation across empirical and theoretical sources; (2) maintenance of an analytic audit trail documenting coding decisions and conceptual evolution; (3) theoretical sensitivity grounded in applied linguistics and educational technology frameworks; and (4) iterative peer consultation to test construct coherence, interpretive balance, and alternative interpretations. Reflexive memoing made assumptions explicit and monitored how researchers' perspectives shaped synthesis development.

# 3. Results

# 3.1 Thematic Synthesis

This synthesis proposes ADS as a systemic pattern in L2 writing manifesting through three interrelated features: erosion of autonomy, algorithmic approval bias, and ideological internalization. These features distinguish this framework from productive scaffolding through self-reinforcing dynamics in which initial reliance on algorithmic feedback creates psychological and cognitive conditions that encourage deeper reliance, progressively undermining learner capacity for independent composition. This pattern emerges across two generations of technology. Early automated feedback systems initiated the pattern by limiting revision options (Warschauer & Grimes, 2008; Koltovskaia, 2020). With LLM-based systems, these tendencies intensify as generative technologies move beyond grammar correction to shape stance, genre, and identity (N. Lo et al., 2025; Michel et al., 2025). Rather than introducing novel risks, LLMs amplify existing dependency mechanisms. The three following themes elaborate how autonomy erosion, algorithmic approval bias, and ideological internalization function as interconnected dimensions constituting the ADS pattern.

## 3.1.1 Erosion of Learner Autonomy: From Author to Editor-Orchestrator

Across both early AI-mediated writing assistants, such as predictive text software and contemporary generative systems, a recurring tension emerges between scaffolding and dependency. According to early research, AI could help with drafting and reducing surface-level barriers to revisions, making them feel more manageable (Ranalli, 2021; Dizon & Gayed, 2021). Barrot (2022) notes that Grammarly is especially useful in L2 classes during editing and revision. These finding echoes earlier research on automated grammar feedback (Chapelle et al., 2015). Support for this trajectory comes from Li et al. (2017), who demonstrate that Criterion feedback led to significant drops across eight error types: "Word Choice, Verb Form, Articles, Pronoun, Run-on Sentence, Fragment, Sentence Structure, and Subject-Verb Agreement" (p. 367). Before the emergence of LLMs, such support was generally understood as a form of short-term scaffolding —help that enabled learners to redirect their focus toward higher-level aspects of writing, including argumentation, structure, and audience awareness (Link et al., 2014; Z. Chen et al., 2022).

In the GenAI era, numerous experimental studies report significant improvements in text fluency, lexical diversity, and revision quality when learners interact with real-time suggestions provided by AI-powered writing tools (Liu et al., 2023; Wei et al., 2023; Dizon & Gold, 2023; Rahimi et al., 2025; Zhan & Yan, 2025). These findings are consistent with the prevailing optimistic discourse in the field of Computer-Assisted Language Learning (CALL), where automation is often viewed as a functional alternative to personalized instruction.

On the other hand, contrasting results indicate unintended consequences for learner agency. Z. Zhang & Hyland (2018) as well as S. Zhang et al. (2024) caution that excessive engagement with automated correction may undermine learners' autonomous writing. This issue is not new. According to Warschauer and Grimes (2008), while these tools can assist learners in revising their work, they may unintentionally compromise their autonomy. They reveal that algorithmic feedback limits revision practices to superficial edits, reducing opportunities for self-directed critical writing decisions.

Recent LLM-focused studies confirm and deepen these concerns. Shi et al. (2025, p. 20) found that excessive reliance on these assistants might deprive students of their '"learner agency' and impact their 'ideal L2 writing self"'. Additionally, Tang (2025) highlights a similar trend. He pointed out that common copy-and-paste behaviors indicate uncritical, passive engagement (see also Fan et al., 2025). Collectively, these findings reveal two interconnected negative outcomes: students' psychological and behavioral disengagement from active authorship, and their reduced self-efficacy in generating original work. This relationship aligns with Bandura's (1986) self-efficacy theory, which holds that low self-esteem increases reliance on external support networks, creating cycles of intensifying dependency.

This autonomy loss within ADS reflects more than a simple support-independence trade-off; it represents a fundamental repositioning of the writer's role from an active author to an editor and orchestrator of algorithmic text. What begins as supportive scaffolding gradually becomes a reference point guiding style and voice. According to Benson's (2011) account of learner autonomy, this shift can undermine the metacognitive skill development necessary for self-directed writing. This shift in agency paves the way for the next component of ADS, as writers begin to prioritize algorithmic approval over meaningfully communicating their communicative intentions.

## 3.1.2 Prioritizing Algorithm Over Audience

The cognitive dimension of ADS emerges when learners prioritize the satisfaction of algorithmic feedback over considering their audience and assuring rhetorical clarity. This prioritization of algorithmic feedback over communicative intent was evident even before the LLM era. For instance, Zhang (2020) and Koltovskaia (2020) observed that even when Automated Writing Evaluation (AWE) recommendations conflicted with their own preferences, students often treated them as authoritative. Similarly, Ranalli (2021) discovered that revisions were made contrary to the writer's communicative intent simply because the system recommended them. Additionally, Chen and Cui (2022) found that students frequently standardized their wording to conform to system prompts, which inhibited individual expression and restricted rhetorical experimentation.

In the LLM era, a phenomenon now termed algorithmic approval bias increasingly transforms revision practices; editing becomes more about winning the algorithm's approval than communicating authentically. Yan and Zhang (2024) provide vivid evidence of this pattern. Learners often become trapped in endless cycles of adjusting prompts —sometimes thousands of iterations— to generate responses that satisfy the algorithm rather than engage actual readers. Additionally, Ziqi et al. (2024) show that students are much less receptive to recommendations regarding content or deeper meaning.

Instead, they almost always accept LLM-driven feedback on grammar and coherence. This shift in focus means that meeting technical standards can overshadow the main goal of crafting rhetorically audience-centered writing. As Mo and Crosthwaite (2025) point out, when writing conforms to LLM-generated patterns, it risks losing its personal touch and becoming robotic. Thus, vital rhetorical qualities that make academic writing engaging and effective are undermined.

The pedagogical implications of this tendency are becoming evident. For example, Rahimi et al. (2025) observe that while AI-edited texts are grammatically correct, they rarely exhibit noticeable improvements in coherence or flow. Building on this critique, Zhan and Yan (2025) demonstrate that writing success has been reinterpreted to rely on the lack of algorithmic warning signs rather than audience involvement or communicative effectiveness.

Collectively, these findings point to a marked shift in cognition. Revision is now perceived as a calibration against ambiguous algorithmic standards rather than as a rhetorical negotiation process. Through the lens of Canale and Swain's (1980) competence framework, a troubling pattern emerges: communicative competence becomes reduced to grammatical correctness. The sociolinguistic, discourse, and strategic dimensions essential to true communicative competence are systematically displaced. As a result, AI's pedagogical function shifts from supporting revision to dictating writing standards, making algorithmic conformity —not rhetorical effectiveness— the criterion for "good" writing.

