Volume 6, Issue 3, October 2026

J-LELC
Journal of Language Education, Linguistics, and Culture
P-ISSN 2774-6003 E-ISSN 2775-099X

Between Professional Judgment and Digital Evidence: A Systematic Review of English Writing Lecturers' Cognitive Conflict in Digital-Age Writing Assessment

Irfan Masrur[1], Naily Inayatul Maghfirah[2]

Universitas Islam KH. Achmad Muzakki Syah[1,2]
irfan.masrur@gmail.com[1], nailymasrur@gmail.com[2]

Received: [date]	Revised: [date]	Accepted: [date]

Abstract

Digital submission platforms, automated writing evaluation, and generative artificial intelligence (GenAI) have unsettled the evidential basis on which English writing lecturers judge student texts. Research on AI-assisted writing, automated feedback, and teacher beliefs is expanding, yet it has rarely been synthesised around the cognitive conflict lecturers experience when their professional judgment collides with digital evidence and institutional expectations. This systematic literature review, reported in line with PRISMA 2020, examined how English writing lecturers experience and manage such conflict, what generates it, and which strategies they use to resolve it. Twenty-six full-text reports were assessed for eligibility, and 19 empirical studies published between 2014 and 2026 in 13 countries were analysed through thematic synthesis. Lecturers experienced conflict chiefly as epistemic uncertainty about authorship, as divergence between holistic judgment and machine output, and as an affective and identity burden marked by anxiety, guilt, fatigue, and a drift from assessor to "detective." Institutional and pedagogical sources appeared in all 19 studies, technological sources in 18, and ethical and student-related sources in 16 each; policy vacuums and unreliable detection were the most decisive triggers. Lecturers relied mainly on verification (16 studies) and communicative strategies (14 studies), whereas institutional responses were largely recommended rather than enacted (13 versus 5 studies). Many conflicts were therefore deferred or displaced rather than resolved. The review proposes a source–discrepancy–management–outcome model and argues that fair digital writing assessment requires process-visible task design, explicit AI-use disclosure, collegial moderation, and an expanded assessment literacy that includes the interpretation of algorithmic evidence.

Keywords: Cognitive Conflict, Digital Writing Assessment, Teacher Cognition, Systematic Literature Review

Citation:
Masrur, I., Maghfirah, N. I. 2026. Between Professional Judgment and Digital Evidence: A Systematic Review of English Writing Lecturers' Cognitive Conflict in Digital-Age Writing Assessment. *J-LELC: Journal of Language Education, Linguistics, and Culture*, 6(3), 160-174. https://doi.org/10.25299/j-lelc.2026.30558

UIR PRESS/J-LELC: Journal Language Education, Linguistics, and Culture
*This work is licensed under a **Creative Commons Attribution 4.0 International License** (CC BY 4.0)*

INTRODUCTION

Assessing student writing has always been an interpretive act. Lecturers read a text, weigh its content, organisation, accuracy, and voice against criteria, and infer from that product something about the writer's competence (Weigle, 2002; Knoch et al., 2021). This inference rests on an assumption that rarely needed stating: the submitted text was produced by the student whose name it carries. In digital higher education that assumption has weakened. Essays now arrive through learning management systems, pass through similarity checkers, are often polished with grammar tools, and may be partly or wholly drafted with GenAI (Warschauer et al., 2023; Perkins et al., 2024). The lecturer still assigns the grade, but the text's production history is largely invisible, and the digital tools that promise to make it visible frequently disagree with each other and with the lecturer.

The evidence on these tools explains why. AI-text detectors perform inconsistently and misclassify writing by non-native English users at a disproportionate rate (Liang et al., 2023; Weber-Wulff et al., 2023). Experienced ESL lecturers asked to identify machine-written essays reached contradictory verdicts and mistook fluent student work for AI output (Alexander et al., 2023). Automated scoring, meanwhile, tends to compress score ranges and reward surface accuracy while missing problems of coherence and elaboration (Anbreen et al., 2026). Lecturers therefore face assessment situations in which several sources of evidence point in different directions and no single source can be trusted on its own. A grade must nevertheless be given, defended, and recorded.

This review frames such situations as cognitive conflict. In the conceptual-change tradition, cognitive conflict refers to a perceived discrepancy between an existing schema and incoming evidence that cannot be ignored and therefore prompts re-evaluation. Cognitive dissonance, as formulated by Festinger (1957) and refined in later work (Harmon-Jones, 2019), is related but narrower: it is the aversive state that arises when a person holds inconsistent cognitions, particularly about their own beliefs and actions, and it motivates efforts to reduce the inconsistency. The two concepts are not interchangeable. Conflict is primarily epistemic (what is true of this text?), whereas dissonance is motivational and self-referential (how can I grade this way and still be the teacher I believe I am?). Both are relevant to assessment. Teacher cognition research has long shown that what language teachers believe and know is mediated by context and does not translate directly into practice (Borg, 2015); in writing feedback, for instance, lecturers trained to favour indirect correction often compromise toward direct correction under student and institutional pressure (Wei & Cao, 2020). Assessment literacy frameworks add that judging student work involves conceptual, praxeological, and socio-emotional dimensions, so that assessor identity and emotion are part of assessment competence rather than noise around it (Xu & Brown, 2016; Pastore & Andrade, 2019).

For the purposes of this review, cognitive conflict in writing assessment is defined as a lecturer's perceived incompatibility between, on the one hand, assessment beliefs or professional judgment and, on the other, digital evidence, institutional rules, or student claims, arising while a written product is being evaluated and requiring a decision. The definition separates conflict from general uncertainty, which may involve no competing cognitions, and from occupational stress, which may have no epistemic object. It also treats conflict as something that lecturers act upon, so that their experience and their management strategies can be studied together.

Relevant evidence is scattered across several research strands. Early work on automated writing evaluation (AWE) documented friction when machine scores contradicted instructors' grades (Link et al., 2014). Studies of emergency online assessment showed EFL teachers doubting the validity of transplanted examinations and struggling with assessor identity (Shao, 2023; Pham & Nguyen, 2026). Research on digital multimodal composing (DMC) exposed tensions between new genres and print-based testing regimes (Zuo & He, 2024). Since 2023, a rapidly growing body of studies has examined teachers' perceptions of ChatGPT, AI feedback, and AI detection (Nguyen, 2023; Zhang, 2025; Mirzayev et al., 2026). A recent narrative review mapped challenges and emerging practices in assessing AI-supported academic writing (Singh et al., 2026), but it focused on assessment design and governance rather than on how lecturers themselves experience and work through conflicting evidence. Consequently, it remains unclear whether the literature has sufficiently addressed (a) how English writing lecturers experience and manage cognitive conflict during digital writing assessment, (b) the technological, pedagogical, ethical, institutional, and student-related sources of that conflict, and (c) the strategies lecturers use to resolve or manage it, together with the outcomes of those strategies. These gaps are treated here as provisional and are tested through systematic searching rather than assumed.

