

Approaches and considerations for the assessment of writing: A review of literature from 2013–2026

Maggie Mroczkowski

June 2026

Contents

Acknowledgements	3
Introduction	3
Literature review	3
Approaches to assessment of writing skill.....	4
Approaches to marking.....	4
Changing the focus of the assessment.....	6
Other themes in writing assessment.....	7
Use of automated tools and AI in writing and writing assessment	9
Perceptions and uses of AI tools for writing.....	10
Balancing use of AI with the value of writing.....	11
Detection of AI-generated content in written work	12
Using AI as a grading or feedback tool on written work	15
Examples of AI use in the development and assessment of writing skills.....	16
Conclusion	23
Further reading	25
References	25

Acknowledgements

Thank you to Rachel Whitford for her contributions to the start of this literature review.

Introduction

This literature review was commissioned to explore how writing should be assessed, primarily in English qualifications but also more generally, following from changes made to conditions of assessment in English writing portfolios and against the backdrop of reform. This is particularly pressing given the rise of generative artificial intelligence (AI). Sources reviewed spanned from earlier publications about approaches to writing assessment to recent publications concerned with AI and writing. The literature was primarily sourced from the Knowledge Exchange and via the Taylor & Francis online database.

Literature review

This literature review begins with an overview of some dominant approaches to the assessment of writing skills within the past decade. While the studies reviewed have been wide ranging in their settings, from early primary-age writing assessments to higher education assessments in various subjects, the approaches examined can be grouped into four main categories:

- criterion-based frameworks
- comparative judgement
- content and authorial voice
- scenario-based assessment

It should also be noted that mixed-method assessment approaches may be employed to enhance the validity and reliability of an evaluation of writing ability (Feenstra, 2012).

The literature review then shifts to a focus on how the proliferation of generative AI is being considered and accounted for in the development and assessment of writing skills.

The guiding research questions are:

- What different models, schools of thought or practices exist for the assessment of writing as a skill?
- How are these models being adapted to the advent and rise of generative AI?
- Can best practice be identified?

Approaches to assessment of writing skill

Approaches to marking

Criterion-based frameworks

Analytical frameworks that scored learners' writing across several different domains were frequently cited in the literature.

[Berge et al. \(2019\)](#) created a 'wheel of writing' framework in Norway to assess the quality of writing of children in primary and secondary school education. In this framework, assessment scripts were rated by professional markers on communication, content, text organisation, language usage, orthography, and punctuation with a rating of one to five. The study showed that the framework was effective, but it focused more on improving writing ability, rather than assessment.

[Stuart and Barnett \(2023\)](#) developed a similar tool called the Writing Quality Scale (WQS), but this one was to identify learners with writing difficulties in higher education. In this study, learners were scored on a scale of one to four on the following criteria: content and development; structure and organisation; vocabulary, sentence structure, punctuation (including capitalisation and apostrophes); and spelling. The tool was found to have good inter-rater reliability and was effective at identifying learners with specific learning difficulties.

A comparable tool was evaluated in a study by [Dunsmuir et al. \(2015\)](#), but this time, the framework was specifically designed for writing assessment. In this study, the authors assessed the reliability and validity of the 'Writing Assessment Measure' (WAM) which was developed to reflect the skills that children of different abilities are expected to achieve in written expression in England and Wales. The written expression areas targeted were ideas development, vocabulary, sentence structure, and technical aspects such as grammar, spelling, handwriting and punctuation. Like in the other studies, markers were presented with scoring criteria across these domains. The tool was found to be reliable in terms of internal consistency, accuracy and agreement between markers. However, the consistency for 'ideas development' was lower than the other domains, suggesting that raters interpreted the criteria differently.

All three studies highlighted the benefits of rating learners' writing on different elements in order to give credence to elements such as content and the development of ideas, alongside more technical aspects such as grammar and spelling. However, the studies also emphasised that these types of tools involved a degree of subjective assessment that could create problems with consistency of marking.

Comparative judgement

Assessing learners' writing through comparative judgement came up frequently in the literature. Comparative judgement is an assessment method where scripts are compiled and then ranked in order according to some predefined attributes ([Hopfenbeck, 2019](#)).

[Bramley and Vitello \(2019\)](#) proposed the use of a computerised 'adaptive comparative judgement' tool whereby the scripts shown to judges depend on the results of previous judgements. This approach aims to save time by avoiding comparisons with predictable outcomes, for example, comparing high quality writing with lower quality writing. Instead, judges can focus on comparisons that require more nuance. In this study, 34 examiners of GSCE English were shown 150 autobiographical essays and asked to compare them using the software NoMoreMarking. However, the findings indicated that the reliability of this assessment method compared to non-adaptive comparative judgement was in doubt. Instead, Bramley and Vitello suggested that researchers could explore non-adaptive, fixed comparison designs, where each object is compared with the same reference criteria.

[Van Daal et al. \(2019\)](#) also examined the validity of the comparative judgement approach. In their research, 41 essays from a pre-master's programme in Belgium were selected and presented anonymously to judges. The judges were given a list of principles on which to assess the writing and were then asked to compare two essays and decide which was better. They were also asked to explain their decision. Van Daal et al. examined two assumptions underpinning comparative judgement, namely its holistic characteristic and how the final rank order reflects shared consensus of what makes a good essay. The findings showed that the almost 70% of arguments put forth by judges were directly related to the competence description given to them, indicating strong construct validity. However, judges relied on their own expertise to varying degrees and tended to focus on different aspects of the writing (for example, some focused more on structure and analysis or synthesis, whereas others focused more heavily on problem definition). Additionally, they differed in terms of broadness in their decision-making. Despite these variations, the authors concluded that the judges' arguments justifying their decisions indicate that the rank order of essays generated by comparative judgement represent 'the whole of academic writing,' and offer a reliable and valid approach to writing assessment.

According to [Ofqual \(2019\)](#), some variations to the comparative judgement process have been suggested to reduce the number of comparisons that need to be made. For example, instead of making multiple paired comparisons, scripts can be ranked in packs of 10, and then that rank order can be converted into multiple sets of paired comparisons. Comparative judgements can

also be used to produce a scale of proficiency in writing, allowing for benchmarks to be identified within that scale.

Changing the focus of the assessment

Creativity and authorial voice

[Humphrey, et al. \(2013\)](#) argued assessments of writing skill in children should focus more on authorial voice, which they define as ‘the unique, personal style that gives the reader a sense of who the author is’. The researchers created a method for assessing authorial voice which looked at four measures of assessment: expressiveness, metanarrative comments, emphasis markers and cultural voice, and compared them with two measures of mechanics: conventions and clarity. They tested this method with 115 primary-age children in the United States, who were given 30 minutes to write a personal narrative in relation to a prompt. Markers then rated the stories on a 3-point scale for the six measures. The study found that markers tended to associate academic prowess with good grammar, spelling and punctuation, neglecting to give academic praise to creativity and individuality in some children’s writing. They concluded that teachers often overlook authorial voice because standardised testing pressures them to focus on state-mandated objectives that emphasise conventions over creative expression, and instead tests must be expanded to include authorial voice.