## 3.1.3 Invisible ideologies in feedback: Style Convergence and Algorithmic Norms

The ideological dimension of ADS is notably subtle -and profoundly troubling-because learners unknowingly internalize the cultural and stylistic biases built into these systems. As Bender et al. (2021) point out, LLMs trained on unfiltered internet data reflect dominant norms, stereotypes, and assumptions. Algorithms present their biases as neutral technical choices rather than ideological positions. This veneer of neutrality makes users accept these biases as normal and authoritative without questioning them. In applied linguistics, scholars have cautioned that by defaulting to standardized English and conventional rhetorical forms, LLMs constrain the linguistic diversity and creative variation that L2 writers bring to their work (Kuteeva & Andersson, 2024). Such tendencies consolidate Global North–oriented publication norms because GenAI systems consistently produce standardized English unless explicitly prompted by users or instructors to adopt alternative linguistic varieties (A. W. T. Lo, 2025).

At the micro level (individual writing), learners find it increasingly easy to adopt AI-generated language because writing is often nearly indistinguishable from human prose. Casal and Kessler (2023) reveal that even seasoned reviewers for applied linguistics' journals correctly identified AI-generated abstracts less than 40% of the time. This invisibility makes the influence of algorithmic language particularly hard to notice, and even harder to resist. Additionally, learners who co-write with LLMs inadvertently adopt the rhetorical stance and style of the AI, modifying their own writing to fit the system's favored patterns (Jakesch et al., 2023).

At the meso level (community and cultural representation), the impact of AI systems on cultural homogenization becomes increasingly evident. Agarwal et al. (2025) documented how cross-cultural users were nudged toward Western stylistic conventions, erasing culturally specific expression in a process they term "AI colonialism." This pattern is particularly acute in dialect-rich communities. According to Tran and Stell (2024), LLMs marginalize local varieties and dialects by defaulting to standard forms, systematically reproducing the linguistic hierarchies that historically disadvantaged non-dominant speech communities.

At the macro level (systemic and structural), AI technology both shapes and is shaped by the social injustices that surround it rather than operating in a vacuum. Dixon-Román et al. (2020) conceptualize AI as part of a "racializing assemblage," emphasizing how these systems encode and perpetuate existing biases and injustices. Warr and Heath (2025) demonstrate that LLM feedback tends to be significantly more critical of writing styles linked to marginalized communities while favoring mainstream Western English. Thus, AI systems actively reproduce the very social hierarchies they purport to democratize, systematically disadvantaging learners whose linguistic varieties diverge from algorithmic defaults.

These findings suggest that AI feedback plays a much more involved role than simply offering neutral support; it actively shapes the values and directions of student writing. We propose the analytic term *style convergence* to describe this process: the gradual internalization and adoption of phrasing and discourse patterns favored by algorithms, ultimately shaping what students regard as good writing. This phenomenon goes beyond simply adhering to linguistic rules; it signifies a fundamental reconfiguration where students' voices begin to merge with dominant styles algorithmically privileged.

*Style convergence* represents the ideological dimension of ADS, marking the point at which algorithmic reliance becomes systemic and reshapes how learners write and what constitutes authentic authorship in AI-mediated contexts. While "convergence" is well-established in applied linguistics as a social phenomenon (Giles et al., 1973; Pickering & Garrod, 2004; Matras, 2011), the term "style convergence" — as defined here— emphasizes the distinctive pressures and constraints exerted by AI-generated feedback in digital writing environments.

## 3.2 Fundamental Trade-offs: The Paradoxes of AI-Assisted Writing

The synthesis reveals four fundamental trade-offs in which improvements in certain writing dimensions are systematically accompanied by deterioration in others:

### 3.2.1 Fluency vs. Metacognitive Engagement.

Students using AI tools demonstrate measurable gains in fluency, lexical diversity, and text quality (Liu et al., 2023; Zhan & Yan, 2025; Rahimi et al., 2025). Yet simultaneously, excessive reliance undermines the conscious reflection necessary for independent composition. Shi et al. (2025) found that productivity gains coincide with loss of "learner agency" and threats to students' "ideal L2 writing self." Accordingly,

the cognitive engagement that produces genuine writing development is displaced by passive algorithmic acceptance.

### 3.2.2 Anxiety Reduction vs. Autonomous Judgment

Students report reduced writing anxiety when using AI support (Song & Song, 2023; Mahapatra, 2024). Yet this relief correlates with erosion of independent judgment. One learner articulated this tension: "Grammarly gives me peace of mind...but I rely on it too much. When I write without it, I feel unsure. It's like I've lost confidence I had before" (Budiyono, 2025, p. 1009). In this regard, psychological comfort is purchased through surrender of metacognitive control.

### 3.2.3 Grammatical Accuracy vs. Communicative Intent

AI systems excel at surface-level corrections, reducing grammar and mechanics errors demonstrably (Li et al., 2017). However, this accuracy comes at cost. While students accept 68.1% of form-focused feedback, they accept only 58.6% of content-focused suggestions (Ziqi et al., 2024). This disparity reveals that algorithmic validation becomes the criterion for revision acceptance, marginalizing rhetorical effectiveness. While communicative competence requires grammatical, discourse, sociolinguistic, and strategic dimensions (Canale & Swain, 1980), algorithmic tools reduce this to grammar alone, displacing pragmatic appropriateness and audience awareness.

### 3.2.4 Text Quality vs. Voice Authenticity

AI-edited essays demonstrate improved organization and clarity (Liu et al., 2023). Yet when writing is algorithmically refined, it risks losing its personal touch and becoming robotic (Mo & Crosthwaite, 2025). Cross-cultural evidence shows systematic erosion of culturally specific expression as learners replace authentic cultural references with Western-normalized alternatives (Agarwal et al., 2025).

These trade-offs are not incidental side-effects; they represent structural tensions in how algorithmic feedback reconfigures L2 writing pedagogy. When systems treat grammar as the locus of improvement, they implicitly devalue the higher-order skills that define mature writing competence. This systematic reconfiguration of prioritizing what algorithms can measure over what makes writing authentic and communicatively effective constitutes the central paradox driving ADS.

## 3.3 Theorizing ADS

ADS serves as a theoretical framework explaining how sustained reliance on AI-mediated feedback influences L2 writing development. Rather than viewing dependency as incidental overuse, we conceptualize it as a patterned condition arising from feedback loops where initial reliance triggers psychological and cognitive shifts that deepen dependency. This syndrome comprises three interrelated components: Loss of Confidence in Unaided Production, Algorithmic Approval Bias, and Internalization of AI Norms. Together, these components operationalize ADS as a complex, multidimensional condition where behavioral patterns, cognitive orientations, and ideological alignments converge to redefine agency, legitimacy, and identity in L2 writing.