Accordingly, this review addresses three questions. RQ1: How do English writing lecturers experience and manage cognitive conflict when assessing students' writing products in the digital age? RQ2: What are the sources of this cognitive conflict? RQ3: What strategies do English writing lecturers use to resolve or manage cognitive conflict, and with what reported outcomes? The review contributes theoretically by extending cognitive conflict and dissonance perspectives into technology-mediated writing assessment, pedagogically by identifying practices that support fair and transparent grading, and institutionally by informing AI-use policy, moderation procedures, and lecturer professional development.

RESEARCH METHOD

Research Design

The study adopted a systematic literature review (SLR) design with qualitative thematic synthesis. Reporting follows the PRISMA 2020 statement (Page et al., 2021), and the review logic follows the configurative approach to systematic reviewing described by Gough et al. (2017), in which the aim is to interpret and arrange concepts across studies rather than to aggregate effect sizes. This design suits the research questions because cognitive conflict is a perceptual and experiential construct reported almost entirely in qualitative and mixed-methods research. A protocol specifying the questions, eligibility criteria, extraction fields, and coding frame was drafted before the formal search.

Data Sources and Search Strategy

Searches were conducted in Scopus, Web of Science, ERIC, and Google Scholar, supplemented by backward and forward citation searching of key studies. The core search string combined four concept blocks: assessment ("writing assessment" OR "assessing writing" OR grading OR feedback), assessors (lecturer OR teacher OR instructor* OR faculty), digital technology ("generative AI" OR ChatGPT OR "automated writing evaluation" OR "AI detection" OR "online assessment" OR "digital multimodal"), and language context (EFL OR ESL OR "English for academic purposes" OR "academic writing"). The time window ran from January 2014 to September 2026. The lower boundary was set to capture pre-GenAI work on AWE, which offers a baseline for judging whether GenAI-era conflicts are new in kind or only in scale.

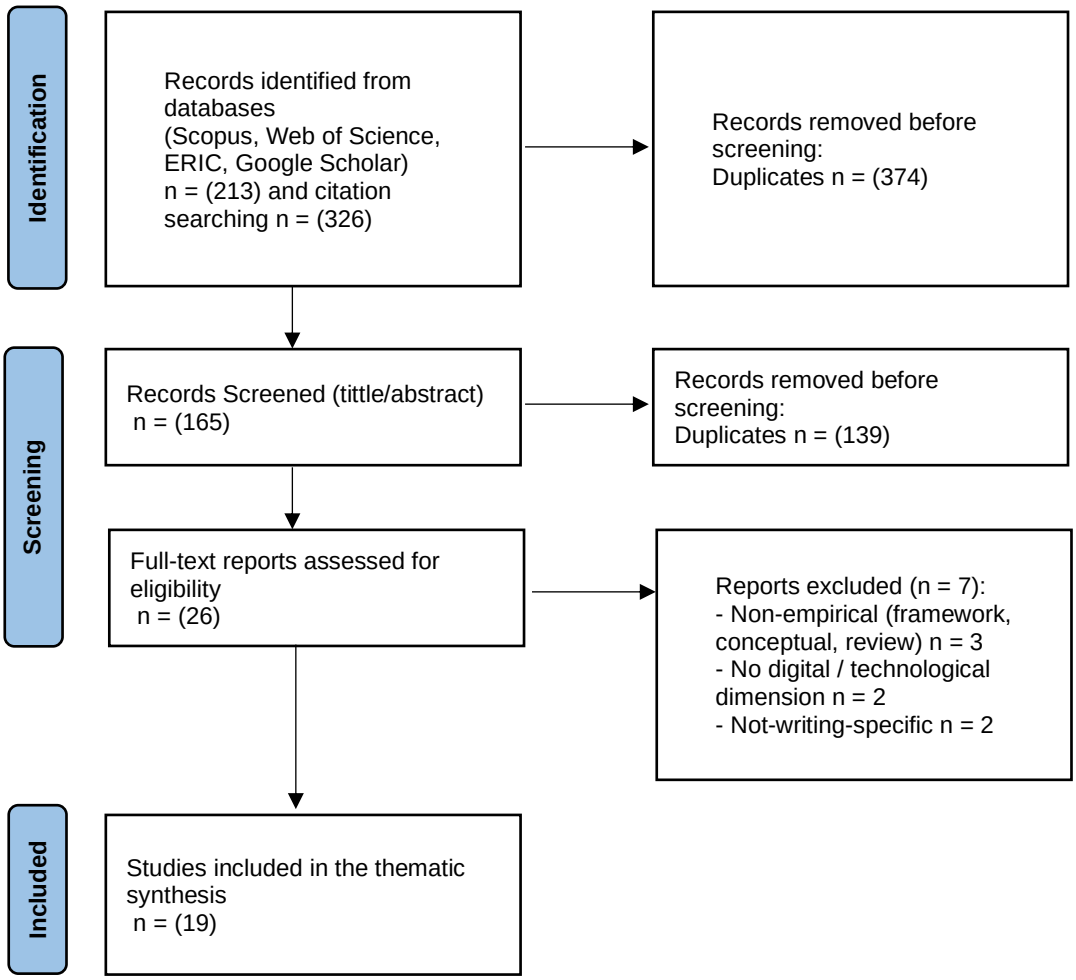
Eligibility Criteria and Study Selection

Eligibility criteria were defined along six dimensions (Table 1). In brief, studies were included if they reported primary empirical data from lecturers, instructors, or faculty who assess English writing in higher education, and if the assessment involved a digital or technological element. Conceptual papers and reviews were excluded from the synthesis but retained, where relevant, as interpretive literature for the discussion.

Table 1. Eligibility Criteria for Study Inclusion

Dimension	Inclusion criteria	Exclusion criteria
Population	Lecturers, instructors, or faculty who assess English (EFL/ESL/EAP/English Studies) writing	Students only; pre-service teachers without assessment roles
Context	Higher education	Primary and secondary schooling; commercial testing without teacher judgment
Phenomenon	Assessment of, or feedback on, written products involving a digital or technological element (GenAI, AWE, AI feedback or scoring, AI or plagiarism detection, LMS-based or online assessment, digital multimodal composing)	Writing assessment with no digital dimension; technology use in instruction without an assessment component
Study type	Primary empirical studies (qualitative, quantitative, or mixed methods)	Conceptual papers, frameworks, editorials, and literature reviews
Language	Full text available in English	Full text unavailable or not in English
Period	January 2014 to September 2026	Published before 2014

Records were screened first on title and abstract and then at full text. Twenty-six full-text reports were assessed for eligibility. Seven were excluded: three were non-empirical (a framework paper by Perkins et al., 2024; a conceptual paper by Warschauer et al., 2023; and a narrative review by Singh et al., 2026); two had no digital or technological dimension (Wei & Cao, 2020; Le & Pham, 2025b); one was set in K-12 schooling (Shao, 2023); and one addressed online EFL assessment in general without a specific focus on writing (Alizadeh Mahmoud Alilo, 2026). Nineteen studies were included in the synthesis. Picture 1 summarises the selection process.