Regarding creativity in writing, [Morris and Sharplin \(2013\)](#) argued that there has been tension in the discourse about the relationship between analytical and creative writing. In 2008, the Western Australian Certificate of Education (WACE) Literature course, for which students primarily focused on analytical writing, included a mandatory creative writing piece for the coursework component. The distinction made between the two in this context is that analytical writing pertains to writing about literary texts while creative writing pertains to the writing of poetry, plays and narrative prose fiction. However, some proponents argue that analytical writing can be considered creative because the assembly of any words in a purposeful way can be considered an act of creativity. Conversely, there are degrees of analysis involved in creative writing. Morris and Sharplin described how in Western Australia, academics in the field of assessment use analytical marking keys to make creative writing criteria explicit, describing categories within each criterion and defining the relative weighting of each criterion to develop a common understanding of standards and to ensure that the assessment of creative writing is fair and consistent.

Despite the emphasis on creative writing in the curriculum, its acceptance continues to be challenged by perceptions that creativity is difficult to assess objectively, fairly and reliably.

Scenario-based assessment

The Educational Testing Service (ETS) conducted research looking into ‘scenario-based assessments,’ which are assessments that simulate reading and writing tasks within an authentic context ([Deane, 2022](#)). For example, they have created assessments that walk students through the process of building an argument to eventually build and present an argument of their own. It is believed that the scenario structure helps to provide scaffolding that reduces learner cognitive load and increases the efficiency of the writing process. According to Graham et al. (2013), people learn to write well when what they are asked to write about matters to them and their community and helps them to accomplish meaningful goals.

[Thake \(2023\)](#) stressed the need for English assessments that consist of tasks simulating the need for written language within real life contexts, to make constructs and assessment tasks relevant to what learners may face once they have left school. He stated that assessments should, for example, focus on workplace needs to help learners develop and demonstrate the communication skills employers often say young people lack. Another example includes tasks that learners might do on their phones and gamified elements to engage and motivate. Furthermore, Thake asserted that marking of writing assessments should be less focused on assigning a grade and more focused on passing a universal standard of proficiency.

While AQA’s Functional Skills English assessments are designed to consist of writing tasks that simulate the need for English language skills within real life contexts, the Functional Skills suite of qualifications¹ is generally poorly perceived ([Ofqual, 2021](#)). Thake commented that the exam and qualifications system is lacking a universal way to properly assess learners meaningfully and in a way that is beneficial for them, particularly for those who do not achieve a good grade in Maths or English GCSEs. Assessments should show what young people can do with clear criteria against which they are measured — similar to a driving test, as Thake noted.

Other themes in writing assessment

The role of content in marking writing assessment

[Bae et al. \(2016\)](#) drew attention to how content relates across different aspects of writing assessment, namely vocabulary, text length, coherence, originality and grammar. They argued that content is the only element of writing that should be assessed, however, to do this, markers need a clear and consistent definition of content. Their definition of content refers to ‘ideas or

¹ Functional Skills Qualifications (FSQ) provide a resit option for those who do not achieve a grade 4 in GCSE Maths and English. Those who achieve a grade 2 or below have the option to study towards an FSQ while those with a grade 3 normally retake GCSE exams.

meaning expressed in writing, and it is assessed in terms of the extent to which those ideas are elaborated, developed, logical, consistent, interesting, and creative as well as relevant to the task requirements; higher levels of content are characterised by more original reflective and interpretive excellence conveying a depth of understanding' (p. 307). While vocabulary, grammar, and so on are still considered important, they are means to achieve the overall goal of writing, which is to create meaning or ideas. In other words, content is the sum of all other elements of writing (that is, vocabulary, coherence, grammar, creativity and text length).

This seems reminiscent of holistic marking, which consists of forming overall judgements on learners' work where criteria are considered simultaneously, often with the use of a rubric for writing assessment ([Tomas et al., 2019](#)). While holistic marking can be more efficient than analytical marking, which focuses on individual criteria, this approach is limited in that:

- Single overall scores do not clearly communicate to the learner where improvements can be made.
- Inter-rater reliability and content validity (the extent to which the assessment assesses what it intends to assess) can be low.

Nevertheless, Bae et al. (2016) differentiate content-alone assessment from holistic marking. In holistic marking, a holistic score is assigned to a particular feature of writing, which is not clearly defined. In content-alone assessment, that feature is content. For this method to be effective, markers need to have the same definition of content in mind during evaluation. Based on their study carried out with primary-age children in South Korea, the authors argued that assessing content only can serve as a comprehensive and practical measure of overall writing ability.

Learners who struggle with writing

For learners who struggle with writing, [Michael et al. \(2022\)](#) suggested the following considerations to effectively assess their writing:

1. Frequently monitor learner progress. They also suggest the use of tools, such as curriculum-based measurement, which they describe as a timed response to a prompt typically scored for words written, words spelled correctly and correct word sequences. While this seems to contradict more holistic approaches to assess overall writing ability, the authors suggest an assessment focus on these elements (word count, spelling, and so on) may help learners who struggle with writing develop foundational writing skills.
2. Carefully consider writing prompts and whether they set out clear expectations for learners. They should contain, according to Calfee and Greitz Miller (2013), a focus statement that

clearly identifies the topic, a clearly stated purpose for the writing task and genre clue words, such as 'persuade', 'explain', and 'describe'.

3. Have learners self-evaluate their own writing progress with specific guidelines, to help them develop as writers. For example, they could identify how many genre elements or transition words their writing contains. Learners can plot this out on a graph or visual.

Curriculum-based measures

[Dockrell et al. \(2018\)](#) considered the extent to which curriculum-based measures (CBM) are a reliable and valid approach to assessing writing ability, in the context of primary aged learners. CBM is intended to capture developmental differences and changes in writing skills over time. Measures include the number of words written, correct word sequences, punctuation marks and words spelled correctly. The appropriateness of these measures varies by age to reflect different developmental levels.

Texts were compared across Year 3, 4 and 5 pupils in England (ages 8–10). The authors concluded that the strengths of CBM included good reliability of scoring and accurate differentiation across year groups, including for learners with additional support needs. However, it can be argued that CBM only provides partial information on writing as it focuses on the product but not the process of writing.

CBM may serve as a helpful formative assessment or progress monitoring tool with young learners, English language learners or learners who may need additional support with writing development, as CBM was found to be sensitive to changes over the five-month period in Dockrell's et al. (2018) study as well as in a more recent, similar study by [Piercy and Dockrell \(2023\)](#). While perhaps not strongly applicable to older learners, other identified merits of CBM are that it provides an alternative to overly bureaucratic assessment schemes and is less time-consuming than other approaches, while maintaining adequate levels of reliability and sensitivity.

Use of automated tools and AI in writing and writing assessment

There is still work to be done in terms of understanding how AI is being accounted for and/or used in writing assessment. A key theme which emerged in the literature reviewed was the use of AI in developing writing and digital literacy skills. This section examines perceptions and evidence around how AI is affecting not only writing assessment, but writing skill development.