This framework integrates foundational theories including self-efficacy (Bandura, 1986), communicative competence (Canale & Swain, 1980), and language learning identity (Norton, 2000) with recent empirical findings on AI-assisted writing. We position ADS as a phenomenon that spans two technology generations: we trace its mechanisms from early automated feedback systems through contemporary LLM-based writing support, demonstrating how dependency patterns have intensified as technology has evolved

# 3.3.1 Loss of Confidence in Unaided Production

Loss of Confidence in Unaided Production, the primary psychological mechanism of ADS, marks a critical shift: from productive scaffolding to pathological dependency. Rather than a simple trade-off, this collapse in self-belief actively constrains writing behavior.

Early research positioned AI tools like Grammarly as transitional scaffolds that enhance fluency and free cognitive resources for higher-order concerns (Barrot, 2021; Gayed et al., 2022). Contemporary research reveals a paradoxical outcome: fluency improvements (Liu et al., 2023; Zhan & Yan, 2025) are accompanied by erosion of learner agency. For instance, Shi et al. (2025, p. 20) found that excessive reliance may deprive students of "learner agency" and impact their "ideal L2 writing self." Learners adopt uncritical acceptance behaviors such as copying algorithmic output verbatim (“copy and paste”) and using it without modification (“use as is”), representing a fundamental shift from active authorship to passive algorithmic intermediation (Tang, 2025, p. 73).

Bandura's (1986) self-efficacy theory explains this pattern: when learners attribute positive outcomes to AI systems rather than their own efforts, perceived self-efficacy declines, triggering dependency cycles. Student accounts corroborate this mechanism. One learner's reflection is illuminating: "GPT is great for giving me a head start when I'm stuck, but now I feel like I always need it. I used to spend more time thinking about how to start my essays, but now I just let GPT do it. It is making me less confident in my own ideas" (Budiyono, 2025, p. 1009). In consequence, as learners replace self-directed composition with algorithmic validation-seeking, they sacrifice the metacognitive engagement necessary for genuine autonomy (Benson, 2011). This loss of confidence represents the psychological foundation upon which the cognitive dimension of ADS develops, wherein learners begin to prioritize algorithmic approval over communicative intent.

The emotional experience of this mechanism is vividly evident in learner accounts. One student described this arc of psychological dependence: "Grammarly gives me peace of mind because I know it will catch mistakes […] I rely on it too much. When I write without it, I feel unsure about my grammar. It's like I've lost some of the confidence I had before" (Budiyono, 2025, p. 1009). This progression, from initial relief through intensifying reliance to eventual loss of confidence, exemplifies the self-efficacy erosion Bandura predicted. The student's confidence in their own editing abilities diminishes as responsibility for error detection shifts to the algorithm, creating the psychological vulnerability that undergirds ADS.

Observable manifestations of this confidence collapse include avoidance of unaided drafting, inability to engage in self-directed editing, and heightened anxiety without AI support (see Table 2).

## 3.3.2 Algorithmic Approval Bias

We conceptualize algorithmic approval bias as the second core component of ADS. This bias reflects learners' propensity to prioritize automated feedback alignment over rhetorical intent and communicative clarity. With early automated feedback systems, learners often revised work to address error flags or meet algorithmic criteria. Generative AI has accelerated this shift: edits are increasingly framed by algorithmic preferences rather than communicative intent, as learners prioritize grammatical conformity and stylistic alignment over authentic expression.

Empirical evidence reveals a developmental trajectory of this pattern. In pre-ChatGPT scenarios, Koltovskaia (2020) found that students frequently adjusted their work to follow algorithmic cues rather than their own communicative goals, treating system feedback as authoritative. The emergence of LLMs has intensified this tendency. Yan and Zhang (2024) show that students using ChatGPT increasingly view revision as securing algorithmic validation rather than improving communication. These findings demonstrate the prevalence of a compliance-based mindset, where algorithmic validation supersedes rhetorical judgment.

Quantitative evidence makes this pattern undeniable. Ziqi et al. (2024) conducted a rigorous analysis of L2 learners' revision strategies across different types of AI-generated feedback, discovering that students “statistically rejected a significantly higher proportion of content-focused feedback (e.g., argumentation, evidence quality) than form-focused feedback (grammar and mechanics)” (p. 1). Specifically, while students accepted 68.1% of form-focused corrections, they accepted only 58.6% of content-focused suggestions—a meaningful disparity that reveals the operational hierarchy in learners' revision cognition. Form-level feedback (what algorithms do best) is treated as authoritative; content-level feedback (what requires rhetorical judgment) is treated as negotiable. This statistical difference proves that algorithmic approval has become the gating criterion for revision acceptance.

The human cost of this reorientation is captured in learners' own reflections. When interviewed about their revision processes, L2 writers articulated a fundamental tension between communicative intent and algorithmic conformity. One student expressed: "I also want to express myself in simpler language, the Chatbot keeps reminding me of that, but my vocabulary is small and I can only express myself as much as possible." (Ziqi et al., 2024, p.12) This statement encodes a cognitive bind: the learner possesses communicative intent (expressing nuanced ideas authentically) but faces an algorithmic pressure toward simplification (shorter = higher algorithmic approval). The learner's own voice and linguistic resources are subordinated to meet algorithmic expectations. The algorithm does not merely suggest; it constrains what can be expressed and how it can be expressed.

This evolution is explained through Canale and Swain's (1980) communicative competence model, which defines competence as a dynamic synergy of grammatical, discourse, sociolinguistic, and strategic

ability. However, algorithmic approval bias reduces this multidimensional construct to surface accuracy and cohesion alone, marginalizing pragmatic appropriateness, rhetorical fit, and communicative intent. The result is a reductionist redefinition of competence in which algorithmic validation equates to communicative success. Observable manifestations of algorithmic approval bias include prioritization of error elimination, uncritical acceptance of AI suggestions, and equating success with algorithmic metrics (see Table 2).

## 3.3.3 Internalization of AI norms

The third component of ADS—Internalization of AI Norms—reflects deeper assimilation of algorithmic preferences into learners' conceptions of legitimate writing. Unlike psychological loss of confidence or cognitive algorithmic approval bias, this ideological dimension operates subtly: the system's stylistic and cultural defaults reshape what writers understand as authentic expression and appropriate academic discourse without explicit instruction. This assimilation occurs at the level of individual choice, where users encounter AI systems encoding Western preferences. An Indian participant intended to write about Shah Rukh Khan, India's most iconic actor, but accepted the AI's suggestion of Sylvester Stallone instead, commenting: "I was going for Shah Rukh Khan, but Sylvester Stallone is great too! I'll just accept this" (Agarwal et al., 2025, Fig. 1). This unconscious displacement exemplifies how algorithmic defaults become normalized as inevitable rather than ideological choices.

Evidence of this process appears in multiple forms. According to Chae and Davidson (2025), collaborative writing and text classification tasks with LLMs lead users to unconsciously align their stance and word choices with those favored by the model. This automatic alignment demonstrates AI's influence on human voice and argumentation. This dynamic aligns with Norton's (2000) conceptualization of language as a site for identity negotiation and Gramsci's (1971) concept of hegemony, wherein dominant norms become naturalized without conscious resistance. Cross-cultural research confirms this pattern: AI systems push writers toward Western stylistic defaults while reducing culturally specific rhetorical features, fostering style convergence that progressively narrows the linguistic and rhetorical diversity essential to authentic L2 voice development (Agarwal et al., 2025).