Picture 1. PRISMA 2020 Flow Diagram of Study Identification, Screening, and Inclusion

Picture 1 shows that the main losses at the full-text stage were conceptual rather than empirical. This matters for interpretation: much of the influential commentary on GenAI and assessment is argumentative, and the empirical base on lecturers' actual assessment experience is considerably smaller than the volume of discussion suggests.

Data Extraction and Quality Appraisal

A structured extraction form recorded bibliographic details; country and institutional context; participants and sample size; design and data-collection methods; the digital technology involved; findings relevant to each research question; reported outcomes and limitations of strategies; and illustrative participant quotations with page references. Methodological quality was appraised with the Mixed Methods Appraisal Tool (Hong et al., 2018), which accommodates qualitative, quantitative, and mixed-methods designs within one instrument. Each study received an overall rating of high, moderate, or low (Table 2). Appraisal was used to weight interpretations, not to exclude studies, because small qualitative studies are often the only source of fine-grained evidence on assessor cognition.

Data Analysis

Analysis followed the three stages of thematic synthesis described by Thomas and Harden (2008) and the reflexive principles of Braun and Clarke (2022). First, the results and discussion sections of each study were coded line by line. Second, codes were grouped into descriptive themes that stayed close to the primary data. Third, analytical themes were developed to answer the research questions. Coding combined deductive and inductive logic. For RQ2, five a priori source categories (technological, pedagogical, ethical, institutional, and student-related) were applied, and each category was coded as dominant, present, or absent in each study. For RQ3, strategies were sorted into four categories (cognitive-reflective, assessment and verification, communicative, and institutional or professional) and coded as enacted when participants reported doing them, or as recommended when they appeared only as proposals by participants or authors. RQ1 themes were derived inductively. Frequencies are reported descriptively to show the breadth of evidence; they do not indicate effect size or prevalence in the wider population. Coding decisions were checked against the primary text, and negative or disconfirming cases were noted explicitly.

RESULTS AND DISCUSSION

Results

Characteristics of the Included Studies

Table 2 summarises the 19 included studies. They were conducted in 13 countries, with Vietnam contributing the largest share (five studies), followed by China and the United States (two each); 14 of the 19 studies were set in Asia, including West and Central Asia. Fourteen studies were qualitative and five used mixed methods. Samples were small: excluding the survey phase of Hang (2025), the median number of lecturer participants was nine, and five studies drew on one to three participants. Most participants taught EFL or EAP writing, but four studies included broader populations, namely multidisciplinary faculty (Dishari, 2026; Sinha et al., 2026), English Studies lecturers (Andrew, 2025), and teachers across disciplines (Adnin et al., 2025). Their findings were weighted accordingly.

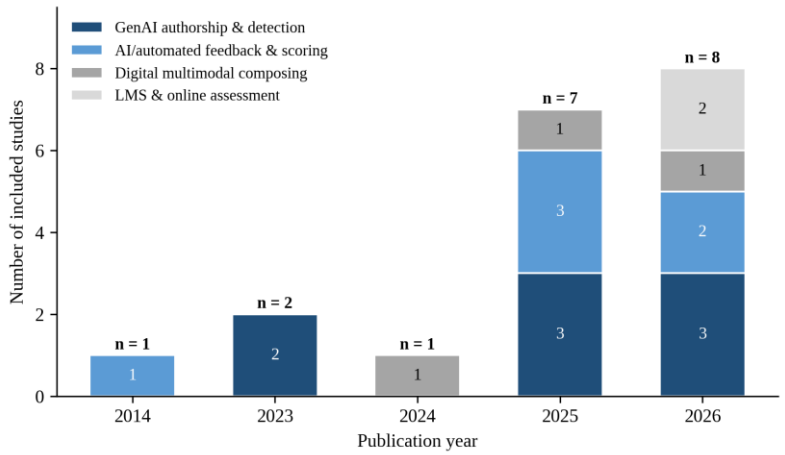
Table 2. Characteristics of the Included Studies (N = 19)

No.	Study	Country	Technology focus	Design and data	Participants	MMAT appraisal
1	Adnin et al. (2025)	USA	GenAI authorship & detection	Mixed methods: survey, interviews, scenarios	9 HE teachers (+97 students surveyed, 15 interviewed)	High
2	Alexander et al. (2023)	Cyprus	GenAI authorship & detection	Qualitative task-based study; detector comparison	6 ESL lecturers	Moderate
3	Anbreen et al. (2026)	Pakistan	AI/automated feedback & scoring	Mixed methods: score comparison, stimulated recall, interviews	2 teachers; 50 IELTS essays	Moderate
4	Andrew (2025)	UK	Digital multimodal composing	Multi-method qualitative: interviews, documents	15 English Studies lecturers	High
5	Demirkol Orak (2025)	Turkey	AI/automated feedback & scoring	Qualitative: open-ended questionnaire	36 EFL instructors	Moderate
6	Dishari (2026)	Lebanon/Gulf	GenAI authorship & detection	Qualitative: semi-structured interviews	20 faculty (3 in English)	High
7	Hang (2025)	Vietnam	AI/automated feedback & scoring	Explanatory sequential mixed methods	233 EFL teachers surveyed; 15 interviewed	High
8	Le & Pham (2025a)	Vietnam	AI/automated feedback & scoring	Duoethnography	2 EFL writing lecturers	Moderate

No.	Study	Country	Technology focus	Design and data	Participants	MMAT appraisal
9	Link et al. (2014)	USA	AI/automated feedback & scoring	Longitudinal qualitative case study	5 ESL writing instructors	Moderate
10	Mirzayev et al. (2026)	Uzbekistan	GenAI authorship & detection	Qualitative: structured interviews	8 EFL teachers (+14 students)	High
11	Mudau et al. (2026)	South Africa	LMS & online assessment	Qualitative case study: interviews	11 academic writing lecturers	Moderate
12	Nguyen (2023)	Vietnam	GenAI authorship & detection	Mixed methods: questionnaire, interviews	20 EFL writing teachers	Moderate
13	Pham et al. (2026)	Vietnam	Digital multimodal composing	Multiple-case study: interviews, metaphor sketches	3 EFL writing lecturers	Moderate
14	Pham & Nguyen (2026)	Vietnam	LMS & online assessment	Qualitative case study: interviews	12 EFL teachers	High
15	Rohmah & Rahayu (2025)	Indonesia	GenAI authorship & detection	Case study: interviews, document analysis	2 EFL writing lecturers	Moderate
16	Shaikenova (2026)	Kazakhstan	GenAI authorship & detection	Mixed methods: questionnaire, interviews	15 EAP/academic writing teachers	Low
17	Sinha et al. (2026)	India	AI/automated feedback & scoring	Interpretive case study: interviews (preprint)	15 faculty assessing writing	Moderate
18	Zhang (2025)	China	GenAI authorship & detection	Qualitative: semi-structured interviews	9 EFL writing teachers	High
19	Zuo & He (2024)	China	Digital multimodal composing	Instrumental case study	1 instructor (+17 students)	High

Note. MMAT = Mixed Methods Appraisal Tool (Hong et al., 2018). Le & Pham (2025a) = the duoethnographic study published in the International Journal of Computer-Assisted Language Learning and Teaching.