Perceptions and uses of AI tools for writing

It seems that increasingly, AI tools are being integrated into writing practice. [McMurtrie \(2022\)](#) suggested that AI tools 'will become part of everyday writing in some shape or form, in the same way calculators and computers have become part of maths and science.'

[Sanz-Tejeda et al. \(2026\)](#) conducted a systematic literature review, examining 136 articles published between 2023 and 2025 on topics including generative AI and reading and writing skills. The results show a clear trend toward integration of AI tools, particularly ChatGPT, into the academic writing process. Opportunities mentioned were offset by a number of challenges raised.

Key perceived benefits of integrating AI tools were enhanced quality of texts and efficiency of the writing process, as well as the role of AI as a feedback or correction tool. Learners reported improvements to the coherence, discursive organisation, lexical richness and argumentation of their texts. In addition, there were positive learner perceptions around using AI for refining research topics, rephrasing sentences and optimising their time as well as for translation, grammar and error checking, and generating ideas. This resonates with research conducted by the Qualifications Scotland Research and Evaluation Team, which identified similar uses of AI among learners in Scotland ([Mroczkowski et al., 2025](#)). Unsurprisingly, the main challenges identified were around detection of AI-generated content, reliability and digital equity.

Sanz-Tejeda et al. (2026) also found varying opinions across the articles they examined as to whether AI helps or reduces creative thinking in written expression. They found concern expressed around the potential for AI to homogenise language used in academic writing, limiting creativity and potentially curtailing higher-order thinking and writing skills.

In a similar vein, [McMurtrie \(2022\)](#) imagined a future in which the academia is seen to be transactional as a result of AI tools, as completion and submission of work, including written work, becomes more important than the learning process. Similarly, [Dickinson and Marshall \(2026\)](#) found recently that the gap between learner production or submission of work and their ability to explain that work, while it has existed before the increasing use of AI, is now wider.

[Alghasab \(2025\)](#) contradicted this view, purporting that AI tools help to mediate the multiple stages of the writing process and provide the scaffolding needed for learners to enrich their knowledge and generate ideas. This study was conducted in the English foreign language learning context and there is a large body of literature supporting the use of AI for English learners (Alsaleem, 2024; [McMurtrie, 2022](#); [Woo et al., 2025](#)). Findings from Alghasab's study of 16–17-year-old English learners in Kuwait showed positive perceptions and a high degree of

familiarity with AI tools such as Grammarly, ChatGPT and Google Translate. Learners felt that use of these tools provided immediate help in improving their ideas, enhancing vocabulary and detecting grammatical errors. However, they also expressed uncertainty regarding the long-term impact in terms of their writing skills development given that, as a result of interacting with the AI tools, they felt they had not internalised any acquired knowledge. This signals back to concerns around the impact on higher-order thinking skills identified in the systematic review by Sanz-Tejeda's et al. (2026). Nevertheless, Alghasab (2025) argued that it was 'futile to prohibit students from using AI writing tools' as learners are clearly 'AI users' (p. 13).

[Van Niekerk et al. \(2025\)](#) conducted an experiment whereby students used ChatGPT to write an academic paper and then analysed the results. They also did a pre- and post-survey to identify changes in students' perceptions about the usefulness of the technology. Results showed that despite the limitations of ChatGPT identified (that is, hallucinations), students continued to value its usefulness as a writing tool after they had completed the exercise. In particular, students felt that ChatGPT was beneficial for improving the structure, flow and tone of their writing. Thus, while students became more aware of the limitations of AI, they were not dissuaded from using it. This supports Alghasab's argument that swaying learners away from using AI is likely to be a futile endeavour.

Balancing use of AI with the value of writing

The prevalence of AI learner use raises questions as to how to approach teaching and assessing of writing skills. Firstly, to mitigate what is within teachers' control regarding use of AI, McMurtrie (2022) amplifies the importance of the following:

- Reasons for AI malpractice must be closely looked into. McMurtrie suggests that the value of the task or intended learning may not be apparent, courses or curricula may be too convoluted, the task may be too complex, and so on.
- Teachers should consider flipped classrooms in which assignments are completed in class. There could be focus given to multimedia assignments or oral presentations. Learners should be asked to write about topics of genuine interest to them and frequent feedback and discussion should be built in.
- McMurtrie asserts that learners have been trained to write 'simulations like an algorithm' in schools. As such, teachers need to communicate the value of writing to represent their thinking, rather than being taught to repeat 'an assemblage of words back to the teacher.' In other words, according to McMurtrie, learners are not taught *why* they write.

- McMurtrie also highlights the importance of teaching digital literacy, emphasising how AI can amplify biases, such as by running a writing prompt through an AI tool several times and looking for patterns in responses.

Of particular interest is the point McMurtrie raises about teaching learners why writing skills are beneficial, suggesting that learners may not see the benefit.

In terms of how AI factors in to writing assessments, three key themes are evident in this review of literature:

1. Detecting AI-generated content in writing
2. Using AI as a grading or feedback tool for written work
3. Integrating AI to promote digital literacy and inform academic integrity while also improving writing skills

Detection of AI-generated content in written work

There seems to be a tension present in the literature available around the level of policing that should be employed around use of AI in writing, and indeed, whether AI is necessary in an optimal learning environment.

McMurtrie (2022) argued that 'if you can create an atmosphere where students are invested in learning, they are not going to reach for a workaround,' or in other words, they are not going to need to use AI, presumably. Dickinson & Marshall (2026) also stated that learners reach for AI less if their course already provides what AI substitutes. While not explicitly saying that AI should not be used, these authors suggest that learners should not need to resort to use of AI if given the proper conditions.

Another school of thought encourages use of AI tools in a way that maximally enhances outcomes as well as engagement. [Giray et al. \(2025\)](#) believed that detection-focused approaches to AI use in writing are a detriment to the learning environment, creating 'an atmosphere of distrust and anxiety' (p. 93). However, others may argue that AI detectors are needed to prevent plagiarism and maintain academic integrity. Nevertheless, Giray et al. stressed the importance of AI in education, arguing that AI tools make lessons more effective for learners and that automation of certain aspects of writing allows learners to focus on more creative aspects. Quillbot has been found helpful for spotting errors and fixing punctuation, while ChatGPT provides timely support in answering questions and suggesting extra resources. The authors also cited tools that help socio-emotional learning and provide personalised learning experiences (Coursera, Khan Academy and edX). The authors acknowledged that the

use of AI for writing requires clear institutional guidelines as well as AI detection. In addition, the need for a focus on process over product is critical in writing and writing assessment.

Despite the reported effectiveness of AI for writing, the difficulty of differentiating between a learner's own work and AI-generated texts is an ongoing challenge. Findings from [Nguyen and Barrot \(2024\)](#) showed that, despite familiarity with common learner errors and previous engagement with AI writing tools, teachers of English as a second language could not consistently identify AI-generated texts.

AI detectors are limited as they 'rely on algorithms that are only as adequate as the data they are trained on' (Giray et al., 2025, p. 87). While there are several AI detectors on the market (Figure 1), each has its own flaws and should not be relied on entirely without human judgement.