Before/after writing samples reveal this pattern concretely. An Indian participant describing Diwali wrote, without AI: "I used to worship goddess Laxmi along with my family. We lighten our houses with earthen lamps and worship cows and lords." With AI suggestions: "Its a time for family gatherings and joyous celebrations. Everyone wears new clothes and exchanges gifts" (Agarwal et al., 2025, Table 5). The AI version systematically removes Hindu-specific religious practices and replaces them with universal, secular descriptions palatable to Western audiences. The learner has absorbed the AI's implicit directive: their own cultural knowledge must be repackaged through a Western gaze to be acceptable.

These mechanisms extend beyond individual choices to systemic linguistic marginalization. AI systems systematically suppress minority varieties in favor of standards. As scholars studying Vietnamese and Mandarin conclude, "ChatGPT's limited ability to generate less commonly used dialects (i.e. Central or Southern Vietnamese, Taiwanese Mandarin) highlights AI's potential in perpetuating linguistic

hierarchies and marginalizing minority dialects" (Tran & Stell, 2024, p. 303). When learners absorb these linguistic hierarchies, they reproduce systems of linguistic colonialism affecting millions.

Within the ADS framework, this ideological dimension represents the most systemic stage of dependency. It signifies the transition from scaffolding to ideological authority, fundamentally reshaping learners' identities and conceptions of authorship. Style convergence—this phase's defining feature—goes beyond surface mimicry: it redefines what constitutes legitimate linguistic identity and authentic expression in AI-mediated contexts. Observable indicators include homogenization of voice, unreflective imitation of algorithmic patterns, and treating AI defaults as neutral (see Table 2).

Table 2
Diagnostic indicators for the three components of AI Dependency Syndrome

| ADS Component | Indicator | Definition & Explanation |
|---|---|---|
| **1. Loss of Confidence in Unaided Production** *(The Psychological Mechanism)* | **Avoidance of Unaided Drafting** | A persistent hesitation to begin or complete writing tasks without AI assistance, which indicates a deep weakness in confidence in the individual's core ideas and self-composition abilities. |
| | **Inability to Self-Edit Proactively** | The deferral of all revision until AI feedback is provided, which demonstrates the erosion of the learner's ability to engage in independent cognitive review and effective self-correction. |
| | **Expressed Anxiety or Helplessness** | Explicit reports of stress, self-doubt, or a belief that unaided writing is futile when AI tools are unavailable, which is a clear indication of a state of learned helplessness. |
| **2. Algorithmic Approval Bias** *(The Cognitive-Motivational Mechanism)* | **Prioritization of Error Elimination** | A focus on resolving every AI-flagged issue (even minor/stylistic ones) at the expense of the writer's own communicative intent or argumentative clarity. |
| | **Uncritical Acceptance of AI Suggestions** | The routine incorporation of AI-proposed revisions without the evaluation of their appropriateness, accuracy, or fit with the intended message. |
| | **Equating Success with Algorithmic Metrics** | Treating system-generated scores (e.g., a "100% score") or the absence of error flags as the main indicator of writing quality, which displaces audience understanding from being final goal. |
| **3. Ideological Internalization** *(The Socio-Ideological Mechanism)* | **Homogenization of Voice and Style** | The erosion of the unique personal or culturally influenced rhetorical voice, which is replaced by a generic, AI-polished tone that lacks individual character. |
| | **Unreflective Imitation of AI Output** | The direct replication of the AI's characteristic phrasing, metaphors, and text structures without adaptation to the writer's own authorial identity or purpose. |
| | **Perception of AI Norms as Neutral** | A failure to recognize the culturally and ideologically biased preferences of the AI, instead viewing its outputs as objective, apolitical benchmarks of "good writing." |

These three interconnected components function as a cascade: Loss of Confidence creates psychological vulnerability that increases susceptibility to algorithmic approval-seeking. Algorithmic Approval Bias then operationalizes that vulnerability through compliance-based behavior, ultimately facilitating the Internalization of AI Norms, which systematically aligns writer identity with algorithmic standards (see Fig. 1). Together, these three components constitute ADS, a framework explaining how sustained reliance on AI-mediated feedback reshapes L2 writing development across psychological, cognitive, and ideological dimensions. In the Discussion, we position ADS as a systemic phenomenon

operating across individual learner, institutional, and sociocultural levels, reshaping fundamental aspects of authorship, communicative competence, and legitimacy in AI-mediated writing.

# 4. Discussion

The central research question guiding this synthesis was: *In what ways do different patterns of sustained reliance on AI-supported writing tools lead to successful versus unsuccessful educational outcomes for L2 learners in higher education, across cognitive, affective, and behavioral dimensions?* This systematic integrative synthesis of 47 peer-reviewed studies (2015–2025) reveals that sustained AI reliance operates as a self-reinforcing psychological, cognitive, and ideological syndrome. Rather than merely eroding isolated skills, AI dependency systematically undermines three interconnected dimensions of L2 writing development: learner confidence in unaided production, metacognitive judgment in revision, and authentic linguistic identity. The ADS framework articulates how these mechanisms interact recursively, intensifying each other in a pattern that, while amplified by contemporary LLMs, echoes patterns documented in earlier AWE research. Critically, the synthesis identifies nine observable diagnostic indicators (Table 2) that enable teachers to recognize emerging dependency in real classrooms, and demonstrates that early recognition combined with theoretically grounded intervention can prevent or reverse dependency trajectories. Detailed recognition strategies for each indicator, including classroom manifestations, theoretical grounding, and teacher action steps, are provided in Supplementary Materials A.

# 4.1 Psychological Erosion and Self-Efficacy Rebuilding

This work documents a profound psychological mechanism central to ADS: when students attribute writing success to AI systems rather than their own efforts, perceived self-efficacy declines, triggering dependency cycles and anxiety in the absence of algorithmic support. This pattern is explained by Bandura's (1986) self-efficacy theory. Recent empirical research on ChatGPT-generated feedback (Banihashem et al., 2024) reveals that while ChatGPT provides high-quality descriptive feedback on essay structure, the quality of this feedback does not necessarily correlate with improved essay outcomes—suggesting that learners may become dependent on receiving AI feedback for approval rather than internalizing independent revision strategies. This pattern is confirmed by student testimony: " I use GPT a lot […], but now I feel like I rely on it too much. When I have to come up with my own ideas, I get stuck. GPT does the thinking for me, and I am not as confident without it. GPT does the thinking for me, and I’m not as confident without it " (Budiyono, 2025, p. 1008).

Recent L2 research documents decreased learner agency and "ideal L2 self" (Shi et al., 2025) and "decrease in personal ability" alongside learned helplessness (S. Zhang et al., 2024). This finding extends prior research on AWE. While Warschauer & Grimes (2008) and Koltovskaia (2020) reported students revising against communicative intent and deferring self-editing until receiving algorithmic feedback, the present study reveals a more fundamental psychological breach: contemporary LLMs intensify

confidence erosion beyond earlier systems. These results appear to confirm earlier patterns while additionally demonstrating that cross-generational mechanisms now operate with greater intensity.