Graph 1 traces publication over time. Apart from a single AWE study in 2014, the corpus is concentrated after the public release of ChatGPT: three studies appeared in 2023–2024, seven in 2025, and eight in 2026 up to the search date. The thematic profile also shifted. Early studies addressed automated scoring or multimodal genres, while from 2025 onward GenAI authorship and detection (eight studies overall) and AI-generated feedback and scoring (six studies overall) dominate. The trend indicates that the problem has moved from whether machines can score writing to whether lecturers can still tell whose writing they are scoring.



Graph 1. Publication Trend of Included Studies by Year and Technology Focus (N = 19)

RQ1: How Lecturers Experience and Manage Cognitive Conflict

Four analytical themes describe the experience of conflict: conflicting professional judgments, epistemic uncertainty about authorship, affective and identity strain, and characteristic patterns of conflict management.

The most direct form of conflict occurred when a lecturer's holistic reading of a text diverged from a machine output. In the earliest study, an instructor found it "really difficult to tell [my student] why I gave him a C-, but he already got a 5 from Criterion" (Link et al., 2014, p. 337); students sometimes regarded the AWE score as more objective than the teacher. The same structure reappears with GenAI. In a duoethnographic account, ChatGPT scored a draft at 68 and flagged surface errors, while the lecturer read an emerging personal voice the tool could not register: "The system reads correctness … but not courage" (Le & Pham, 2025a, p. 9). Teachers in China found AI scores from different platforms "confusing and contradictory" (Zhang, 2025), and Indian faculty described a "pedagogical dissonance" between "the teacher's eye" and algorithmic scoring, compounded by the inability to explain scores they did not understand: "I cannot defend a score I do not understand" (Sinha et al., 2026, p. 6). Anbreen et al. (2026) offer quantitative support for this experience: human and ChatGPT-5 scores on 50 IELTS essays correlated moderately ($\rho = .685$) with similar means, but the GenAI scores clustered within a narrow band (SD = 0.58 versus 1.44 for the human rater) and overlooked coherence and elaboration problems. Agreement at the level of averages thus concealed disagreement at the level of individual judgments, which is precisely where lecturers experience conflict.

A second form of conflict concerned the provenance of the text. Kazakhstani EAP teachers encountered essays that read as "advanced academic work" whose authors "could not provide an adequate justification for the organization of argumentation" (Shaikenova, 2026, p. 579). Uzbek EFL teachers graded under what Mirzayev et al. (2026) call profound uncertainty, relying on informal cues such as sudden improvement, error-free prose, missing personal voice, or failure to use course readings. All eight teachers acknowledged that strong genuine writing could be misjudged as AI-generated, which produces a paradox specific to second-language contexts: good L2 writing itself triggers suspicion. Alexander et al. (2023) showed how fragile such cues are. Lecturers applied a deficit model in which errors signalled human authorship and polish signalled AI, and they consequently judged a fully AI-written essay as the work of a learner "still in the process of mastering English" (p. 35). In the United States, teachers felt confident recognising AI "patterns" but would not act on them: "on my part, it's a guess. I don't want to accuse them of something [if] they didn't do it" (Adnin et al., 2025, Section 6.1).

Conflict was rarely only cognitive. Dishari (2026) documented an explicit dissonance between "I value student trust" and "I think they're cheating," resolved for some by disillusionment and for others by stricter rules. Grading became "a constant investigation … It's draining" (p. 275), and lecturers described guilt, "ethical fatigue," and complicity when they used AI for their own work while penalising students for doing the same. An Uzbek teacher admitted, "Sometimes I feel guilty about the fact that I may be biased" (Mirzayev et al., 2026, p. 13). Turkish instructors reported feeling "inadequate" before near-perfect machine output and warned that passing AI feedback to students unfiltered "undermin[es] your role as a teacher" (Demirkol Orak, 2025, p. 84). Identity strain also appeared in the gap between what lecturers were asked to teach and what they were required to examine. A Vietnamese lecturer found it "illogical to teach digital skills but have paper-based final exams … It feels irresponsible" (Pham et al., 2026, p. 12), and Hang (2025) described a state of "aware incompetence" in which teachers recognised ethical risks but lacked the pedagogical and technical readiness to act on them.

Across studies, lecturers' responses followed three recognisable patterns. The first was investigative: comparing the suspect text with earlier work, re-prompting ChatGPT with the assignment question, or questioning the student orally (Adnin et al., 2025; Mirzayev et al., 2026). The second was accommodative: lowering the stakes of the judgment through leniency, mutual trust, or a decision not to pursue the matter. One Vietnamese teacher saw "no point in trying to dig deep and find the truth" (Pham & Nguyen, 2026), and a Chinese teacher who found that detectors labelled his own pre-AI papers as 30–70% AI-generated abandoned detection in favour of trust (Zhang, 2025). The third was structural: redesigning the assessment so that the conflict would not arise again, for instance by returning to supervised or paper-based writing (Pham et al., 2026) or reweighting grades toward closed-book examinations (Zhang, 2025). Responses also differed by experience. Senior teachers in Zhang's (2025)

study felt able to "catch the mole," whereas younger colleagues did not, and Link et al. (2014) found that instructors' confidence with AWE grew over two semesters. Practice nevertheless remained, in Mirzayev et al.'s (2026) phrase, informal and individualistic: "Right now, everyone seems to be figuring it out by themselves" (p. 13).

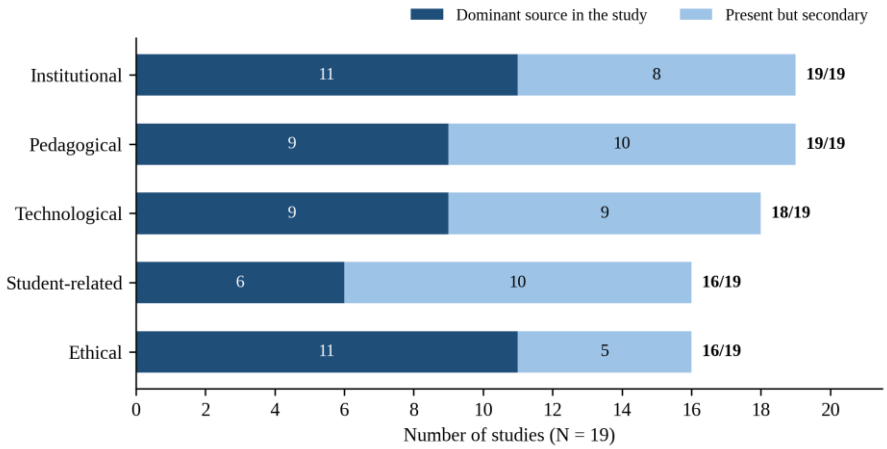
RQ2: Sources of Cognitive Conflict

Table 3 maps the five source categories across the 19 studies, and Graph 2 summarises their frequency.