Figure 1: Popular writing detectors available on the market

Writing detector	Description	Key features	Price	Website
Turnitin	A popular plagiarism and AI detection tool used by educational institutions.	AI-generated text detection, plagiarism checker, detailed reports.	Custom pricing for institutions.	turnitin.com
Grammarly (Plagiarism Detection)	Known primarily for grammar checking, also offers a plagiarism detection tool.	Grammar correction, plagiarism detection, writing style enhancement.	Free; Premium plans from \$12/mo.	grammarly.com
Copyleaks	Detects plagiarism and AI-generated content in student and business writing.	Plagiarism, AI detection, multiple file formats, integrations with LMS.	Free for up to 10 pages; paid plans from \$10.99/mo.	copyleaks.com
Quetext	Simple and effective plagiarism and AI content checker with an easy-to-use interface.	DeepSearch™ technology, AI-generated text detection, citation assistance.	Free for basic; paid plans from \$9.99/mo.	quetext.com
Originality.AI	Designed to detect GPT-3, GPT-4, and other AI-generated content in blogs, websites, etc.	AI detection, team management tools, scanning large documents.	\$0.01 per 100 words.	originality.ai
Sapling AI Detector	Primarily an AI writing assistant with added AI-generated content detection.	AI assistance for writing, AI-generated text detection.	Free plan; custom pricing for enterprise.	sapling.ai
GPTZero	Specialized in detecting AI-generated content, especially for academic use.	Detailed breakdowns of AI vs. human-written sections, real-time analysis.	Free basic use; custom pricing for educators.	gptzero.me
Crossplag AI Detector	Offers a cross-language plagiarism detection tool with AI detection capabilities.	Multilingual support, AI detection, detailed reports.	Free trial; paid plans from €12/mo.	crossplag.com
Writer.com (AI Content Detector)	AI detection specifically designed for marketing and business writing.	AI-generated text detection, grammar checks, style guide enforcement.	Free trial; business plans from \$18/mo.	writer.com

Source: Giray et al., 2025, p. 88

Common issues with AI detectors, in addition to perpetuating harmful biases and data quality issues, include lack of contextual understanding, overreliance on certain linguistic patterns and syntax (rigidity), and susceptibility to evasion by workarounds (tools like Undetectable AI).

In an effort to overcome the limitations of current AI detection tools, [Aburass and Abu Rumman \(2024\)](#) have proposed a 'writer's integrity framework,' designed to authenticate the human authorship of a text and distinguish it from AI-generated text. The software is intended to be a standalone word processor that can run as a website or as an application on any operating system. The framework focuses on the dynamics of the writing process rather than concentrating on textual analysis alone. The stages of the proposed framework and its intended use is as follows:

- Real-time writing activity logging — monitors and logs real-time writing activities in a designated text area (things like keystroke, edits (additions and deletions), and paste actions performed by the writer)
- Change detection — as the writer composes or modifies text, the framework detects and logs changes between the previous and current drafts of the text
- Paste action logging — special attention given to text pasted into the document, records length of the pasted text and its positioning within the document
- Analysis of writing behaviour — all the data is analysed to calculate key metrics that characterise an individual's writing behaviour, including typing speed, frequency of edits and the ratio of pasted text to overall content
- Log cleaning and data condensation — the framework cleans the data, removing redundant information and summarising writing activity to optimise storage

This framework is not without its own limitations. Writers might find the system intrusive and additionally, the system would need to ensure that strict data security measures are in place. The authors have considered that future work in developing the framework may involve a built-in paraphraser to determine whether the writer's paraphrased content is appropriate and to give due credit. Like Turnitin, the framework is also expected to have a similarity measure. What this framework does point out is the importance of teachers having good familiarity with their learners' writing quality.

In sum, there is currently no AI detector that can differentiate between human and AI-generated texts with 100% accuracy. Human oversight as well as knowledge of the writer's writing history and behaviours must be accompanied with any use of an AI detection tool.

Using AI as a grading or feedback tool on written work

Since the late 1990s and early 2000s, the ETS has found that automated writing and evaluation tools can be used effectively to assess the quality of learner writing when used alongside human judgement. Given that human raters are subjective in their evaluations to a degree, and different raters place greater emphasis on different aspects of writing, automated or AI marking tools may act as a complement to human marking in the absence of a mutually agreed upon standard of writing quality (Deane, 2022).

Ofqual (2019) also acknowledged the limitations of marker reliability and suggested that automatic essay scoring (AES) potentially 'reduces' the need for human markers, though it does not 'replace' human markers. AES may struggle to recognise written features such as creativity, reader-based prose and persuasiveness, so while it may help as a marker monitoring tool (for example, flagging human-computer mark discrepancies), concerns over validity may be too great to warrant its use.

The same can be said for today's advanced AI models. Educational service 'No More Marking' reflected on the reliability and validity of their AI judges in a recent [online article](#). While they acknowledge that humans will 'always need to keep monitoring and ensuring the AI is aligned with humans', they advertise a '[90% AI – 10% human model](#)', in which AI performs comparative judgement that reduces the time traditionally taken to mark a class set of essays from 5 hours to 15 minutes. This may seem like a high proportion of AI to human involvement.

Findings from recent studies seem to support the use of AI in feedback and marking. In a recent quasi-experimental study, 13–14-year-old learners were placed into two groups. One group received generative AI-supplemented feedback in addition to teacher feedback and peer assessment on weekly writing topics for a period of six weeks. The other group received only the teacher feedback and peer assessment ([Huang et al., 2026](#)). While outcomes for both groups improved, the group that received the AI-supplemented feedback showed greater gains in writing proficiency and were able to score their peers' writing more accurately. The inclusion of AI in the writing feedback process also resulted in more favourable learner attitudes toward perceived usefulness of feedback approaches. The researchers concluded that AI-supplemented feedback and teacher feedback complemented each other.

Similar findings were noted by researchers who compared different approaches to feedback for medical students. Students who received combined AI support with human feedback performed best, while those using AI support alone not only scored lowest, but were also the most confident, despite being the least skilled (Stephenson & Armstrong, 2026). Without the support of human feedback, students who leaned too heavily on the AI tool missed key nuances,

leading to overconfidence. The results suggest that AI can be transformative only when used appropriately and under strong oversight.

Another study in the higher education context found that English as a foreign language (EFL) teachers positively perceived the use of CoGrader, specifically, an assistant tool in scoring written essays and in providing feedback (Alsalem, 2024). However, the authors stressed that the software should not be accepted as a complete substitute to teacher grading but encouraged teachers to 'engage cautiously' with CoGrader, and other grading tools.

Examples of AI use in the development and assessment of writing skills

AI Assessment scale

The AI Assessment Scale (AIAS) was first developed by Perkins et al. (2023) and has since been updated to account for more nuance and complexity in AI use for written assessment.

The original framework was designed to guide communications with learners about the permissibility of AI tools in assessment and ensure responsible use. It categorised AI use in assessments into three areas ([Meakin, 2024](#)):

1. AI use not permitted — AI tools may not be used to complete any portion of an assignment.
2. AI can be used in specific ways — AI tools may be used by learners in certain ways, but not in others; educators should clearly communicate expectations.
3. AI is permitted — Learners are permitted to use AI tools to support their learning as they complete the assignment (for example, brainstorming, planning, drafting, revising).