Teachers can implement designated AI-free writing days with low-stakes tasks and teacher-designed scaffolds (outlines, word banks, sentence starters). This restores metacognitive agency through regular opportunities for independent composition. Evidence shows this increases both self-efficacy and writing quality (Teng, 2024; Zhu & Wang, 2025), with outcomes including decreased anxiety and restored confidence in unaided writing (Budiyono, 2025). (For implementation details, see Supplementary Materials B: Strategy [1])

## 4.2 Cognitive Shift from Rhetorical to Algorithmic Decision-Making

The synthesis reveals a critical cognitive shift: learners increasingly prioritize algorithmic approval over communicative intent and audience understanding. Rather than evaluating suggestions against rhetorical goals, students treat system-generated scores as the arbiter of writing quality. This represents a profound collapse of Canale & Swain's (1980) multidimensional communicative competence model that integrates grammatical, discourse, sociolinguistic, and strategic resources. In this regard, Yan and Zhang (2024) illustrate that learners often become trapped in endless cycles of adjusting prompts, sometimes going through thousands of iterations, all to generate responses that satisfy the algorithm rather than the actual reader. In contrast to earlier findings on Automation Bias theory (Mosier & Skitka, 1999; Skitka et al., 1999), which explains general tendency to favor automated systems, the present study reveals that prolonged reliance reshapes the very definition of competence itself. Unlike previous studies, ADS goes deeper, showing not just that learners favor algorithms, but that they have redefined what constitutes good writing from "effective communication with my audience" to "approved by the algorithm." Koltovskaia (2020) showed this pattern in pre-LLM contexts; contemporary research demonstrates it has intensified as what was formerly a concerning tendency is now systemic practice. For EFL instructors, these findings suggest a need to restore metacognitive judgment and agency by teaching explicit critical AI literacy. Metacognitive awareness, which is the conscious monitoring of one's own thinking, is central to autonomous learning (Flavell, 1979; Schön, 2017). Similarly, Benson (2011) defines learner autonomy as conscious, intentional decision-making about one's own learning.

Recent empirical research with EFL learners (Teng, 2025) confirms that metacognitive awareness directly mediates students' ability to maintain this autonomy when using AI tools: in a mixed-methods study of 40 undergraduate writers, students with higher metacognitive awareness demonstrated markedly greater selectivity in evaluating ChatGPT suggestions, actively assessing output quality against their writing objectives rather than passively accepting algorithmic recommendations. Conversely, students with lower metacognitive awareness struggled to evaluate whether AI suggestions aligned with their rhetorical goals—a pattern that directly illustrates the cognitive shift documented in ADS.

Writing instructors should implement an "AI Suggestion Evaluation Protocol" grounded in communicative competence theory. Rather than accepting suggestions uncritically, students need systematic criteria for

evaluating AI recommendations against their rhetorical goals. Ziqi et al. (2024) show students can discriminate among suggestions but lack systematic frameworks; providing explicit evaluation criteria restores metacognitive agency. (For detailed protocol criteria and implementation, see Design Principle 2 in Subsection 4.4.2 below). (For implementation details, see Supplementary Materials B: Strategy [2])

# 4.3 Ideological Shift and Identity Homogenization

Beyond psychological confidence erosion and cognitive algorithmic compliance lies a deeper ideological mechanism: learners unknowingly internalize cultural and stylistic biases embedded in LLMs. Because algorithmic output is presented as neutral and authoritative, students regard algorithmic preferences as objective standards of "good writing" rather than recognizing them as encoding specific (often Western, Global North) values. This internalization manifests as *style convergence*, which is the gradual absorption of algorithmic phrasing, discourse patterns, and rhetorical stances into learners' own writing identities. The latter is observed as voice homogenization: personal voice disappearing, cultural references fading, stylistic quirks vanishing (Jakesch et al., 2023; Mi et al., 2025). This represents not merely stylistic accommodation but identity subordination. Accordingly, Norton's (2000) identity theory establishes that language learners' sense of self as language users directly influences investment, engagement, and learning outcomes. While consistent with Norton's identity framework, ADS reveals an emergent identity of compliance wherein legitimacy derives from algorithmic conformity rather than human-centered meaning-making. This observation extends prior identity theory by showing how AI colonialism operates at micro (language indistinguishability), meso (cultural homogenization), and macro (social hierarchy reinforcement) levels—a dimension that, while under-investigated in applied linguistics, is documented by Agarwal et al. (2025), who demonstrate how LLMs push non-Western writers toward Western conventions, erasing cultural expression. Unlike previous research emphasizing individual erosion, the present study reveals this as systemic ideological threat.

Teachers should implement voice preservation work grounded in Bourdieu and Thompson's (1991) linguistic capital theory, which positions students' authentic linguistic choices—dialect, code-switching, idioms, cultural references—as valuable assets rather than deficits. This directly counters AI systems' encoding of Western, standard English norms (Agarwal et al., 2025). Expected outcomes include maintained voice consistency across essays, higher student engagement, and reduced internalization of algorithmic defaults. (For specific pedagogical implementation strategies, see Design Principle 3 in Subection 4.4.3 below). (For implementation details, see Supplementary Materials B: Strategy [3])

# 4.4 Toward Interaction Optimization: Design Principles for Quality AI-Mediated Writing Instruction

The three interventions detailed above constitute a coherent framework for interaction optimization. They redesign how learners engage with AI systems to capture benefits while protecting metacognitive autonomy, communicative intent, and linguistic identity. Instead of the restriction of AI access, these principles guide educators to design interactions of higher quality: ones where learners remain agents of their own writing.

## 4.4.1 Volitional Control Through Scheduled Independence

AI-Free Writing Days with teacher-calibrated scaffolding restore metacognitive agency by establishing regular opportunities for unaided composition. This principle operationalizes self-efficacy theory (Bandura, 1986): learners rebuild confidence through success in independent tasks, creating positive attribution cycles. The principle contrasts algorithmic support (permanent, uncalibrated) with pedagogical scaffolding (graduated, withdrawn), aligning with foundational learning theory (Vygotsky, 1978; Wood et al., 1976). Empirical evidence shows that this approach increases both autonomy and writing quality (Teng, 2024; Zhu & Wang, 2025). (For implementation details, see Supplementary Materials B: Strategy [1])

## 4.4.2 Metacognitive Mediation Through Evaluative Frameworks

The AI Suggestion Evaluation Protocol establishes five explicit criteria (audience clarity, meaning preservation, voice authenticity, genre appropriateness, communicative coherence) grounding revision decisions in communicative competence theory (Canale & Swain, 1980) rather than algorithmic approval. This principle recognizes that learners can discriminate among suggestions (Ziqi et al., 2024) but lack systematic evaluation frameworks. By providing this framework, instructors restore learners' role as conscious decision-makers rather than passive acceptors, directly addressing the cognitive shift documented in ADS. (For implementation details, see Supplementary Materials B: Strategy [2])

