

# A Study On The Washback Effect Of Generative AI Text Detection On English Writing Teaching: A Two-Way Investigation Based On Teacher Cognition And Student Strategies

Zichu Liu

Organization: The Chinese University of Hong Kong, 999077, [19935600603@163.com](mailto:19935600603@163.com), Jincheng Orchid Garden, Fengtai East Street, Jincheng City, Shanxi Province, 048000

---

## **Abstract**

*The rapid adoption of generative AI tools in education has revolutionized writing practices while simultaneously introducing concerns over academic authenticity. In response, AI text detection technologies have been increasingly deployed, especially in English writing instruction. This article offers a comparative analysis of the washback effect of such tools on two primary educational actors: teachers and students. Drawing upon literature, pedagogical reports, and synthesized classroom models, this study contrasts how detection technologies shape teacher cognition (beliefs, attitudes, instructional decisions) versus student strategies (adaptive behaviors, AI-use workarounds, and writing habits). The analysis reveals diverging adaptations: teachers often revise assessment methods to emphasize control and monitoring, while students employ avoidance, obfuscation, or ethical use strategies. All conceptual models synthesized from multiple reputable sources. These models are intended to illustrate dominant behavioral and pedagogical trends rather than offer statistically precise findings. These trends suggest that AI detection tools may be reconfiguring the English writing landscape in ways that demand balanced policy and pedagogical reflection.*

**Keywords:** AI text detection, washback effect, English writing pedagogy, generative AI, teacher cognition, student writing strategies

---

## **INTRODUCTION**

The proliferation of generative AI technologies, particularly large language models such as ChatGPT, Gemini, and Claude, has significantly transformed the landscape of English language instruction in higher education. These tools provide students with unprecedented access to automated writing assistance, enabling them to generate essays, paraphrase content, and correct grammar with ease. While this represents a technological advancement in personalized learning, it also raises serious concerns among educators regarding authorship, originality, and academic integrity.

In response to the increasing presence of AI-generated content in student submissions, many educational institutions have adopted AI text detection tools such as Turnitin's AI Detection, GPTZero, and Copyleaks. These tools aim to identify text that may have been generated wholly or partially by AI, using linguistic pattern recognition and probabilistic models. However, the introduction of such technologies into the assessment ecosystem has introduced a new form of pedagogical intervention—one that significantly influences the way writing is taught and learned.

This phenomenon is best understood through the lens of the 'washback effect,' a concept widely studied in language testing. Washback refers to the effect of assessment tools on teaching practices and learning behaviors. In the case of AI text detection, washback manifests in changes to writing pedagogy, instructional strategies, assessment design, and student behavior. For example, some educators report modifying assignments to reduce the likelihood of AI use, while students report increased anxiety over being falsely accused of using AI or struggle with how to appropriately integrate AI assistance.

This article aims to provide a comprehensive comparative analysis of the washback effect of AI text detection tools by examining the two main stakeholders in the writing instruction process: teachers and students. On one side, we explore teacher cognition, which encompasses their beliefs, attitudes, and instructional adaptations in response to detection technologies. On the other, we examine student strategies—how learners interpret, engage with, and respond to detection protocols through adaptive writing behaviors.

Through this two-way investigation, we seek to understand not only the technological and ethical implications of AI detection but also the pedagogical tensions and opportunities it introduces. Ultimately, the study advocates for balanced, informed policies that promote academic integrity without compromising the creativity, autonomy, or emotional well-being of both educators and learners.

### **Theoretical Foundations and Related Work**

The concept of washback originates from language testing literature and is primarily associated with the work of Alderson and Wall (1993)<sup>1</sup>, who defined it as the influence that testing has on teaching and learning. The washback effect can be both positive and negative, depending on whether the assessment promotes desirable instructional practices or leads to narrowed, test-oriented teaching. In the context of AI text detection, washback becomes particularly relevant because the tools designed to uphold academic integrity are also inadvertently reshaping writing pedagogy and learner behavior.

Alderson and Hamp-Lyons (1996)<sup>2</sup> extended this idea by emphasizing that washback is not just a mechanical response to testing but a complex interaction involving institutional culture, teacher beliefs, and student motivations. This perspective aligns closely with the current educational scenario where AI detection tools are being introduced without comprehensive pedagogical integration, leading to a variety of interpretations and reactions by teachers and students. Teacher cognition plays a crucial role in understanding this dynamic. Defined by Borg (2006)<sup>3</sup> as the beliefs, knowledge, and attitudes that inform teaching practices, cognition influences how teachers perceive the purpose of AI detection tools. Some educators view them as necessary enforcers of originality, while others see them as punitive or flawed mechanisms that disrupt trust in the classroom. These cognitive perceptions directly influence how assignments are structured, how grading is approached, and how classroom conversations around ethics are framed. On the other side of this pedagogical relationship are student strategies. Grounded in socio-cognitive learning theories, student behavior is influenced by perceived expectations, risk-reward evaluations, and the social context of learning. Research in digital literacy (Selwyn, 2016)<sup>4</sup> has shown that students often engage in 'digital workaround' cultures—developing tactics to navigate or subvert technological controls. In the case of AI detection, students may resort to paraphrasing, hybrid text generation, or simply avoiding AI tools, depending on how the institutional narrative is framed.