The latest iteration of the AI Assessment Scale can be seen in Figure 2.

Figure 2: The AI Assessment Scale

1	NO AI	The assessment is completed entirely without AI assistance in a controlled environment, ensuring that students rely solely on their existing knowledge, understanding, and skills You must not use AI at any point during the assessment. You must demonstrate your core skills and knowledge.
2	AI PLANNING	AI may be used for pre-task activities such as brainstorming, outlining and initial research. This level focuses on the effective use of AI for planning, synthesis, and ideation, but assessments should emphasise the ability to develop and refine these ideas independently. You may use AI for planning, idea development, and research. Your final submission should show how you have developed and refined these ideas.
3	AI COLLABORATION	AI may be used to help complete the task, including idea generation, drafting, feedback, and refinement. Students should critically evaluate and modify the AI suggested outputs, demonstrating their understanding. You may use AI to assist with specific tasks such as drafting text, refining and evaluating your work. You must critically evaluate and modify any AI-generated content you use.
4	FULL AI	AI may be used to complete any elements of the task, with students directing AI to achieve the assessment goals. Assessments at this level may also require engagement with AI to achieve goals and solve problems. You may use AI extensively throughout your work either as you wish, or as specifically directed in your assessment. Focus on directing AI to achieve your goals while demonstrating your critical thinking.
5	AI EXPLORATION	AI is used creatively to enhance problem-solving, generate novel insights, or develop innovative solutions to solve problems. Students and educators co-design assessments to explore unique AI applications within the field of study. You should use AI creatively to solve the task, potentially co-designing new approaches with your instructor.

Source: Perkins et al. (2024)

As the text in the image may be difficult to see, the five levels are reiterated as follows:

1. **No AI** — ‘The assessment is completed entirely without AI assistance in a controlled environment, ensuring that students rely solely on existing knowledge, understanding, and skills.’
2. **AI Planning** — ‘AI may be used for pre-task activities such as brainstorming, outlining and initial research. This level focuses on the effective use of AI for planning, synthesis, and ideation, but assessments should emphasise the ability to develop and refine these ideas independently.’
3. **AI Collaboration** — ‘AI may be used to help complete the task, including idea generation, drafting, feedback, and refinement. Students should critically evaluate and modify the AI suggested outputs, demonstrating their understanding.’
4. **Full AI** — ‘AI may be used to complete any elements of the task, with students directing AI to achieve the assessment goals. Assessments at this level may also require engagement with AI to achieve goals and solve problems.’
5. **AI Exploration** — ‘AI is used creatively to enhance problem-solving, generate novel insights, or develop innovative solutions to solve problems. Students and educators co-design assessments to explore unique AI applications within the field of study.’

A pilot study conducted by [Furze et al. \(2024\)](#) explored use of the AI Assessment Scale in educational assessments in a higher education context where most students are English as a second language learners. Exams and assessments set by professional bodies or activities requiring live demonstrations were restricted to Level 1 on the scale. English language programmes were restricted to Level 2 and take-home assessments were restricted to Level 3. The report does not state how many assessments were submitted by level, but the total number of assessments submitted can be seen in Figure 3. Faculty were encouraged to consider setting assessments at Levels 4 and 5, however there were no reported outcomes around this. The authors noted an overall reduction in AI-related misconduct cases over a one-year period and believed this can be attributed to increased awareness among students of the risks they faced when using AI tools in their assessment.

Figure 3: Academic misconduct results January 2023 – January 2024

Semester start date	Number of assignment submissions	Detected cases of misconduct	Misconduct cases as % of assignment submissions	AI-related misconduct cases	AI-related misconduct cases as % of misconduct cases	AI misconduct as % of assignment submissions
January 2023	1,722	145	8.42%	112	77.24%	6.50%
April 2023	5,255	99	11.57%	86	86.87%	1.64%
July 2023	1,576	7	3.46%	4	57.14%	0.25%
October 2023	3,,996	1	0.13%	0	0.00%	0.00%
January 2024	4159	24	0.67%	4	16.67%	0.10%

Source: Furze et al., 2024, p. 9

There was also a rise in academic performance, which may be in part due to AI helping students overcome English language barriers. As a result of this study, the authors are in strong support of the use of AI to enhance learner engagement and stated that ‘the narrative surrounding GenAI (in higher education) must shift beyond cheating.’ (p. 11).

With regard to academic integrity, it is important to recognise that some forms of assessment may lend themselves to academic dishonesty more than others (Furze et al., 2024), and indeed, some assessments may lend themselves better to use of AI than others. When AI is deemed inappropriate for certain assessments, Meakin (2024) recommended that the assessment should be challenging to complete by using AI tools or by copying from external sources. These assessments could potentially take any of the following formats:

- Assessments that demonstrate critical thinking, problem solving, analysis and projects that demand application of knowledge in unique ways make it less likely that learners will use AI

- Projects, presentations or performances can encourage collaboration and discourage individual cheating
- Conducting (formative) assessments in class, both written and oral, which include analysis that draws on class discussions; tasks that allow educators to measure understanding in real-time and to monitor progress
- Implementing time limits to limit time available for learners to consult AI-generated content
- Open-book exams with a twist that include questions requiring higher-order thinking, require learners to work with reference materials solely
- Oral assessments that require learners to defend a position
- Adaptive testing tailored to a learner's performance
- Peer review and evaluation processes for assignments and projects that can help to identify discrepancies in group work, lead to revisions of work and encourage learners to hold each other accountable
- Staged submission of writing assessment (require an outline, research notes and drafts, and so on, and have learners respond to feedback)
- Opportunities to teach-back, communication confirmation methods whereby learners demonstrate their understanding through speech

Alternatives to essays

Furze (2024) acknowledged that while essays have traditionally been a straightforward way to allow students to demonstrate their knowledge, easy access to automated writing tools warrant an exploration of alternative methods. However, the use of AI is not the only rationale for considering alternative assessment methods, as essays can also be time-consuming, subjective, stressful and pose risks such as a writing skill bias or plagiarism.

Alternatives proposed include:

- performance-based assessment that encourages creative writing, visual storytelling or public speaking
- portfolios or writing journals
- projects (such as creating a literary magazine with different genres of literature along with a literary analysis)
- observations or peer evaluations
- 'visual essays' providing a critical commentary on a specific topic, accompanied by a sequence of images

Corbin et al. (2025) advocated 'disrupting' today's standard academic essay, rather than abandoning the essay entirely. Essays often follow a formulaic structure resembling a research report or scientific article. The authors argued that society has lost sight of what essays historically set out to do — allowing understanding to emerge through the process of essay writing itself. The authors distinguished essay writing, which is exploratory in nature, from 'systematic writing', which is described as beginning with a conclusion and working backwards to support that conclusion (Corbin et al., 2025, p. 200).