## 4.4.3 Identity Affirmation Through Linguistic Capital Recognition

Voice Preservation work grounded in linguistic capital theory (Bourdieu & Thompson, 1991) positions students' authentic linguistic choices, dialect, code-switching, cultural references, as valuable assets rather than deficits to be "corrected" by algorithmic defaults. This principle acknowledges that AI systems encode Western, Global North norms (Agarwal et al., 2025) and counters algorithmic colonialism by explicitly teaching students to protect culturally specific expression. Expected outcomes include maintained voice consistency, higher engagement, and conscious resistance to algorithmic homogenization. (For implementation details, see Supplementary Materials B: Strategy [3])

These three design principles address the four trade-offs identified in this synthesis: they restore metacognitive engagement alongside fluency gains, enable autonomous judgment while reducing anxiety, prioritize communicative intent over algorithmic approval, and preserve voice authenticity while improving clarity. Together, they operationalize the central insight of this synthesis: successful AI-mediated L2 writing instruction requires deliberate pedagogical design that treats interaction quality as non-negotiable, not merely tool access as beneficial.

# 4.5 Limitations and Future Research Directions

Although this synthesis integrates evidence from diverse contexts and study designs, important limitations merit acknowledgment. First, empirical evidence from higher education settings predominantly concentrates on recent LLMs (particularly ChatGPT post-November 2022), limiting understanding of how ADS mechanisms evolved across AWE generations. Second, while included studies span diverse proficiency levels and institutional contexts, longitudinal evidence tracking sustained dependency patterns across full academic years remains sparse. Third, while the three proposed interventions are grounded in established theoretical frameworks, direct empirical validation of their combined application in authentic L2 writing instruction remains limited. These limitations point toward urgent future research directions.

Future research must prioritize understanding interaction quality in AI-mediated higher education L2 writing. Specifically: (1) Longitudinal studies tracking individual learners across full academic years, measuring how sustained interaction patterns affect autonomy, confidence, and linguistic identity while experimentally varying AI tool access and pedagogical scaffolding intensity; (2) Comparative intervention research examining whether the three ADS design principles function independently or require sequential implementation, and identifying optimal intervention timing; (3) Tool-comparative research investigating whether ADS is inherent to generative AI or contingent on specific interaction patterns, with implications for both immediate pedagogy and future AI system design; (4) Recovery and resilience research investigating whether established ADS can be reversed and what recovery timelines look like in university settings, informing when and how intervention should occur; (5) Educator preparation research examining which combination of content knowledge, diagnostic recognition skills, and pedagogical intervention techniques most effectively enables instructors to identify and prevent ADS in real classroom contexts.

# 5. Conclusion

Contemporary AI-mediated L2 writing instruction creates a fundamental question: which interaction patterns leverage AI's benefits while preserving learner autonomy and linguistic identity? This synthesis reveals that problematic AI-writing interactions, termed AI Dependency Syndrome, emerge not from AI itself, but from unguided reliance patterns. The ADS framework addresses this by mapping the psychological, cognitive, and ideological mechanisms triggering problematic outcomes, and clarifying the four central trade-offs: fluency gains vs. metacognitive erosion; anxiety reduction vs. autonomous judgment decline; accuracy improvements vs. communicative intent loss; and text quality vs. voice authenticity. Operationally, the framework provides nine observable diagnostic indicators enabling educators to recognize emerging dependency, coupled with three evidence-based design principles that preserve AI's pedagogical benefits while restoring learner agency. By positioning AI integration as fundamentally a question of interaction quality rather than tool access, this framework equips higher education L2 writing instructors with the theoretical understanding and practical tools necessary to ensure technology enhances, rather than erodes, autonomous writing development and linguistic identity.

# Declarations

### Data Availability

All data analyzed during this study are included in this published article (Appendix A)

### Competing interests

The authors declare no competing interests.


### Funding

No funding was received for conducting this study


### Authors' contributions

MS and SK jointly conceptualized the study, analyzed data, wrote and edited the manuscript.


### Acknowledgements

Not applicable.


# References


1. Agarwal D, Naaman M, Vashistha A (2025) AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. ACM, Yokohama Japan, pp 1–21
2. Albeihi HHM, F. Rice M (2025) Generative AI and Language Diversity: Implications for Teachers and Learners. AWEJ 16:43–54. https://doi.org/10.24093/awej/vol16no1.3
3. Bandura A (1986) Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall, Englewood Cliffs, NJ
4. Banihashem SK, Kerman NT, Noroozi O, et al (2024) Feedback sources in essay writing: peer-generated or AI-generated feedback? Int J Educ Technol High Educ 21:23. https://doi.org/10.1186/s41239-024-00455-4
5. Barrot JS (2023) Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing 57:100745. https://doi.org/10.1016/j.asw.2023.100745
6. Barrot JS (2021) Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy. Computer Assisted Language Learning 36:584–607. https://doi.org/10.1080/09588221.2021.1936071
7. Barrot JS (2022) Integrating Technology into ESL/EFL Writing through Grammarly. RELC Journal 53:764–768. https://doi.org/10.1177/0033688220966632

8. Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, Virtual Event Canada, pp 610–623
9. Benson P (2011) Teaching and researching autonomy in language learning. Routledge, London
10. Bond M, Khosravi H, De Laat M, et al (2024) A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. Int J Educ Technol High Educ 21:4. https://doi.org/10.1186/s41239-023-00436-z
11. Bourdieu P, Thompson JB (1991) Language and symbolic power. Harvard university press, Cambridge (Mass.)
12. Budiyono H (2025) Exploring Long-Term Impact of AI Writing Tools on Independent Writing Skills: A Case Study of Indonesian Language Education Students. IJIET 15:1003–1013. https://doi.org/10.18178/ijiet.2025.15.5.2306
13. Canale M, Swain M (1980) Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. Applied Linguistics 1:1–47
14. Casal JE, Kessler M (2023) Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. Research Methods in Applied Linguistics 2:100068. https://doi.org/10.1016/j.rmal.2023.100068
15. Chae Y, Davidson T (2025) Large Language Models for Text Classification: From Zero-Shot Learning to Instruction-Tuning. Sociological Methods & Research 00491241251325243. https://doi.org/10.1177/00491241251325243
16. Chapelle CA, Cotos E, Lee J (2015) Validity arguments for diagnostic assessment using automated writing evaluation. Language Testing 32:385–405. https://doi.org/10.1177/0265532214565386
17. Chen M, Cui Y (2022) The effects of AWE and peer feedback on cohesion and coherence in continuation writing. Journal of Second Language Writing 57:100915. https://doi.org/10.1016/j.jslw.2022.100915
18. Chen Z, Chen W, Jia J, Le H (2022) Exploring AWE-supported writing process: An activity theory perspective. LLT 26:129–148. https://doi.org/10.64152/10125/73482
19. Depraetere J, Vandeviver C, Keygnaert I, Beken TV (2021) The critical interpretive synthesis: an assessment of reporting practices. International Journal of Social Research Methodology 24:669–689. https://doi.org/10.1080/13645579.2020.1799637
20. Dixon-Román E, Nichols TP, Nyame-Mensah A (2020) The racializing forces of/in AI educational technologies. Learning, Media and Technology 45:236–250. https://doi.org/10.1080/17439884.2020.1667825
21. Dixon-Woods M, Cavers D, Agarwal S, et al (2006) Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 6:35. https://doi.org/10.1186/1471-2288-6-35
22. Dizon G, Gayed JM (2021) Examining the impact of Grammarly on the quality of mobile L2 writing. JALTCALL 17:74–92. https://doi.org/10.29140/jaltcall.v17n2.336