Table 3. Sources of Cognitive Conflict Reported in the Included Studies

Study	Technological	Pedagogical	Ethical	Institutional	Student-related
Adnin et al. (2025)	●	○	●	○	●
Alexander et al. (2023)	●	●	○	○	–
Anbreen et al. (2026)	●	○	●	●	○
Andrew (2025)	○	●	–	●	○
Demirkol Orak (2025)	●	○	○	○	●
Dishari (2026)	○	●	●	●	○
Hang (2025)	○	●	●	●	–
Le & Pham (2025a)	●	○	○	●	○
Link et al. (2014)	●	○	–	○	●
Mirzayev et al. (2026)	○	○	●	●	○
Mudau et al. (2026)	○	●	●	●	○
Nguyen (2023)	○	○	●	○	●
Pham et al. (2026)	○	●	○	●	○
Pham & Nguyen (2026)	●	○	●	○	●
Rohmah & Rahayu (2025)	○	○	●	●	●
Shaikenova (2026)	○	●	●	○	○
Sinha et al. (2026)	●	○	●	●	○
Zhang (2025)	●	●	○	○	○
Zuo & He (2024)	–	●	–	●	–
Reported (any)	**18**	**19**	**16**	**19**	**16**
Dominant	**9**	**9**	**11**	**11**	**6**

Note. ● = dominant source in the study; ○ = present but secondary; – = not reported.



Graph 2. Frequency of Cognitive Conflict Sources Across Included Studies (N = 19)

Two patterns stand out. First, conflict was almost always multi-sourced: 14 of the 19 studies reported all five categories and four reported four; only Zuo and He (2024), a single-instructor case, reported two. Second, breadth and salience did not coincide. Institutional and pedagogical sources were present in every study, but ethical sources were the most often dominant (11 studies, tied with institutional sources), whereas student-related sources, although common, were dominant in only six. In other words, lecturers rarely located the origin of conflict in students themselves; they located it in the rules, tools, and ethical stakes surrounding the student.

Technological sources were reported in 18 studies and dominant in nine. Three mechanisms recurred. Detection tools were unavailable, unaffordable, or unreliable and could not support a defensible decision (Alexander et al., 2023; Zhang, 2025; Mirzayev et al., 2026). Automated scores and feedback were opaque or inconsistent with human readings (Link et al., 2014; Anbreen et al., 2026; Sinha et al., 2026). And GenAI produced fluent text that blurred the line between assistance and outsourcing, so that "polished, empty prose" (Shaikenova, 2026, p. 579) could pass surface checks. In COVID-era studies, connectivity failures and the limits of camera proctoring played a similar role (Pham & Nguyen, 2026).

Pedagogical sources centered on construct drift, meaning uncertainty about what the assessment now measured. Lecturers asked, in effect, "What am I even assessing now?" (Dishari, 2026). In DMC contexts the question was whether a project tested digital design or writing (Pham et al., 2026; Andrew, 2025), and Andrew (2025) found that lecturers who declared they marked "substance over style" could not in practice separate design quality from the persuasiveness of the argument. Product-based grading conventions sat uneasily with process-oriented beliefs about writing development (Mudau et al., 2026), and existing rubrics were judged inadequate for AI-assisted texts (Zhang, 2025; Alexander et al., 2023).

Ethical sources concerned academic integrity, fairness, and the risk of wrongful accusation. Lecturers struggled to distinguish "legitimate assistance and unacceptable assistance" (Anbreen et al., 2026, p. 25) and worried about fairness to students who did not use AI (Nguyen, 2023). Equity concerns extended to unequal access to paid tools (Zhang, 2025), bias against non-standard Indian English and data-privacy risks (Sinha et al., 2026), and the ethics of surveillance itself: "if the solutions were to increase surveillance and loss of privacy, then I definitely don't want such solutions" (Pham & Nguyen, 2026, p. 10).

The most consistent institutional source was a policy vacuum. Uzbek teachers worked with no policy, no shared protocol, and no tools (Mirzayev et al., 2026); Indian faculty observed that one colleague encouraged AI for brainstorming while the next banned it entirely (Sinha et al., 2026); and Indonesian lecturers found that AI regulation was happening "at the classroom level rather than through formal policy" (Rohmah & Rahayu, 2025, p. 543). Hang (2025) reported that 76.4% of 233 Vietnamese teachers had received no training in AI-generated feedback, a condition the author called "institutional abandonment." Where policy did exist, it often pulled against lecturers' practice, as with high-stakes print-based examinations that marginalised multimodal work (Zuo & He, 2024; Pham et al., 2026). Large classes and heavy marking loads further narrowed the space for careful judgment (Mudau et al., 2026).

Student-related sources included undisclosed AI use and evasion of detection (Adnin et al., 2025), uncritical acceptance of AI feedback as "absolute truth" (Demirkol Orak, 2025), over-reliance on tools (Nguyen, 2023), and misunderstanding of permitted use. The latter is illustrated by a student who submitted a fully AI-generated essay "because you told us we can use it" (Rohmah & Rahayu, 2025, p. 542). This example shows that student-related conflict often originated in ambiguous rules rather than in deliberate dishonesty, which links it back to the institutional category.

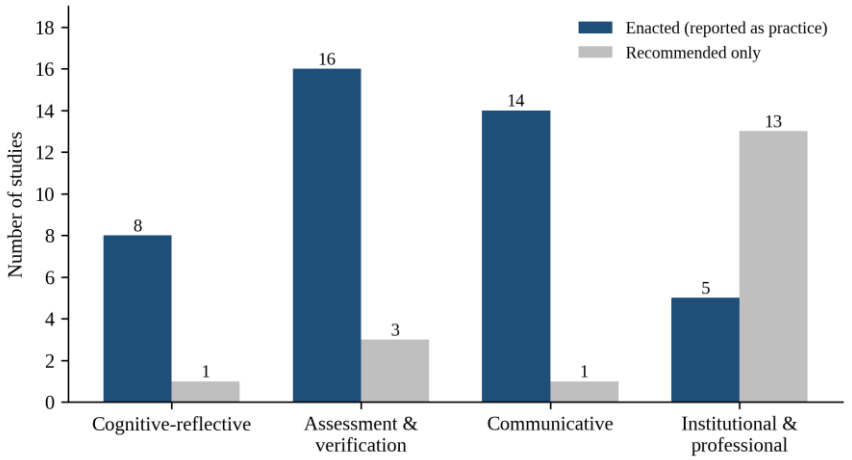
RQ3: Strategies for Resolving or Managing Cognitive Conflict

Table 4 summarises the strategies identified, and Graph 3 contrasts strategies that lecturers reported enacting with those that were only recommended.