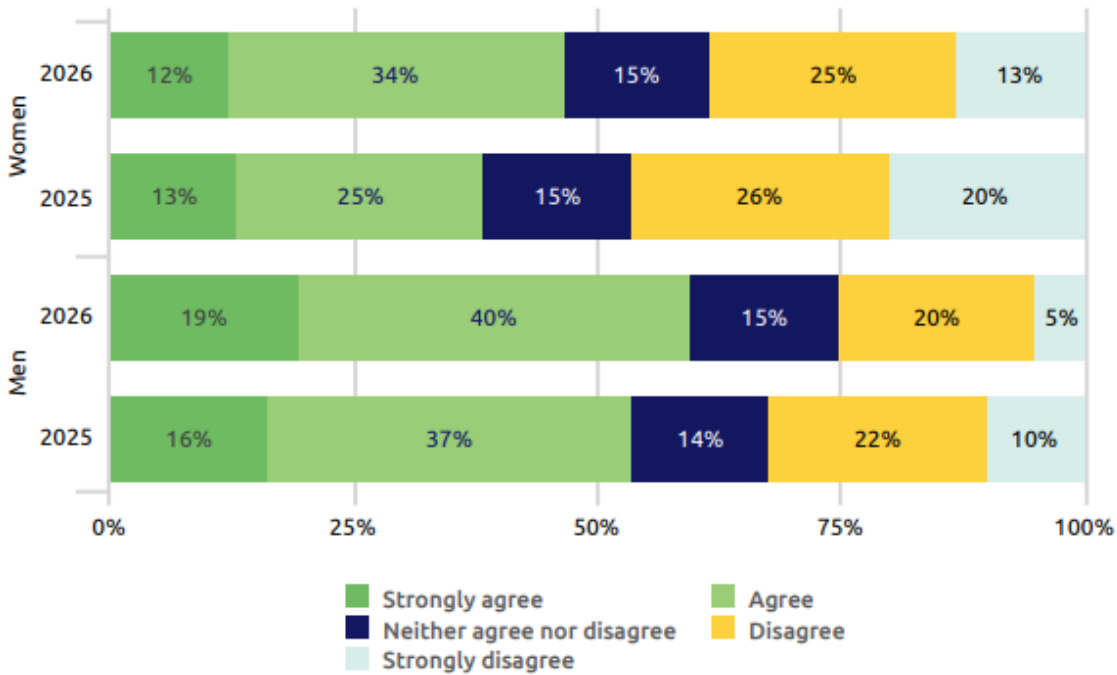
Integrating AI into the writing process to promote digital literacy and inform academic integrity

Notwithstanding assessment design considerations that curtail the use of AI, several recent studies make a case for why allowing for use of AI to enhance writing ultimately helps to inform academic integrity, as well as digital literacy. However, first and foremost, learners must be appropriately supported by being made aware of the limitations of using AI. These include incorrect or inaccurate information (or misinformation); biased, prejudiced and stereotypical responses; and when AI lacks understanding of the context of the prompts or queries (Meakin, 2024).

A recent report published by the Higher Education Policy Institute (HEPI) recommended that clear, accessible and assessment-specific guidance on AI use will maintain academic integrity and allow students to be assessed on the AI skills they will need in the future ([Stephenson & Armstrong, 2026](#)). The researchers administered a survey to 1,054 undergraduate students covering questions around student AI use and institutional approaches to AI. Below are some key findings from the survey:

- 37% of students reportedly used AI to enhance or edit their writing, using tools such as Grammarly, and cited that using AI tools for writing assignments helped 'take the pressure off'. (p. 23)
- Students also expressed concerns about being falsely accused of cheating, describing this as a 'constant worry' despite some having never used AI to write assignments. (p. 21)
- Men were more likely to have had experience of using AI at school before arriving in higher education. (See Figure 4)

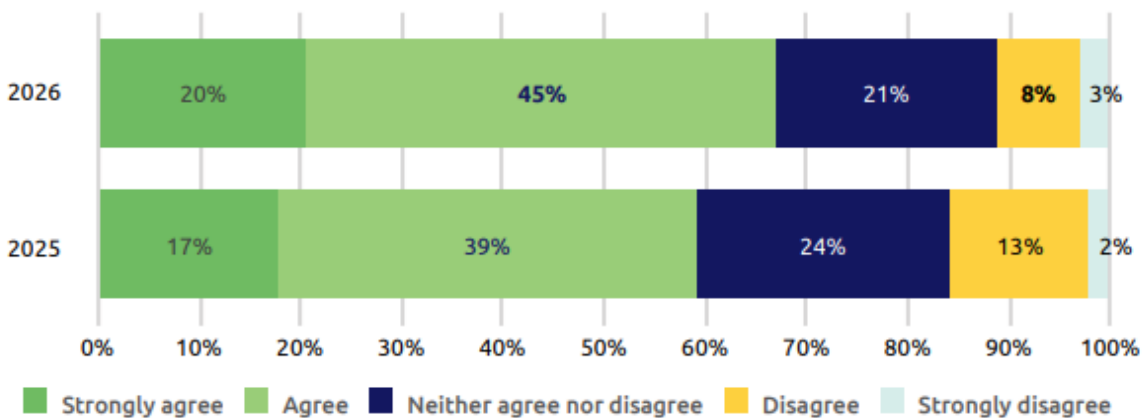
Figure 4: Proportion of students who had experience using AI at school before arriving in higher education — differences by gender



Source: Stephenson & Armstrong, 2026, p. 27

Transformation of assessments as a result of AI has increased between the previous survey cycle (2025) and the 2026 survey, with 9% more students agreeing that assessments had changed in 2026 than in 2025 (Figure 5)

Figure 5: Proportion of students agreeing or disagreeing that AI tools have changed the way their institutions assess



Source: Stephenson & Armstrong, 2026, p. 30

The results show that institutions are becoming less sceptical of AI for learning and assessment (though still not wholly supportive). The authors recommended that students should have

provision for practising AI skills for assessment alongside AI-free assessments to maintain academic integrity and standards. The authors also cautioned that ‘institutions should ensure that they are continuing to specifically teach traditional skill sets, such as essay writing, that may be underdeveloped in some students who have already become reliant on AI’ (pp. 41–42).

Using AI to enhance digital literacy while improving academic writing skills

[Woo et al. \(2025\)](#) developed two co-curricular interventions aimed at addressing both academic writing and AI literacy gaps among EFL secondary students in Hong Kong. The interventions differed only in structure: one comprised one-hour lessons delivered over a period of eight weeks, while the other was a single four-hour workshop. Modules taught writing concepts, including genre-based and process-based approaches to academic writing, alongside development of AI prompt engineering skills. The interventions included two final academic writing tasks, in which students had a choice between two topics: computational thinking and English language learning; or computational thinking and AI education. Students primarily engaged with chatbots available to them in Hong Kong, such as HuggingChat, Copilot, Mistral and Perplexity.

Results from formative Kahoot assessments and evaluation of the final writing task showed improvements, particularly in prompt engineering skills, critical evaluation of AI-generated content and ethical considerations. Students also showed knowledge of genre writing but a lack of knowledge around some aspects of process writing, such as adding, deleting and re-arranging ideas. This could indicate that more time spent on explicitly teaching planning and reviewing skills before introducing AI-supported writing tasks may be beneficial for students. There were no significant differences found between the intervention formats, suggesting that condensed workshops can be equally effective as extended courses.