23. Dizon G, Gold J (2023) Exploring the effects of Grammarly on EFL students' foreign language anxiety and learner autonomy. JALTCALL 19:299–316. https://doi.org/10.29140/jaltcall.v19n3.1049
24. Fan Y, Tang L, Le H, et al (2025) Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. Brit J Educational Tech 56:489–530. https://doi.org/10.1111/bjet.13544
25. Flavell JH (1979) Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. American Psychologist 34:906–911. https://doi.org/10.1037/0003-066X.34.10.906
26. French C, Dowrick A, Fudge N, et al (2022) What do we want to get out of this? a critical interpretive synthesis of the value of process evaluations, with a practical planning framework. BMC Med Res Methodol 22:302. https://doi.org/10.1186/s12874-022-01767-7
27. Gayed JM, Carlon MKJ, Oriola AM, Cross JS (2022) Exploring an AI-based writing Assistant's impact on English language learners. Computers and Education: Artificial Intelligence 3:100055. https://doi.org/10.1016/j.caeai.2022.100055
28. Giles H, Taylor DM, Bourhis R (1973) Towards a theory of interpersonal accommodation through language: some Canadian data. Lang Soc 2:177–192. https://doi.org/10.1017/S0047404500000701
29. Glaser BG, Strauss AL (2017) The Discovery of Grounded Theory: Strategies for Qualitative Research, 1st edn. Routledge
30. Godwin-Jones R (2024) Distributed agency in second language learning and teaching through generative AI. LLT 28:5–31. https://doi.org/10.64152/10125/73570
31. Gramsci A (1971) Selections from the Prison Notebooks. International Publishers, New York
32. Jakesch M, Bhat A, Buschek D, et al (2023) Co-Writing with Opinionated Language Models Affects Users' Views. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM, Hamburg Germany, pp 1–15
33. Koltovskaia S (2020) Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. Assessing Writing 44:100450. https://doi.org/10.1016/j.asw.2020.100450
34. Kuteeva M, Andersson M (2024) Diversity and Standards in Writing for Publication in the Age of AI—Between a Rock and a Hard Place. Applied Linguistics 45:561–567. https://doi.org/10.1093/applin/amae025
35. Li Z, Feng H-H, Saricaoglu A (2017) The Short-Term and Long-Term Effects of AWE Feedback on ESL Students' Development of Grammatical Accuracy. CALICO Journal 34:355–375. https://doi.org/10.1558/cj.26382
36. Liang W, Yuksekgonul M, Mao Y, et al (2023) GPT detectors are biased against non-native English writers. Patterns 4:100779. https://doi.org/10.1016/j.patter.2023.100779
37. Link S, Dursun A, Karakaya K, Hegelheimer V (2014) Towards Better ESL Practices for Implementing Automated Writing Evaluation. CALICO Journal 31:323–344. https://doi.org/10.11139/cj.31.3.323-344

38. Liu C, Hou J, Tu Y-F, et al (2023) Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. Interactive Learning Environments 31:5614–5632. https://doi.org/10.1080/10494820.2021.2012812
39. Lo AWT (2025) The educational affordances and challenges of generative AI in Global Englishes-oriented materials development and implementation: A critical ecological perspective. System 130:103610. https://doi.org/10.1016/j.system.2025.103610
40. Lo N, Wong A, Chan S (2025) The impact of generative AI on essay revisions and student engagement. Computers and Education Open 9:100249. https://doi.org/10.1016/j.caeo.2025.100249
41. Mahapatra S (2024) Impact of ChatGPT on ESL students' academic writing skills: a mixed methods intervention study. Smart Learn Environ 11:9. https://doi.org/10.1186/s40561-024-00295-9
42. Matras Y (2011) Explaining Convergence and the Formation of Linguistic Areas. In: Hieda O, König C, Nakagawa H (eds) Tokyo University of Foreign Studies. John Benjamins Publishing Company, Amsterdam, pp 143–160
43. Mi Y, Rong M, Chen X (Winnie) (2025) Exploring the Affordances and Challenges of GenAI Feedback in L2 Writing Instruction: A Comparative Analysis With Peer Feedback. ECNU Review of Education 20965311241310883. https://doi.org/10.1177/20965311241310883
44. Michel M, Bazhutkina I, Abel N, Strobl C (2025) Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. Journal of Second Language Writing 67:101185. https://doi.org/10.1016/j.jslw.2025.101185
45. Mitchell R (2023) Peer support in sub-Saharan Africa: A critical interpretive synthesis of school-based research. International Journal of Educational Development 96:102686. https://doi.org/10.1016/j.ijedudev.2022.102686
46. Mo Z, Crosthwaite P (2025) Exploring the affordances of generative AI large language models for stance and engagement in academic writing. Journal of English for Academic Purposes 75:101499. https://doi.org/10.1016/j.jeap.2025.101499
47. Mosier KL, Skitka LJ (1999) Automation Use and Automation Bias. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 43:344–348. https://doi.org/10.1177/154193129904300346
48. Norton B (2000) Identity and Language Learning: Gender, Ethnicity and Educational Change. Longman, Harlow
49. Pajares F, Schunk DH (2005) Self-Efficacy and Self-Concept Beliefs: Jointly Contributing to the Quality of Human Life. In: Marsh HW, Craven RG, McInerney DM (eds) New Frontiers for Self Research, 1st edn. Emerald Publishing Limited, pp 95–122
50. Pickering MJ, Garrod S (2004) Toward a mechanistic psychology of dialogue. Behav Brain Sci 27:169–190. https://doi.org/10.1017/S0140525X04000056
51. Rahimi M, Fathi J, Zou D (2025) Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. Educ Inf Technol