Table 4. Strategies Used to Resolve or Manage Cognitive Conflict

Category	Illustrative practices	Representative studies	Reported outcomes and limitations
Cognitive-reflective	Treating AI scores as advisory; filtering AI feedback through own criteria; dialogic reflection with a colleague; shifting from policing to mentoring	Zhang (2025); Demirkol Orak (2025); Le & Pham (2025a); Anbreen et al. (2026)	Restored confidence in own judgment; time-intensive; depends on individual disposition

Category	Illustrative practices	Representative studies	Reported outcomes and limitations
Assessment and verification	Staged drafts and outlines; version histories and prompt logs; comparison with in-class writing; brief oral explanation; integrated or source-based tasks; reflective commentaries; adapted rubrics	Adnin et al. (2025); Mirzayev et al. (2026); Shaikenova (2026); Pham & Nguyen (2026); Andrew (2025); Zuo & He (2024)	Most credible evidence of authorship; increased workload; end-of-term comparisons cannot quantify AI use
Communicative	Stating AI-use rules and acknowledgement requirements; talking with students before grading; using flawed AI feedback as a teaching point; asking students to justify accepted AI suggestions	Rohmah & Rahayu (2025); Mirzayev et al. (2026); Link et al. (2014); Hang (2025)	Clarified expectations; occasional reversal of unfair marks; rules alone did not prevent misuse
Institutional and professional	Colleague mentoring; division-level procedures; joint invigilation; peer learning; (recommended) AI policy, moderation, training, psychological support	Andrew (2025); Pham & Nguyen (2026); Hang (2025); Dishari (2026); Sinha et al. (2026)	Enacted in 5 studies, recommended in 13; largely untested



Graph 3. Strategies for Managing Cognitive Conflict: Enacted Versus Recommended Only (N = 19)

Cognitive-reflective strategies (enacted in eight studies) involved reappraising one's own judgment and the status of machine output. Lecturers treated AI scores as "just for reference. I have my judgement" (Zhang, 2025, p. 10), filtered AI feedback through their own criteria (Demirkol Orak, 2025), and, in the most developed example, used dialogic reflection with a colleague to "dwell in dissonance" before reclaiming their judgment (Le & Pham, 2025a). Anbreen et al. (2026) described a reflective shift from policing students to mentoring them in ethical AI use.

Assessment and verification strategies were the most widely enacted (16 studies). They fall into two groups. Evidential strategies gathered additional evidence about authorship: staged drafts and outlines (Nguyen, 2023; Shaikenova, 2026), version histories and prompt logs (Adnin et al., 2025), comparison with earlier work or in-class writing (Mirzayev et al., 2026; Zhang, 2025), and short oral explanations or defences (Shaikenova, 2026; Mirzayev et al., 2026). Design strategies changed the task so that the evidence problem was reduced: integrated and source-based writing (Pham & Nguyen, 2026), accompanying reflective essays for multimodal projects (Andrew, 2025), rubrics adapted from examination criteria (Zuo & He, 2024), and permitted AI use limited to brainstorming and outlining with explicit marking-up of AI material (Rohmah & Rahayu, 2025). Notably, several lecturers rejected detection as a basis for decisions. One used a detector only so that "it's not just my opinion, but also verified by the tool" (Adnin et al., 2025, Section 6.1), while others re-checked flagged essays case by case rather than accepting detector output (Pham & Nguyen, 2026).

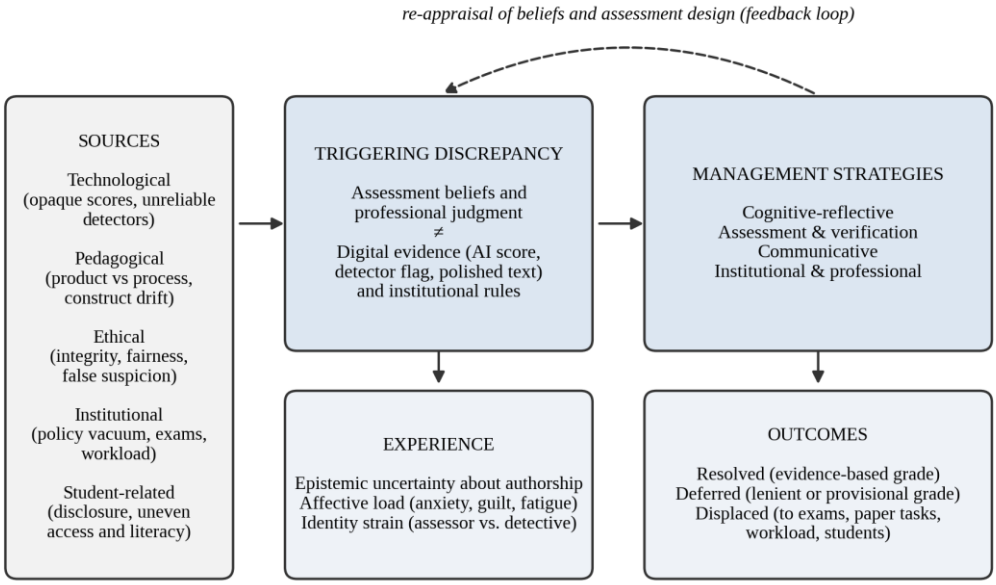
Communicative strategies (enacted in 14 studies) made expectations and judgments negotiable. They included setting AI-use rules at the start of a course and requiring acknowledgement of ChatGPT (Rohmah & Rahayu, 2025), talking with a suspected student before deciding on a mark (Mirzayev et al., 2026), turning inaccurate AI feedback into lessons in critical evaluation (Link et al., 2014; Demirkol Orak, 2025), and asking students to justify each AI suggestion they accepted or rejected (Hang, 2025). Communication also served as a check on the lecturer: in Mirzayev et al. (2026), one teacher reversed a reduced mark after a convincing oral explanation.

Institutional and professional strategies show the starkest asymmetry. They were recommended in 13 studies but enacted in only five, and the enacted cases were modest: colleague mentoring and "parallel" essay-plus-multimodal assessment to secure committee approval (Andrew, 2025), division-level procedures and jointly invigilated examinations (Pham & Nguyen, 2026), peer learning among teachers (Hang, 2025), consultation with an educational technology specialist (Zhang, 2025), and tool training or self-exploration (Link et al., 2014). The recommended agenda was far broader, covering clear and educative AI policies co-developed with staff and students, assessment moderation, AI and assessment literacy training, and psychological support (Mirzayev et al., 2026; Dishari, 2026; Sinha et al., 2026).

Outcomes and limitations of strategies. Few studies evaluated outcomes, and those that did reported partial success at a cost. Verification improved lecturers' confidence that students understood their work (Adnin et al., 2025) and allowed teachers to "track intellectual growth" (Shaikenova, 2026), but it increased workload and in some cases reduced motivation (Pham & Nguyen, 2026). Comparing assignments with examinations only worked at the end of term and could not show the extent of AI use (Zhang, 2025). Accompanying essays did not remove the difficulty of marking multimodal work (Andrew, 2025), and permission-based AI policies still produced over-reliance (Rohmah & Rahayu, 2025). In many cases the conflict was not resolved but deferred, through a lenient or provisional grade, or displaced, onto examinations, paper-based tasks, students, or lecturers' own time.

Discussion

Taken together, the findings suggest that cognitive conflict in digital writing assessment follows a recognisable sequence, summarised in Picture 2. Multiple sources converge on a triggering discrepancy between the lecturer's beliefs and judgment and the digital evidence or institutional rule in front of them. The discrepancy is experienced simultaneously as epistemic uncertainty, affective load, and identity strain. Lecturers then draw on reflective, verification, communicative, or institutional strategies, and the outcome may be a resolved, deferred, or displaced conflict. Resolved conflicts can feed back into revised beliefs and assessment designs.