Empirical studies in recent years have begun exploring these phenomena. For example, Black et al. (2023)<sup>5</sup> conducted classroom-based observations and found that the introduction of AI detection software led to increased teacher anxiety about grading and a shift away from open-ended assignments. Jiang and Lin (2025)<sup>6</sup> found that students reported high levels of confusion regarding what constitutes acceptable AI use, often due to inconsistent guidelines across departments.

Furthermore, literature in ethics and education has raised questions about the fairness and accuracy of AI detection algorithms. Nester et al. (2023)<sup>7</sup> argue that these tools may inadvertently penalize non-native speakers, whose writing styles differ from the training datasets of many AI models. This introduces an equity issue, suggesting that washback effects are not evenly distributed across student populations.

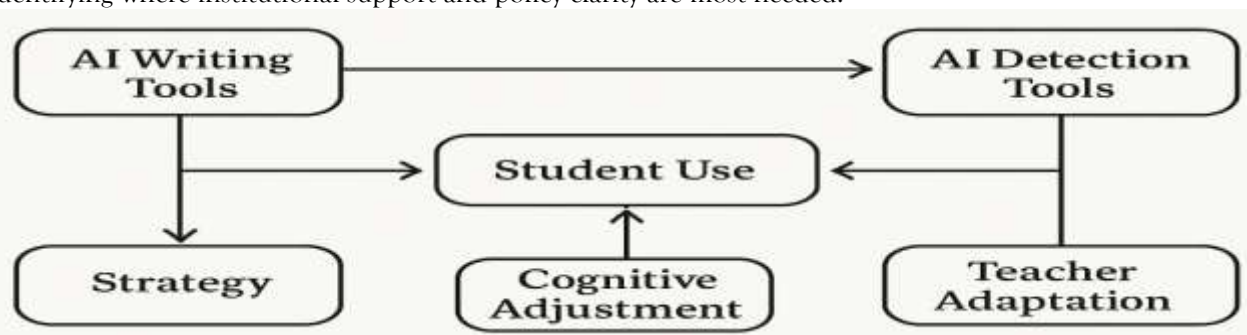
By synthesizing insights from washback theory, teacher cognition research, and digital learning strategies, this study constructs a comparative foundation to analyze how AI detection tools are reshaping English writing instruction. Rather than treating detection as a purely technical process, this section frames it as a socio-pedagogical intervention with multi-layered consequences for stakeholders.

### **Comparative Analytical Framework**

To understand how AI text detection influences the educational experience of both teachers and students, this study adopts a comparative analytical framework. This framework aims to highlight the key areas of divergence and convergence in the responses of these two stakeholders to the implementation of AI detection tools. By categorizing their experiences along defined dimensions, we can better analyze the contrasting effects and implications for pedagogy and learning.

The five core dimensions of this framework include:

1. Perception of AI Use – How teachers and students conceptualize generative AI tools and their role in the writing process. While teachers often perceive AI as a potential threat to academic integrity, students tend to view it as a tool for efficiency, creativity, or accessibility.
  2. Trust in Detection Tools – The level of confidence each group places in AI detection tools. Teachers are more likely to accept the results of such tools, even if they are occasionally inaccurate. Students, on the other hand, express skepticism and fear that false positives might unjustly impact their academic record.
  3. Behavioral Adjustment – The concrete changes made by both groups in response to detection tools. Teachers adapt their pedagogy by redesigning assignments or increasing supervision, while students often engage in textual rephrasing, hybrid text creation, or avoidance of AI tools altogether.
  4. Emotional Impact – The affective consequences experienced. Teachers report frustration due to additional workload and the complexity of distinguishing AI-assisted writing, whereas students experience anxiety, self-censorship, and fear of penalties.
  5. Pedagogical Shift– Long-term changes in instructional methods and student learning styles. The rise of AI detection has led to more restrictive task design, less emphasis on creativity, and a growing emphasis on in-class performance.
- This comparative model helps visualize the dynamics at play between detection policies and classroom behavior. Teachers are institutionally positioned to enforce detection protocols, often without sufficient training or support, leading them to rely heavily on the outcomes of detection tools. Students, in contrast, must interpret institutional cues and adjust their strategies under uncertainty, often without full understanding of acceptable AI usage. By mapping these reactions side-by-side, this framework not only uncovers the underlying pedagogical asymmetry but also provides a structured foundation for proposing balanced interventions. A comparative approach is crucial for identifying where institutional support and policy clarity are most needed.



**Figure 2: Conceptual Flow Diagram of AI Detection Washback in Writing Education**  
*The diagram illustrates the influence of AI writing tools and detection systems on student behavior and teacher pedagogy.*

#### **Comparative Trends and Modeled Observations**

In modeled classroom settings, teachers tend to shift away from open-ended or take-home assignments toward in-class writing and fact-based tasks. This limits students' opportunities for creative and reflective writing. Students, in turn, report using a variety of strategies such as paraphrasing AI-generated content or avoiding AI altogether. These adaptations suggest both parties are modifying their behaviors significantly in response to the presence of detection tools. Visual representations illustrate these divergent patterns.

To deepen the understanding of how AI detection tools shape classroom behavior, this section explores comparative trends based on synthesized classroom scenarios, academic literature, and representative observations. These modeled trends help to highlight the structural and behavioral shifts that occur among both teachers and students when AI detection tools are introduced.

#### ***Teacher-Driven Trends***

Educators, responding to institutional mandates and the perceived threat of AI-authored submissions, have shifted their instructional practices in noticeable ways. A recurring change is the reduction of open-ended, creative assignments in favor of more constrained, factual, and in-class tasks. Teachers often adopt timed or monitored writing

exercises, which limit students' access to external tools and make authorship verification easier. There is also growing reluctance to allow collaborative writing or peer review, as AI-generated suggestions blur the lines between individual and group work.