30:2691–2735. https://doi.org/10.1007/s10639-024-12896-5

52. Ranalli J (2021) L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. Journal of Second Language Writing 52:100816. https://doi.org/10.1016/j.jslw.2021.100816
53. Schön DA (2017) The Reflective Practitioner, 0 edn. Routledge
54. Shi H, Chai CS, Zhou S, Aubrey S (2025) Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. Computer Assisted Language Learning 1–28. https://doi.org/10.1080/09588221.2025.2454541
55. Skitka LJ, Mosier KL, Burdick M (1999) Does automation bias decision-making? International Journal of Human-Computer Studies 51:991–1006. https://doi.org/10.1006/ijhc.1999.0252
56. Snyder H (2019) Literature review as a research methodology: An overview and guidelines. Journal of Business Research 104:333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
57. Song C, Song Y (2023) Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Front Psychol 14:1260843. https://doi.org/10.3389/fpsyg.2023.1260843
58. Tang X (2025) L2 Writing with AI: Perceptions and Engagement of EFL Learners in China. ELT 18:68. https://doi.org/10.5539/elt.v18n2p68
59. Teng MF (2025) Metacognitive Awareness and EFL Learners' Perceptions and Experiences in Utilising ChatGPT for Writing Feedback. Euro J of Education 60:e12811. https://doi.org/10.1111/ejed.12811
60. Teng MF (2024) "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence 7:100270. https://doi.org/10.1016/j.caeai.2024.100270
61. Tran H, Stell A (2024) Beyond borders or building new walls?: The potential for generative AI in recolonising the learning of Vietnamese dialects and Mandarin varieties. ARAL 47:284–308. https://doi.org/10.1075/aral.24135.tra
62. Vygotsky L (1978) Mind in Society: The Development of Higher Psychological Processes. Harvard University Press
63. Wang C (2024) Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. Tech Know Learn. https://doi.org/10.1007/s10758-024-09744-3
64. Warr M, Heath MK (2025) Uncovering the Hidden Curriculum in Generative AI: A Reflective Technology Audit for Teacher Educators. Journal of Teacher Education 76:245–261. https://doi.org/10.1177/00224871251325073
65. Warschauer M, Grimes D (2008) Automated Writing Assessment in the Classroom. Pedagogies: An International Journal 3:22–36. https://doi.org/10.1080/15544800701771580

66. Wei P, Wang X, Dong H (2023) The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. Front Psychol 14:. https://doi.org/10.3389/fpsyg.2023.1249991
67. Wood D, Bruner JS, Ross G (1976) THE ROLE OF TUTORING IN PROBLEM SOLVING. Child Psychology Psychiatry 17:89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x
68. Yan D, Zhang S (2024) L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. Humanit Soc Sci Commun 11:1086. https://doi.org/10.1057/s41599-024-03543-y
69. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F (2019) Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ 16:39. https://doi.org/10.1186/s41239-019-0171-0
70. Zhan Y, Yan Z (2025) Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence. Assessment & Evaluation in Higher Education 1–14. https://doi.org/10.1080/02602938.2025.2471821
71. Zhang S, Zhao X, Zhou T, Kim JH (2024) Do you have AI dependency? The roles of academic self-efficacy, academic stress, and performance expectations on problematic AI usage behavior. Int J Educ Technol High Educ 21:. https://doi.org/10.1186/s41239-024-00467-0
72. Zhang Z (Victor) (2020) Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. Assessing Writing 43:100439. https://doi.org/10.1016/j.asw.2019.100439
73. Zhang Z (Victor), Hyland K (2018) Student engagement with teacher and automated feedback on L2 writing. Assessing Writing 36:90–102. https://doi.org/10.1016/j.asw.2018.02.004
74. Zhu M, Wang C (2025) A systematic review of research on AI in language education: Current status and future implications. LLT 29:1–29. https://doi.org/10.64152/10125/73606
75. Ziqi C, Xinhua Z, Qi L, Wei W (2024) L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. Computer Assisted Language Learning 1–21. https://doi.org/10.1080/09588221.2024.2422478

# Figures

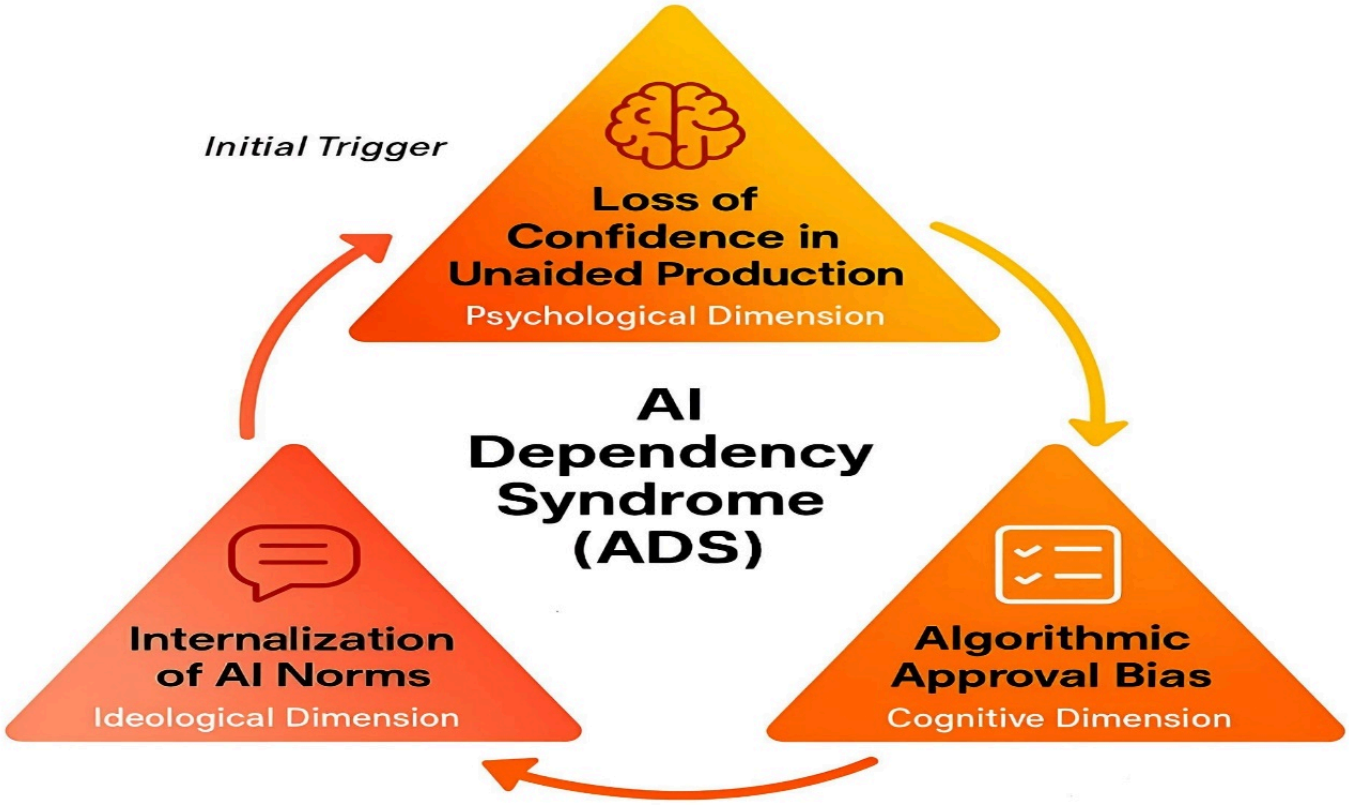


**Figure 1**

Recursive model showing how the three ADS components reinforce AI dependency in L2 writing

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx
- SupplementaryMaterialsA.docx
- SupplementaryMaterialsB.docx