Picture 2. Source–Discrepancy–Management–Outcome Model of Cognitive Conflict in Digital Writing Assessment

The model makes two points that individual studies do not. First, the epistemic and the affective components are entangled. Lecturers felt guilt because they were uncertain (Mirzayev et al., 2026), and fatigue because verification was the only available response to uncertainty (Dishari, 2026). Second, the frequency of deferred and displaced outcomes indicates that much observable assessment practice is dissonance reduction rather than conflict resolution in the epistemic sense. When a teacher decides there is "no point in trying to dig deep" (Pham & Nguyen, 2026) or adopts mutual trust after detectors fail (Zhang, 2025), the discomfort is reduced, as Festinger (1957) would predict, but the question of authorship remains open. This distinction matters for validity, because the grade is recorded as if the question had been answered.

The review offers little support for the view that digital tools simply reduce assessment uncertainty. They appear to do both: they reduce some uncertainty while creating new forms of it. AWE and AI feedback reduced uncertainty about surface accuracy and saved time (Demirkol Orak, 2025; Link et al., 2014), yet they introduced a second assessor whose reasoning was hidden and whose output students sometimes trusted more than the lecturer. Detection tools promised to answer the authorship question but in practice relocated it, because their output itself required interpretation and was known to be biased against L2 writers (Liang et al., 2023; Weber-Wulff et al., 2023). It is therefore necessary to distinguish between what a tool can do and how lecturers interpret what it produces. Much of the conflict documented here arose not from technical failure alone but from the absence of an interpretive framework for machine evidence. Where lecturers positioned that evidence as advisory, as one input to be triangulated (Zhang, 2025; Demirkol Orak, 2025), they described regaining control of the decision; where students or institutions expected it to be decisive, conflict sharpened (Link et al., 2014; Sinha et al., 2026).

Three theoretical implications follow. First, the review supports keeping cognitive conflict and cognitive dissonance analytically distinct. Conflict describes the evidential problem; dissonance describes what happens when lecturers act against their own assessment beliefs to manage that problem, for example by penalising suspected AI use without proof or by grading leniently despite doubts. Dishari's (2026) use of dissonance theory captures the second process well, while studies of detection and scoring (Alexander et al., 2023; Anbreen et al., 2026) document the first. Second, the findings extend teacher cognition research (Borg, 2015) by showing that contextual mediation now includes algorithmic actors. The belief–practice gaps Wei and Cao (2020) attributed to students, colleagues, and culture are joined by gaps produced by detectors, scoring engines, and platform logics. Third, the findings support the socio-emotional dimension of assessment literacy proposed by Pastore and Andrade (2019) and suggest a further component: the capacity to interpret, question, and explain algorithmic evidence. Sinha et al.'s (2026) participant who "cannot defend a score I do not understand" illustrates that explainability is now an assessment literacy issue as well as a technical one.

The findings agree with conceptual work warning that suspicion falls hardest on capable L2 writers (Warschauer et al., 2023) and that detection-first approaches are the least defensible option when misclassification risk is high (Singh et al., 2026). They also agree with Perkins et al.'s (2024) argument for graded and explicit AI-use levels: lecturers valued rules stated at the start of a course, although rules alone did not prevent misuse (Rohmah & Rahayu, 2025). The review departs from earlier AWE research in one respect. Link et al. (2014) found that instructor satisfaction rose and conflict eased with experience, whereas GenAI-era studies report no comparable learning curve, probably because the authorship question cannot be settled by familiarity with the tool. Emergency online assessment research (Shao, 2023; Alizadeh Mahmoud Alilo, 2026) anticipated the identity struggles found here, but GenAI moves the source of doubt from the testing environment to the text itself, which makes it harder to remove through proctoring or scheduling. Finally, studies of assessment practice without a digital element, such as Wei and Cao (2020) and Le and Pham (2025b), show that belief–practice compromises long predate AI. Digital tools did not create cognitive conflict in writing assessment, but they have multiplied its sources and made it more visible.

For assessment practice, the evidence favours designs that make the writing process visible over attempts to police the product. Staged submissions, brief reflective commentaries, version histories where appropriate, and short oral explanations generated the most credible evidence in the reviewed studies, and they do so without relying on detectors. Analytic rubrics that name acceptable AI use and reward disclosure and critical evaluation of AI output (Rohmah & Rahayu, 2025; Singh et al., 2026) can turn an integrity dispute into an assessable learning outcome. Because verification is labour-

intensive, it should be applied proportionately: built into a few high-stakes tasks rather than added to every assignment. For professional development, lecturers need more than tool training. The reviewed studies point to four capacities: interpreting algorithmic evidence and its error profiles; holding difficult conversations about authorship without accusation; calibrating judgments with colleagues through moderation; and recognising the emotional costs of suspicion-based assessment. Institutions, for their part, need to close the gap shown in Graph 3 by turning recommended measures into routine structures: written AI-use policies co-developed with staff and students, moderation panels for contested cases, and workload allocation that recognises verification as part of assessment work.

The review also exposes gaps. Only a few studies examined cognitive conflict as their explicit object; most addressed it indirectly through perceptions or challenges. Longitudinal designs were almost absent after 2014, so it is unknown whether GenAI-era conflict diminishes as lecturers gain experience. Most studies relied on self-report, and only Anbreen et al. (2026) and Alexander et al. (2023) observed lecturers making actual judgments on texts. Southeast and Central Asian contexts were comparatively well represented, yet Indonesia, for example, was represented by a single two-lecturer case study, and no study used a comparative cross-national design. No study tested whether a particular strategy improved grading consistency or fairness. Future research could use think-aloud or stimulated-recall protocols during real grading, follow lecturers across semesters as institutional policies change, compare institutions with and without moderation procedures, and evaluate interventions such as disclosure rubrics or oral-defence protocols.

The review has limitations of its own. The corpus is restricted to English-language publications and to the databases searched, and it includes one preprint (Sinha et al., 2026), one doctoral thesis (Andrew, 2025), and one conference paper (Shaikenova, 2026), which have not all undergone conventional peer review. Many included studies had very small samples, and four involved participants beyond English writing lecturers. Frequency counts were based on reported findings and may reflect what authors chose to emphasise. Finally, the classification of strategies as enacted or recommended required interpretive judgment where studies blurred the two.

CONCLUSION

This review synthesised 19 empirical studies on how English writing lecturers experience, explain, and manage cognitive conflict when assessing student writing in digital environments. In answer to RQ1, lecturers experience conflict as a divergence between their holistic judgment and machine output, as epistemic uncertainty about authorship, and as an affective and identity burden; they manage it through investigative, accommodative, and structural responses that remain largely individual. In answer to RQ2, conflict is multi-sourced: institutional and pedagogical sources were present in every study, technological sources in nearly all, and ethical concerns were the most often dominant, with policy vacuums and unreliable detection acting as the most decisive triggers. In answer to RQ3, lecturers rely on verification and communicative strategies, while institutional and professional responses are mostly proposed rather than implemented, so that many conflicts are deferred or displaced rather than resolved.