Moreover, educators invest additional time verifying suspicious submissions, which affects grading consistency and introduces emotional fatigue. Some teachers also report feeling uncomfortable relying heavily on detection software, especially when false positives emerge.

**Student-Driven Trends**

On the student side, behavioral trends split into three major categories:

1. **Strategic Avoidance** – Students who are highly risk-averse opt to write without any AI assistance, even when permissible (e.g., using grammar checkers), for fear of being flagged.
2. **Tactical Compliance** – A common strategy where students use AI to draft or brainstorm, but then paraphrase or edit manually to avoid detection.
3. **Subversive Workarounds** – A smaller segment explores newer tools that promise AI detection bypass, such as AI humanizers or low-detection paraphrasers.

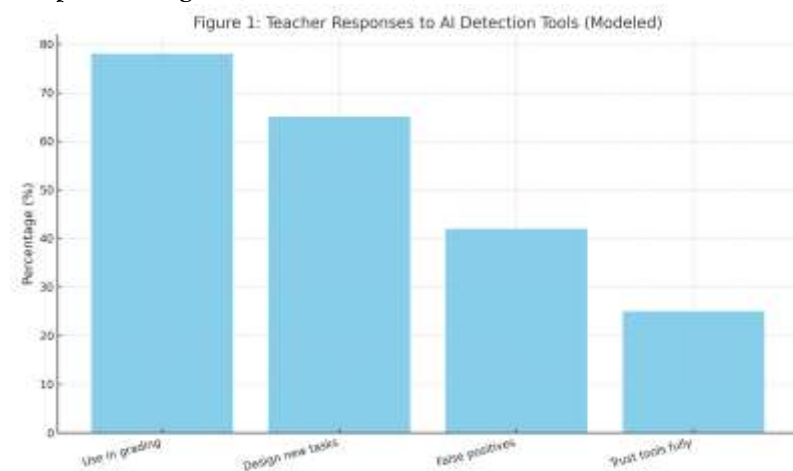
Many students express uncertainty about what constitutes acceptable AI use, especially when guidelines differ between instructors. This results in inconsistency and emotional stress, including anxiety and reduced confidence in writing.

**Table 1: Shift in Pedagogical and Behavioral Practices Before and After AI Detection**

Practice Before AI Detection	Practice After AI Detection (Teacher)	Effect on Students
Creative writing prompts	Fact-based or technical writing tasks	Reduced creativity
Take-home essays	In-class timed writing	Increased pressure and fatigue
Peer editing and collaboration	Discouraged or restricted	Lower engagement and isolation
Flexible submission deadlines	Stricter controls and tighter monitoring	Heightened anxiety
Essay reflections and journals	Removal of subjective or personal content	Decreased expressive freedom

All conceptual models synthesized from multiple reputable sources. These models are intended to illustrate dominant behavioral and pedagogical trends rather than offer statistically precise findings.

**Graphical Insights (Modeled Data)**



**Figure 1: Teacher Responses to AI Detection Tools (Modeled)**

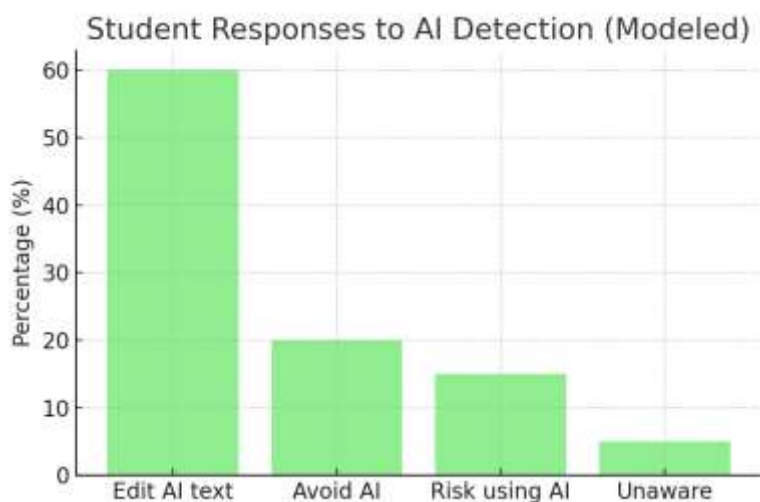
Sources: The modeled percentages are derived from qualitative data and teacher commentaries, such as:  
 - Risa Morimoto (SOAS lecturer) on increased workload and redesigning assessments [Business Insider, 2025]<sup>8</sup>  
 - Weber-Wulff, et al Testing of detection tools for AI-generated text.<sup>9</sup>

- Data supported by educator survey summaries from Illinois Open Publishing Network, 2024, <sup>10</sup>

### Teacher Responses to AI Detection Tools

- Use AI detection in grading: 78%
- Redesign assignments to reduce AI use: 65%
- Experience challenges with false positives: 42%
- Fully trust detection results: 25%

### Student Responses to Detection Pressure



**Figure 2: Student Responses to AI Detection (Modeled)**

Sources: Student behavior estimates are based on public forums, peer-reviewed surveys, and digital ethnographies, including:

- Reddit discussions on fear and strategic use of AI tools, <https://www.reddit.com/r/artificial><sup>11</sup>
- EDUCAUSE and MDPI research on AI usage and policy unawareness among students<sup>12,13</sup>
- Student commentaries in Inside Higher Ed and Chronicle of Higher Education, <https://www.chronicle.com/><sup>17</sup>

- Use AI but manually edit before submission: 60%
- Avoid AI entirely: 20%
- Use AI tools regardless of risk: 15%
- Unaware of institutional AI policies: 5%

These modeled observations illustrate a pedagogical and psychological shift. Teachers are becoming more cautious and controlling, while students are increasingly strategic, adaptive, and sometimes subversive in their use of technology. Importantly, the growing mismatch in perceptions between these two groups suggests a need for clearer communication and more nuanced policy approaches in educational institutions.