The authors recommended added scaffolding for future interventions to allow students to develop their understanding of writing and AI concepts before guided application and independent practice. It was also suggested that engaging with a variety of tools can help inform students of individual strengths and weaknesses, allowing them to select the most appropriate tool for their needs.

Effect of AI on learners’ identity and teachers’ practices in the high school English classroom

[Riser \(2025\)](#) also examined how AI was embedded into curricula, with a focus on preserving learner voice and maintaining academic integrity in the United States secondary context. Analysis of classroom videos, teacher interviews and learner reflective journals showed that

learners displayed greater autonomy when working collaboratively to evaluate AI-generated text. Teachers played the role of ethical facilitators by highlighting transparency, accountability and reflective dialogue. The findings suggest that cultivating shared moral responsibility can encourage ethical and responsible use of AI. During class activities, learners worked together to weigh the authenticity of AI-generated content and ‘co construct AI limits.’ The authors recommended explicit protocols for AI tasks, class communities that co-author ethical guidelines and the positioning of teachers as ‘ethics coaches.’

Conclusion

This literature review concludes with some broad considerations in light of how writing skills, as well as writing assessment, are approached and evolving with the spread of generative AI use.

Enduring approaches to the assessment of writing include criterion-based frameworks, holistic methods and scenario-based assessments that reflect authentic settings and purpose. The literature also highlights themes such as authorial voice in writing, the use of comparative judgement and the role of content in evaluating writing ability. While analytical approaches offer structure and reliability, they often involve subjectivity and may overlook creativity or the broader purpose of writing. Formative tools like curriculum-based measures can support specific learner needs, particularly for those developing foundational skills. As generative AI gains prominence, these established approaches need to be reconsidered to ensure they remain valid, reliable and aligned with evolving definitions of writing competence.

Effective use of AI in writing, and writing assessment, may require a shift from viewing AI as a threat or shortcut. However, AI tools must be considered in light of where they can add value and where independent, human effort remains essential. A strategic approach may involve aligning use of AI with the assessment purpose. Frameworks such as the AI Assessment Scale demonstrate how structured levels of AI use can help ensure that AI enhances rather than undermines validity. Clarity around expectations is key. This includes explicit guidelines on permitted and non-permitted uses of AI.

Live demonstrations and modelling of AI use can help to enhance informed and reflective learner engagement with the tool. Demonstrations can emphasise how prompts shape outputs, how errors, hallucinations or biases emerge, and how outputs can be refined through structured iteration. This can help to develop learners’ critical evaluation skills and digital literacy skills.

A key tension identified in the literature lies between maintaining integrity and avoiding a ‘policing’ approach. An excessive focus on detection may create an environment of distrust,

which can ultimately be a detriment to the learning process. A balanced approach that encourages shared responsibility, where learners understand expectations and the rationale for using or not using AI, could encourage proportionate and appropriate use of AI tools.

Assessments can also be designed with reducing reliance on AI detection in mind (that is, process-based or staged, as well as in-class components). The literature supports reframing integrity as a pedagogical issue rather than a compliance one, encouraging trust and transparency. As stated by Giray et al. (2025), 'instead of using AI writing detectors just to catch students out, instructors should consider how these AI tools fit into the bigger picture of teaching and learning the writing process' (p. 99).

AI detectors can be positioned more strategically as diagnostic or formative tools. For example, they can highlight sections of text that require further scrutiny, prompt dialogue between educators and learners, and support reflection over writing processes.

As a feedback tool, AI can help to identify surface-level issues, provide suggestions for structure or development, and support drafting and revision cycles. The literature suggests that AI feedback can be particularly effective when complemented by human feedback. Treating AI as part of a process-oriented model of writing development can emphasise writing as a process rather than simply a product.

Finally, nurturing learner curiosity, experimentation and critical engagement can support metacognition and critical thinking, particularly when experimentation is guided. This also aligns with broader goals of preparing learners for a context in which AI is already a part of professional and academic practice.

Further reading

Department for Education, Writing Framework in England (July 2025)

https://assets.publishing.service.gov.uk/media/68bec95444fd43581bda1c86/The_writing_framework_092025.pdf.

References

- Aburass, S. & Abu Rumman, M. (2024). Authenticity in authorship: the Writer's Integrity framework for verifying human-generated text. *Ethics and Information Technology*, 26(62), 1–12. <https://doi.org/10.1007/s10676-024-09797-z>.
- Alghasab, M. B. (2025). English as a foreign language (EFL) secondary school students' use of artificial intelligence (AI) tools for developing writing skills: Unveiling practices and perceptions. *Cogent Education*, 12(1), 1–17. <https://doi.org/10.1080/2331186X.2025.2505304>.
- Alsalem, M. S. (2024). EFL teachers' perceptions of the use of an AI grading tool (CoGrader) in English writing assessment at Saudi universities: an Activity Theory Perspective. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2430865>.
- Bae, J., Bentler, P.M. & Lee, Y. (2016). On the role of content in writing assessment. *Language Assessment Quarterly*, 13(4), 302–328. <http://dx.doi.org/10.1080/15434303.2016.1246552>.
- Berge, K. L., Skar, G. B., Matre, S., Solheim, R. Evensen, L. S., Otnes, H. & Thygesen, R. (2019). Introducing teachers to new semiotic tools for writing instruction and writing assessment: Consequences for students' writing proficiency. *Assessment in Education: Principles, Policy & Practice*, 26(1), 6–25. <https://doi.org/10.1080/0969594X.2017.1330251>.
- Bramley, T. & Vitello, S. (2019). The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(1), 43–58. <https://doi.org/10.1080/0969594X.2017.1418734>.
- Calfee, R. C. & Greitz Miller, R. (2013). Best practices in writing assessment for instruction. In S. Graham, C. A. MacArthur & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (pp. 351–377). Gilford Press.