The central implication is that fair assessment of digital-age writing cannot rest on lecturers' individual vigilance or on detection technology. It requires process-visible task design, explicit and educative AI-use policies, collegial moderation of contested judgments, and an assessment literacy that includes the interpretation of algorithmic evidence. Future empirical work should examine conflict episodes as they occur during grading and test whether the strategies identified here improve the consistency, fairness, and credibility of lecturers' decisions.

REFERENCES

Adnin, R., Pandkar, A., Yao, B., Wang, D., & Das, M. (2025). Examining student and teacher perspectives on undisclosed use of generative AI in academic work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713393

Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology, 23*(2), 25–43. https://doi.org/10.56297/BUKA4060/XHLD5365

Alilo, H.A.M. (2026). Evaluating assessment strategies for EFL teacher pre-, during-, and post-COVID-

19: Teachers' reflections on challenges and opportunities. *Language, Identity and the Digital Realm, 2*(1), 23–48. https://doi.org/10.22034/lidr.2025.2072655.1020

Anbreen, T., Özturan, T., Shrestha, P. N., & Maqsood, A. (2026). Human vs. generative artificial intelligence in writing assessment: Investigating feedback alignment, score validity, and teacher agency. *Educational Assessment, Evaluation and Accountability*. Advance online publication. https://doi.org/10.1007/s11092-026-09486-z

Andrew, M. (2025). *Exploring digital multimodal approaches to teaching, learning, and assessment through the perspectives of English lecturers in higher education* [Doctoral thesis, University of Leicester].

Borg, S. (2015). *Teacher cognition and language education: Research and practice*. Bloomsbury Academic.

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Demirkol Orak, S. (2025). Turkish EFL teachers' perspectives on AI-generated feedback: Negotiating trust, control, and pedagogical adaptation in writing instruction. *Applied Linguistics: Research, Measurement and Practice, 1*(1), 74–94. https://doi.org/10.65334/n40kdg46

Dishari, S. (2026). Teaching in the age of generative AI: A qualitative analysis of faculty identity, emotion, and assessment concerns. *Journal of Artificial Intelligence and Technology, 6*, 270–283. https://doi.org/10.37965/jait.2025.0941

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

hyGough, D., Oliver, S., & Thomas, J. (Eds.). (2017). *An introduction to systematic reviews* (2nd ed.). SAGE.

Hang, N. T. (2025). Vietnamese EFL teachers' readiness and challenges in integrating AI-generated feedback. *Language, Technology, and Social Media, 4*(1), 89–103. https://doi.org/10.70211/ltsm.3026-7196.315

Harmon-Jones, E. (Ed.). (2019). *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed.). American Psychological Association. https://doi.org/10.1037/0000135-000

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Knoch, U., Fairbairn, J., & Jin, Y. (2021). *Scoring second language spoken and written performance: Issues, options and directions*. Equinox.

Le, T. T., & Pham, T. T. (2025a). A duoethnographic study of Writing-With as a mode of EFL teacher professional development practice in the age of AI. *International Journal of Computer-Assisted Language Learning and Teaching, 15*(1), 1–16. https://doi.org/10.4018/IJCALLT.388557

Le, T. T., & Pham, T. T. (2025b). Writing-With assessment: Dialogic reflection as a bridge between teacher feedback and student learning. *Language Testing in Asia, 15*, Article 53. https://doi.org/10.1186/s40468-025-00399-0

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards best ESL practices for implementing automated writing evaluation. *CALICO Journal, 31*(3), 323–344. https://doi.org/10.11139/cj.31.3.323-344

Mirzayev, A., Abdullah Kamal, S. S. L., Turdiyeva, G., Tojiboyeva, M., & Irgasheva, N. (2026). AI detection and grading fairness in EFL assessment: A qualitative study of student and teacher perceptions at a university in Uzbekistan. *Cogent Education, 13*(1), Article 2742806. https://doi.org/10.1080/2331186X.2026.2742806

Mudau, P. P., Cekiso, M., & Mandende, I. P. (2026). Integrating technology into academic writing instruction: Lecturers' practices and perspectives. *Literator, 47*(1), Article a2274. https://doi.org/10.4102/lit.v47i1.2274

Nguyen, T. T. H. (2023). EFL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. *International Journal of Language Instruction, 2*(3), 1–47. https://doi.org/10.54855/ijli.23231

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

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(6). https://doi.org/10.53761/q3azde36

Pham, L.-K.-T., Huynh, N.-T., & Tran, T.-N. (2026). The peripherality of digital multimodal composing in Vietnamese EFL writing: A metaphor-based study of teacher cognition. *Australian Journal of Applied Linguistics, 9*, Article 104271. https://doi.org/10.29140/ajal.2026.104271

Pham, T. T., & Nguyen, H. H. (2026). EFL online assessment under abrupt disruption: Vietnamese tertiary teachers' adaptive strategies for sustainable digital assessment. *Sustainable Education and Digital Transformation, 2*(1), Article e44169. https://doi.org/10.33902/SEDT.202644169

Rohmah, F. N., & Rahayu, S. P. (2025). Ethical AI in EFL writing instruction: A case study of lecturers' communication and pedagogical strategies. *Voices of English Language Education Society, 9*(3), 536–550. https://doi.org/10.29408/veles.v9i3.30723

Shaikenova, D. (2026). Teachers' views on academic writing assignments generated by artificial intelligence: How do they affect the teaching methods in higher education institutions. In *The use of modern digital technologies in linguistics and foreign language teaching: International scientific and practical conference* (pp. 576–583). Zenodo. https://doi.org/10.5281/zenodo.19781196

Shao, S. (2023). Understanding K-12 in-service EFL teachers' online assessment practices and assessor identity: A Chinese study. *SAGE Open, 13*(4). https://doi.org/10.1177/21582440231216617

Singh, R. K. V., Hashim, H., Jamaludin, K. A., & Jamil, M. R. M. (2026). Assessing AI-supported English academic writing in higher education: A narrative review of challenges, emerging practices, and prospective strategies. *Arab World English Journal*, Special Issue on Artificial Intelligence (3), 436–451. https://doi.org/10.24093/awej/A3.28

Sinha, A., Goswami, R., Nag, A., & Nayak, R. (2026). *Assessing teacher readiness and ethical awareness of AI tools for writing assessment: A case from Indian higher education* [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-9698381/v1

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*, Article 45. https://doi.org/10.1186/1471-2288-8-45

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

Wei, W., & Cao, Y. (2020). Written corrective feedback strategies employed by university English lecturers: A teacher cognition perspective. *SAGE Open, 10*(3). https://doi.org/10.1177/2158244020934886

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Zhang, J. (2025). ChatGPT or "CheatGPT"? EFL teachers' perceptions of scoring AI-assisted English writing through the lens of assessment literacy. *Digital Applied Linguistics, 2*, Article 102486. https://doi.org/10.29140/dal.v2.102486

Zuo, Y., & He, F. (2024). Negotiating polycentric power dynamics in China through digital multimodal composing. *TESOL Quarterly, 58*(2), 751–774. https://doi.org/10.1002/tesq.3252