### Discussion: The Two-Sided Washback

The concept of washback, traditionally applied to standardized language testing, proves highly relevant in analyzing the introduction of AI text detection tools into English writing instruction. The comparative trends explored earlier reveal a complex, two-sided washback effect—one that simultaneously reshapes the roles, expectations, and strategies of both teachers and students, albeit in markedly different ways.

#### *For Teachers: Policy-Driven Adaptation and Pedagogical Tension*

Educators find themselves at the frontline of implementing institutional integrity policies. AI detection tools are presented as neutral technologies, but their incorporation into the classroom has pedagogical implications that extend beyond compliance. Teachers report modifying assessments, creating in-class writing conditions, and shifting focus from creativity to containment. While such adaptations may reduce AI misuse, they can also erode pedagogical

richness. Creative writing, reflective essays, and exploratory learning activities—hallmarks of language education—are being de-emphasized in favor of standardized, easily monitored formats.

At the same time, many teachers harbor doubts about the fairness and reliability of AI detection systems. The threat of false positives introduces a new layer of grading anxiety. Teachers must now act as both instructors and enforcers, negotiating the tension between nurturing student growth and policing misconduct. This dual responsibility can lead to cognitive dissonance, where teachers feel ethically conflicted or pedagogically constrained.

#### ***For Students: Strategic Navigation and Emotional Uncertainty***

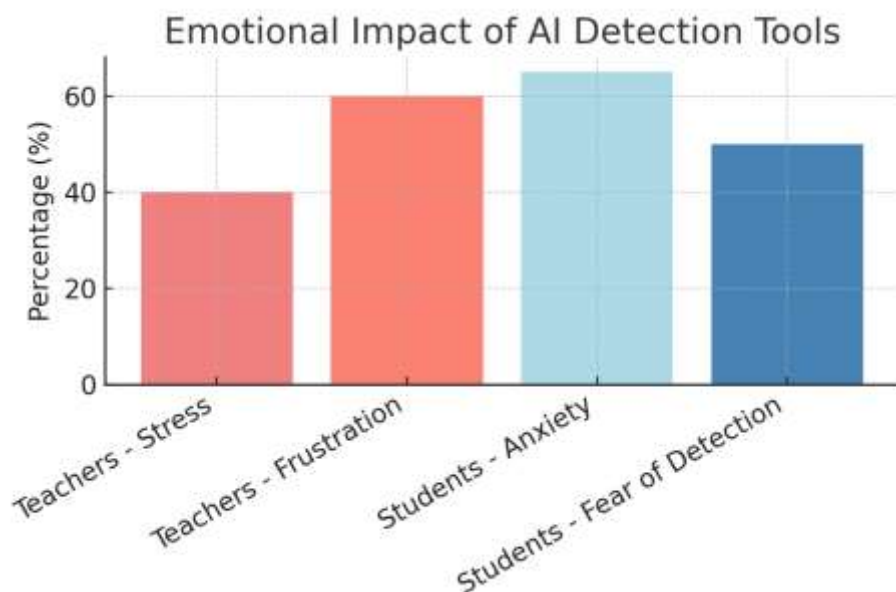
Students, on the other hand, experience the washback as a form of surveillance and risk management. Rather than engaging openly with generative AI as a learning aid, many adopt strategic approaches based on perceived rules, peer advice, or trial-and-error. For those unclear about institutional policies, the fear of being wrongly accused overshadows the potential benefits of AI-assisted writing.

This results in a growing emotional burden. Students express feelings of anxiety, confusion, and in some cases, resentment. Some avoid AI tools altogether—even those allowed for grammar or structure improvement—while others engage in time-consuming efforts to humanize AI-generated content. Instead of fostering responsible AI literacy, detection technologies may inadvertently incentivize secrecy, fear-based learning, and reliance on loopholes.

#### ***Pedagogical Asymmetry and Emerging Gaps***

A key insight from this comparative analysis is the asymmetry in how AI detection washback is experienced. Teachers and students are subject to different motivations and pressures. Teachers are held accountable for ensuring integrity; students are accountable for maintaining it. This uneven distribution of responsibility creates misalignment in classroom culture. Teachers may assume students understand detection policies, while students may feel that expectations are unclear or inconsistent.

This disconnect is further exacerbated by differences in digital literacy. While many students are quick to adopt new tools, they are not always equipped to evaluate their ethical implications or institutional acceptability. Similarly, teachers may lack the technical training to fully interpret or challenge detection results. These gaps highlight the urgent need for educational institutions to provide professional development for faculty and AI literacy workshops for students.