- Corbin, T., Walton, J., Bannister, P. & Deranty, J. P. (2025). On the essay in a time of GenAI. *Educational Philosophy and Theory*, 58(3), 198–210.
<https://doi.org/10.1080/00131857.2025.2572802>.
- Deane, P. (2022). The Importance of Assessing Student Writing and Improving Writing Instruction. ETS. Available from: <https://files.eric.ed.gov/fulltext/ED622583.pdf>.
- Dickinson, J. & Marshall, M. (2026) Trained to stop learning? How students are experiencing assessment and learning in an age of AI. WonkHE. Available from:
<https://wonkhe.com/wp-content/wonkhe-uploads/2026/03/Trained-to-stop-learning-F.pdf>.
- Dockrell, J., Connelly, V., Walter, K. & Critten, S. (2018). The role of Curriculum Based Measures in assessing writing products. In B. Miller, P. McCardie & V. Connelly (Eds.), *Writing development in struggling learners: Understanding the needs of writers across the life course*, Studies in Writing, vol. 35, Brill Academic Publishers.
<https://pure.coventry.ac.uk/ws/files/22329700/Binder7.pdf>.
- Dunsmuir, S., Kyriacou, M., Batuwitage, S., Hinson, E. Ingram, V. & O’Sullivan, S. (2015). An evaluation of the Writing Assessment Measure (WAM) for children’s narrative writing. *Assessing Writing*, 23, 1–18. <https://doi.org/10.1016/j.asw.2014.08.001>.
- Feenstra, H. (2012). Mixed Methods: Using a Combination of Techniques to Assess Writing Ability. Available from: <https://scispace.com/pdf/mixed-methods-using-a-combination-of-techniques-to-assess-33luze03q6.pdf>.
- Furze, L. (2024). Rethinking assessment for Generative AI: Beyond the essay. Available from: <https://rethinkingassessment.com/rethinking-blogs/rethinking-assessment-for-generative-ai-beyond-the-essay/>.
- Furze, L., Perkins, M., Roe, J. & MacVaugh, J. (2024) The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, 40(4), 1–18. <https://doi.org/10.14742/ajet.9434>.
- Giray, L., Sevnarayan, K. & Ranjbaran Madiseh, F. (2025). Beyond policing: AI writing detection tools, trust, academic integrity and their implications for college writing. *Internet Reference Services Quarterly*, 29(1), 83–116.
<https://doi.org/10.1080/10875301.2024.2437174>.

- Graham, S., Harris, K. R. & McKeown, D. (2013). The writing of students with LB and meta-analysis of SRSD writing intervention studies: Redux. In L. Swanson, K. R. Harris & S. Graham (Eds.), *Handbook of Learning Disabilities* (2nd ed., pp 405–438). Guilford Press.
- Huang, B., Xie, S. & Hew, T. K. F. (2026). Effects of integrating generative AI, teacher and peer feedback on Middle school students' attitudes, writing proficiency and peer-assessment scoring ability: A quasi-experimental study. *Educational Psychology*, 1–23.
<https://doi.org/10.1080/01443410.2025.2611980>.
- Hopfenbeck, T. N. (2019). Writing assessment, comparative judgement and students' evaluative expertise. *Assessment in Education: Principles, Policy & Practice*, 26(1), 1–5.
<https://doi.org/10.1080/0969594X.2019.1577959>.
- Humphrey, R. C., Walton, M. D. & Davidson, A. J. (2013). “I’m gonna tell you about it”: Authorial voice and conventional skills in writing assessment and educational practice. *The Journal of Educational Research*, 107(2), 111–122.
<https://doi.org/10.1080/00220671.2013.788990>.
- McMurtrie, B. (2022). AI and the future of undergraduate writing. *The Chronicle of Higher Education*. Available from: <https://www.chronicle.com/article/ai-and-the-future-of-undergraduate-writing?sra=true>.
- Meakin, L. A. (2024). AI and assessment: Rethinking assessment strategies and supporting students in appropriate use of AI. Available from:
https://my.chartered.college/impact_article/ai-and-assessment-rethinking-assessment-strategies-and-supporting-students-in-appropriate-use-of-ai/.
- Michael, E. L., Ray, A. & McKeown, D. (2022) Assessment for effective strategy instruction in writing. *Preventing School Failure: Alternative Education for Children and Youth*, 66(4), 327–332. <https://doi.org/10.1080/1045988X.2022.2076642>.
- Morris, G. & Sharplin, E. (2013). The assessment of creative writing in senior secondary English: A colloquy concerning criteria. *English in Education*, 47(1), 49–65.
<https://doi.org/10.1111/eie.12004>.
- Mroczkowski, M., Lawson, J. & Green, J. (2025). Generative artificial intelligence in Scottish education: Learner experiences and attitudes in schools and colleges. *Qualifications Scotland* (previously SQA). Available from: https://www.sqa.org.uk/sqa/files_ccc/ai-learner-experiences-and-attitudes.pdf.

- Nguyen, L. & Barrot, J. S. (2024). Detecting and assessing AI-generated and human-produced texts: The case of second language writing teachers. *Assessing Writing*, 62. <https://doi.org/10.1016/j.asw.2024.100899>.
- Ofqual. (2019). A review of approaches to assessing writing at the end of primary education: Drawing upon historical and international approaches. Available from: https://assets.publishing.service.gov.uk/media/5c9cdfd0e5274a5278f2755d/International_primary_writing_review_-_FINAL_28.03.2019.pdf.
- Ofqual. (2021). Ofqual public sentiment survey 2021. Available from: <https://www.gov.uk/government/publications/ofqual-public-sentiment-2021/ofqual-public-sentiment-survey-2021>.
- 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>.
- Piercy, W. & Dockrell, J. E. (2023). Assessing children's writing products using curriculum-based measures of writing (CBM-W). *Educational and Child Psychology*, 40(1), 44–60. <https://doi.org/10.53841/bpsecp.2023.40.1.44>.
- Riser, P. (2025). Generative Artificial Intelligence (AI) in the high school English classroom and its effect on students' identity and teachers' practice. *English in Education*, 59(2), 176–192. <https://doi.org/10.1080/04250494.2025.2491361>.
- Sanz-Tejeda, A., Domínguez-Oller, J. C., Baldaquí-Escandell, J. M., Gómez-Díaz, R. & García-Rodríguez, A. (2026). The impact of generative AI on academic reading and writing: A synthesis of recent evidence (2023-2025). *Frontiers in Education*, 10, 1–14. <https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1711718/full>.
- Stephenson, R. & Armstrong, C. (2026). Student Generative AI Survey 2026. HEPI. Available from: <https://www.hepi.ac.uk/wp-content/uploads/2026/03/HEPI-Report-199-Gen-AI-Survey-2026.pdf>.
- Stuart, N. J. & Barnett, A. L. (2023). The Writing Quality Scale (WQS): A new tool to identify writing difficulties in students. *British Journal of Special Education*, 50(2), 195–332. <https://doi.org/10.1111%2F1467-8578.12464>.

- Thake, A. S. (2023). "A, B, C, it's as easy as 1, 2, 3": Towards new assessments for numeracy, literacy and digital fluency. AQA. Available from:
<https://filestore.aqa.org.uk/content/about-us/education-policy/AQA-TOWARDS-NEW-ASSESSMENTS-NLD.PDF>.
- Tomas, C., Whitt, E., Lavelle-Hill, R. & Severn, K. (2019). Modelling holistic marks with analytic rubrics. *Frontiers in Education*, 4(89), 1–19. <https://doi.org/10.3389/feduc.2019.00089>.
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V. & De Maeyer, S. (2019). Validity of comparative judgement to assess academic writing: Examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, 26(1), 59–74. <https://doi.org/10.1080/0969594X.2016.1253542>.
- van Niekerk, J., Delport, P. M. J. & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. *Computers and Educations: Artificial Intelligence*, 8, 1–11. <https://doi.org/10.1016/j.caeai.2024.100342>.
- Woo, D., Kizilates, S., Guo, K. & Wang, D. (2025). Developing English academic writing and AI literacy in an EFL co-curricular secondary school context. *Innovation in Learning and Teaching*, 1–17. <https://doi.org/10.1080/17501229.2026.2636962>.