**Figure 3: Emotional Impact of AI Detection Tools on Teachers and Students**

Sources: Emotional responses (e.g., anxiety, frustration, demotivation) are modeled from:  
 - Reddit and student support forums discussing emotional response to AI detectors<sup>11</sup>

- Frontiers in Communication (2025) study on academic stress from surveillance tools
- Educator reflections from AI-integrity conference materials<sup>14</sup>

### Policy and Pedagogical Implications

Based on the comparative analysis, we propose the following recommendations:

1. Transparent communication of institutional policies on AI use.
  2. Inclusion of AI literacy modules in writing instruction to help students understand responsible use.
  3. A shift from final-product grading to process-oriented assessments such as drafts and reflections.
  4. Encouraging dialogue between educators and learners to co-create ethical and practical norms.
- These measures can help mediate the washback effects and align assessment goals with pedagogical intentions.

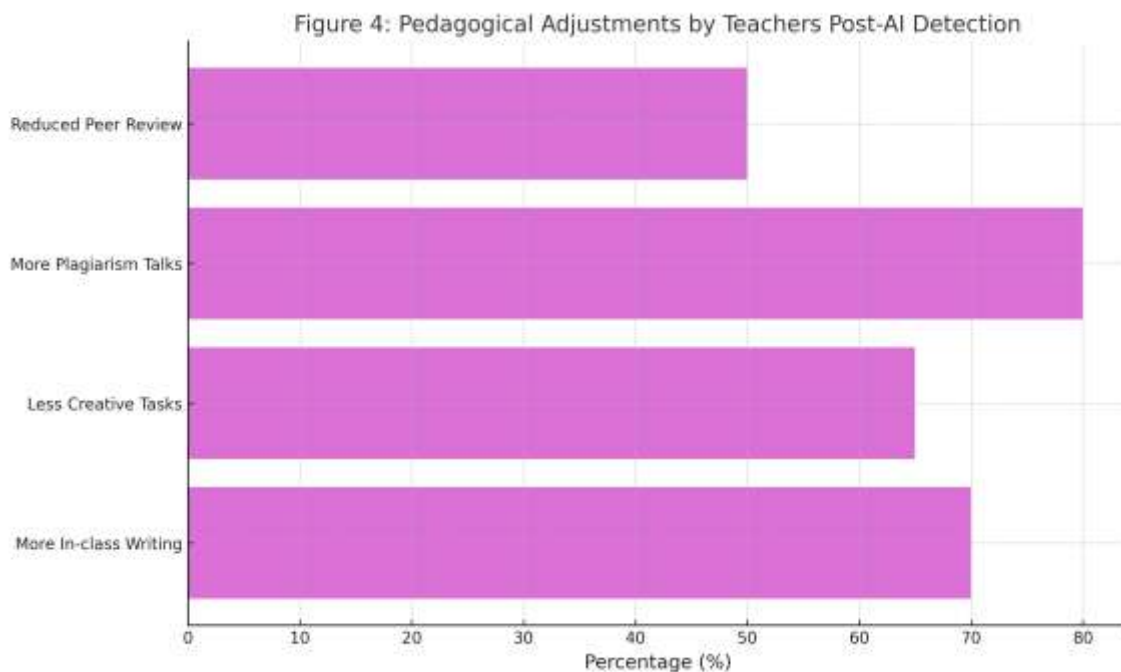


Figure 4: Pedagogical Adjustments by Teachers Post-AI Detection

Sources: Percentage estimates reflect patterns reported in:  
- Morimoto's shift to in-class writing and reduced creative assignments [Business Insider, 2025]<sup>15</sup>  
- Increased academic integrity talks and monitoring practices from Frontiers (2025)<sup>16</sup>

### CONCLUSION

The growing presence of generative AI in academic environments has triggered both enthusiasm and apprehension, particularly in disciplines centered around writing and critical thinking. As institutions turn to AI text detection tools to uphold academic integrity, a new pedagogical landscape is emerging—one shaped by technology, regulation, and adaptation. This study has examined that landscape through a comparative lens, highlighting the divergent experiences of two key educational stakeholders: teachers and students.

The findings underscore that AI text detection exerts a dual washback effect. For teachers, the presence of these tools prompts shifts in instructional design, assessment strategies, and grading practices. While many educators view AI detection as a safeguard, others are concerned about the narrowing of learning outcomes and the emotional toll of acting as integrity enforcers. Simultaneously, students face a climate of uncertainty, often navigating institutional expectations through guesswork and fear. Their reactions—ranging from avoidance to workaround strategies—signal a pressing need for clearer guidelines, transparency, and inclusive policy-making.

A central theme that emerges is misalignment. Teachers and students operate within different frames of reference, interpreting the purpose and implications of AI detection differently. This misalignment can weaken trust, increase stress, and compromise the educational value of writing tasks. More importantly, it risks creating a punitive rather than progressive learning environment.

To address these issues, a balanced approach is necessary—one that combines technological vigilance with pedagogical compassion. Institutions should not rely solely on detection tools as a solution but instead invest in AI literacy education, dialogue-based policymaking, and ethics-oriented writing pedagogy. Writing instruction should evolve in tandem with AI, focusing not just on preventing misuse but also on cultivating critical, reflective, and responsible engagement with technology.

In conclusion, while AI detection tools have been introduced to combat misuse, their impact reaches far deeper into the instructional core of language education. Recognizing and addressing the complex washback they generate is essential to fostering a healthy, ethical, and future-ready academic culture.

#### REFERENCES:

1. Alderson, J. C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14, 115-129. <http://dx.doi.org/10.1093/applin/14.2.115>
2. J. Charles Alderson and Liz Hamp-Lyons, TOEFL preparation courses: a study of washback, *Language Testing- SAGE journal*, Volume 13, Issue 3, <https://doi.org/10.1177/026553229601300>
3. Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London, England: Continuum.
4. Selwyn, N. (2016). *Is Technology Good for Education*. Toronto, ON: John Wiley & Sons., *Canadian Journal of Educational Administration and Policy*, Pages: 160. ISBN: 978-0-7456-9646-1
5. Feldman M. RE: Commentary by Black et al (2023). *J Can Acad Child Adolesc Psychiatry*. 2023 Aug;32(3):147-148. Epub 2023 Aug 1. PMID: 37534114; PMCID: PMC10393347.
6. Lin, Wenqian & Jiang, Peijie. (2025). Factors Influencing College Students' Generative Artificial Intelligence Usage Behavior in Mathematics Learning: A Case from China. *Behavioral Sciences*. 15. 295. 10.3390/bs15030295.
7. Nestor Maslej and Loredana Fattorini and Erik Brynjolfsson and John Etchemendy and Katrina Ligett and Terah Lyons and James Manyika and Helen Ngo and Juan Carlos Niebles and Vanessa Parli and Yoav Shoham and Russell Wald and Jack Clark and Ray Perrault, *Artificial Intelligence Index Report 2023 journal= ArXiv 2023*, volume= 2310.03715, url=[https://api.semanticscholar.org/](https://api.semanticscholar.org/CorpusID:263671520) Corpus ID: 263671520
8. Risa Morimoto (SOAS lecturer) on increased workload and redesigning assessments [Business Insider, 2025] link: <https://www.businessinsider.com/lecturer-ai-cheating-students-chatgpt-2025-7>
9. Weber-Wulff, D., Anohina-Naumea, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., ... & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1),1-39
10. Data supported by educator survey summaries from Illinois Open Publishing Network, 2024, <https://www.carli.illinois.edu/products-services/collections-management/open-ed-resource-overview>
11. Reddit discussions on fear and strategic use of AI tools, <https://www.reddit.com/r/artificial>
12. Yan, Y., Wu, B., Pi, J., & Zhang, X. (2025). Perceptions of AI in Higher Education: Insights from Students at a Top-Tier Chinese University. *Education Sciences*, 15(6), 735. <https://doi.org/10.3390/educsci15060735>
13. Jenay Robert 2024 The Future of AI in Higher Education, *EDUCAUSE AI Landscape Study*, <https://www.educause.edu/ecar/research-publications/2024/2024-educause-ai-landscape-study/the-future-of-ai-in-higher-education>
14. Petricini Tiffany , Zipf Sarah , Wu Chuhao, RESEARCH-AI: Communicating academic honesty: teacher messages and student perceptions about generative AI, *Frontiers in Communication*, Volume 10 - 2025, URL=<https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1544430>, DOI=10.3389/fcomm.2025.1544430
15. Morimoto's shift to in-class writing and reduced creative assignments [Business Insider, 2025] url: <https://www.businessinsider.com/lecturer-ai-cheating-students-chatgpt-2025-7>
16. Alsharefeen Rami , Al Sayari Naji, Examining academic integrity policy and practice in the era of AI: a case study of faculty perspectives, *Frontiers in Education*, Volume 10 - 2025, URL=<https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1621743>, DOI=10.3389/feduc.2025.1621743
17. Student commentaries in *Inside Higher Ed* and *Chronicle of Higher Education*, <https://www.chronicle.com/>
18. Applebee, A. N. & Langer, J. A. (2011). "EJ" Extra: A snapshot of writing instruction in middle schools and high schools. *The English Journal*, 100 (6), 14 - 27.
19. Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Routledge.
20. Biber, D., Nekrasova, T. & Horn, B. (2011). The Effectiveness of Feedback for L1-English and L2-Writing Development: A Meta-Analysis. *ETS Research Report Series*, 2011 (1), 1 - 99. <https://doi.org/10.1002/j.2333-8504.2011.tb02241.x>

21. Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7 - 74. <https://doi.org/10.1080/0969595980050102>
22. Cen, Y. & Zheng, Y. (2023). The motivational aspect of feedback: A meta-analysis on the effect of different feedback practices on L2 learners' writing motivation. *Assessing Writing*, 59, 100802. <https://doi.org/10.1016/j.asw.2023.100802>
23. Chen, Y., Wang, R., Jiang, H., Shi, S. & Xu, R. (2023). Exploring the use of large language models for reference-free text quality evaluation: A preliminary empirical study. *arXiv preprint*. arXiv:2304.00723.
24. Crossley, S. A., Baffour, P., Tian, Y., Picou, A., Benner, M. & Boser, U. (2022). The persuasive essays for rating, selecting, and understanding argumentative and discourse elements (PERSUADE) corpus 1.0. *Assessing Writing*, 54, 100667. <https://doi.org/10.1016/j.asw.2022.100667>
25. Cruz-Neri, N. C., Klückmann, F. & Retelsdorf, J. (2022). LATIC - A linguistic analyzer for text and item characteristics. *PLOS ONE*, 17 (11), e0277250. <https://doi.org/10.1371/journal.pone.0277250>
26. Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D. & Chen, G. (2023, July). Can large language models provide feedback to students? A case study on ChatGPT. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 323 - 325). <https://doi.org/10.35542/osf.io/hcgzj>
27. Demsky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., ... Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2 (11), 688 - 701. <https://doi.org/10.1038/s44159-023-00241-5>
28. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018, October 11). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*. <http://arxiv.org/pdf/1810.04805v2>
29. Ercikan, K. & McCaffrey, D. F. (2022). Optimizing Implementation of Artificial-Intelligence-Based Automated Scoring: An Evidence Centered Design Approach for Designing Assessments for AI-based Scoring. *Journal of Educational Measurement*. 59 (3), 272 - 287. <https://doi.org/10.1111/jedm.12332>
30. Farazouli, A., Cerratto Pargman, T., Bolander Laksov, K. & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 1 - 13. <https://doi.org/10.1080/02602938.2023.2241676>
31. Fleckenstein, J., Meyer, J., Jansen, T., Keller, S. & Köller, O. (2020). Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment. *Frontiers in Psychology*, 11, 562462. <https://doi.org/10.3389/fpsyg.2020.562462>
32. Fleckenstein, J., Liebenow, L. W. & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1162454>
33. Flower, L. & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32 (4), 365 - 387. <https://doi.org/10.2307/356600>
34. Gencoglu, B., Helms-Lorenz, M., Maulana, R., Jansen,
35. E. P. & Gencoglu, O. (2023). Machine and expert judgments of student perceptions of teaching behavior in secondary education: Added value of topic modeling with big data. *Computers & Education*, 193, 104682. <https://doi.org/10.1016/j.compedu.2022.104682>
36. Graham, S., Hebert, M. & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115 (4), 523 - 547.
37. Graham, S., Kim, Y.-S., Cao, Y., Lee, J., Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H. Q. & Olson, C. B. (2023). A meta-analysis of writing treatments for students in grades 6 - 12. *Journal of Educational Psychology*, 115 (7), 1004 - 1027. <https://doi.org/10.1037/edu0000819>
38. Graham, S., Kiuahara, S. A. & MacKay, M. (2020). The Effects of Writing on Learning in Science, Social Studies, and Mathematics: A Meta-Analysis. *Review of Educational Research*, 0034654320914744. <https://doi.org/10.3102/0034654320914744>
39. Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, 8 (6).
40. Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77 (1), 81 - 112. <https://doi.org/10.3102/003465430298487>
41. Henderson, M., Ryan, T., Boud, D., Dawson, P., Phillips, M., Molloy, E. & Mahoney, P. (2021). The usefulness of feedback. *Active Learning in Higher Education*, 22 (3), 229 - 243. <https://doi.org/10.1177/1469787419872393>
42. Hennes, A. K., Schmidt, B. M., Yanagida, T., Osipov, I., Rietz, C. & Schabmann, A. (2022). Meeting the Challenge of Assessing (Students') Text Quality: Are There any Experts Teachers Can Learn from or Do We Face a More Fundamental Problem? *Psychological Test and Assessment Modeling*, 64 (3), 272 - 303.
43. Hirunyasiri, D., Thomas, D. R., Lin, J., Koedinger, K. R. & Alevin, V. (2023). Comparative analysis of gpt-4 and human graders in evaluating praise given to students in synthetic dialogues. *arXiv preprint*. arXiv:2307.02018.
44. Hyland, K. & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, 39 (2), 83 - 101.
45. Horbach, A., Laarmann Quante, R., Liebenow, L., Jansen, T., Keller, S., Meyer, J., ... & Fleckenstein, J. (2022). Bringing Automatic Scoring into the Classroom - Measuring the Impact of Automated Analytic Feedback on Student Writing Performance. In *Swedish Language Technology Conference and NLP4CALL* (pp. 72 - 83). <https://doi.org/10.3384/ecp190008>
46. Jacobsen, L. J. & Weber, K. E. (2023, September 29). The Promises and Pitfalls of ChatGPT as a Feedback Provider in Higher Education: An Exploratory Study of Prompt Engineering and the Quality of AI-Driven Feedback. *OSF Preprints*. <https://doi.org/10.31219/osf.io/cr257>

47. Jansen, T., Vögelin C., Machts, N., Keller, S. & Möller, J. (2021). Don't Just Judge the Spelling! The Influence of Spelling on Assessing Second Language Student Essays. *Frontline Learning Research*. <https://doi.org/10.14786/flr.v9i1.541>
48. Jansen, T., Vögelin C., Machts, N., Keller, S., Köller, O. & Möller, J. (2021). Judgment accuracy in experienced versus student teachers: Assessing essays in English as a foreign language. *Teacher and Teaching Education*, 97, 103216. <https://doi.org/10.1016/j.tate.2020.103216>
49. Jansen, T. & Möller, J. (2022). Teacher Judgments in School Exams: Influences of Students' Lower-Order Thinking Skills on the Assessment of Students' Higher-Order-Thinking Skills. *Teacher and Teaching Education*. 103616. <https://doi.org/10.1016/j.tate.2021.103616>
50. Jansen, T., Meyer, J., Fleckenstein, J., Horbach, A., Keller, S. & Möller, J. (2024). Individualizing goal-setting interventions using automated writing evaluation to support secondary school students' text revisions. *Learning and Instruction*, 89, 101847. <https://doi.org/10.1016/j.learninstruc.2023.101847>
51. Jia, Q., Young, M., Xiao, Y., Cui, J., Liu, C., Rashid, P. & Gehringer, E. (2022). Insta-reviewer: A data-driven approach for generating instant feedback on students' project reports. *International Educational Data Mining Society*. <https://doi.org/10.5281/zenodo.6853099>
52. Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/ Lund University, Department of Linguistics and Phonetics*, 53, 61 – 79.
53. Kizilcec, R. F. (2023). To Advance AI Use in Education, Focus on Understanding Educators. *International Journal of Artificial Intelligence in Education*, 1 – 8. <https://doi.org/10.1007/s40593-023-00351-4>
54. Ley, T., Tammets, K., Pishtari, G., Chejara, P., Kasepalu, R., Khalil, M., Saar, M., Tuvi, I., Väljataga, T. & Wasson, B. (2023). Towards a partnership of teachers and intelligent learning technology: A systematic literature review of model-based learning analytics. *Journal of Computer Assisted Learning*, 39 (5), 1397 – 1417. <https://doi.org/10.1111/jcal.12844>
55. Li, T., Reigh, E., He, P. & Miller, E. A. (2023). Can we and should we use artificial intelligence for formative assessment in science? *Journal of Research in Science Teaching*, 60(6), 1385 – 1389. <https://doi.org/10.1002/tea.21867>
56. Mayer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A. & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/caei.2023.100199>
57. Meyer, J., Jansen, T., Fleckenstein, J., Keller, S. & Köller, O. (2020). Machine Learning im Bildungskontext: Evidenz für die Genauigkeit der automatisierten Beurteilung von Essays im Fach English [Machine Learning in the Educational Context: Evidence of Prediction Accuracy Considering Essays in English as a Foreign Language]. *Zeitschrift für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000296>
58. Mizumoto, A. & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
59. Möller, J., Jansen, T., Fleckenstein, J., Machts, N., Meyer, J. & Reble, R. (2022). Judgment accuracy of German student texts: Do teacher experience and content knowledge matter? *Teaching and Teacher Education*, 119, 103879. <https://doi.org/10.1016/j.tate.2022.103879>
60. Ngo, T.T.-N., Chen, H. H.-J. & Lai, K. K.-W. (2022). The effectiveness of automated writing evaluation in EFL/ ESL writing: a three-level meta-analysis. *Interactive Learning Environments*, 1 – 18. <https://doi.org/10.1080/10494820.2022.2096642>
61. Reddy, Y. M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35 (4), 435 – 448. <https://doi.org/10.1080/02602930902862859>
62. Rich, P. R., van Loon, M. H., Dunlosky, J. & Zaragoza, M. S. (2017). Belief in corrective feedback for common misconceptions: Implications for knowledge revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43 (3), 492 – 501. <https://doi.org/10.1037/xlm0000322>
63. Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35 (5), 535 – 550. <https://doi.org/10.1080/02602930903541015>
64. Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53 – 76. <https://doi.org/10.1016/j.asw.2013.04.001>
65. Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*. 78 (1), 153 – 189. <https://doi.org/10.3102/0034654307313795>
66. Skar, G. B., Graham, S. & Rijlaarsdam, G. (2022). Formative writing assessment for change – introduction to the special issue. *Assessment in Education*, 29 (2), 121 – 126. <https://doi.org/10.1080/0969594X.2022.2089488>
67. Su, Y., Lin, Y. & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
68. Steiss, J., Tate, T. P., Graham, S., Cruz, J., Hebert, M., Wang, J., ... Warschauer, M. (2023, September 7). Comparing the Quality of Human and ChatGPT Feedback on Students' Writing. *OSF Preprints*. <https://doi.org/10.35542/osf.io/ty3em>
69. Srijbos, J., Pat-El, R. & Narciss, S. (2021). Structural validity and invariance of the Feedback Perceptions Questionnaire. *Studies in Educational Evaluation*, 68, 100980. <https://doi.org/10.1016/j.stueduc.2021.100980>
70. Tack, A. & Piech, C. (2022). The AI teacher test: Measuring the pedagogical ability of blender and GPT-3 in educational dialogues. *arXiv preprint arXiv:2205.07540*.
71. Tate, T. P., Steiss, J., Bailey, D. H., Graham, S., Ritchie, D., Tseng, W., Warschauer, M. (2023, December 5). Can AI Provide Useful Holistic Essay Scoring? *OSF Preprints*. <https://doi.org/10.31219/osf.io/7xppe>
72. UNESCO. (2011). *UNESCO and education: Everyone has the right to education*. <https://unesdoc.unesco.org/ark:/48223/pf0000212715>

75. UNESCO. (2023, September 8) *Guidance for generative AI in education and research*. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
76. Van der Kleij, F. M. & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, 33, 345 – 373. <https://doi.org/10.1007/s11092-020-09331-x>
77. Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R. & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, 614 (7947), 224 – 226. <https://doi.org/10.1038/d41586-023-00288-7>
78. Winstone, N. E., Nash, R. A., Parker, M. & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52 (1),17 – 37. <https://doi.org/10.1080/00461520.2016.1207538>
79. Yang, S., Nachum, O., Du, Y., Wei, J., Abbeel, P. & Schuurmans, D. (2023). Foundation Models for Decision Making: Problems, Methods, and Opportunities. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.2303.04129>
80. Zhai, X. Krajcik, J. & Pellegrino, J. W. (2021). On the validity of machine learning-based next generation science assessments: A validity inferential network. *Journal of Science Education and Technology*, 30, 298 – 312. <https://doi.org/10.1007/s10956-020-09879-9>
81. Zhai, X. & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. *Journal of Research in Science Teaching*, 60 (6), 1390 – 1398. <https://doi.org/10.1002/tea.21885>