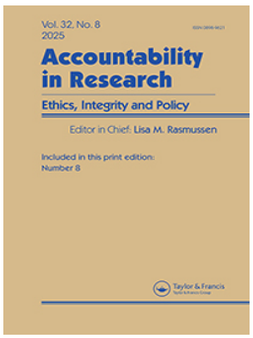

Taylor & Francis
Taylor & Francis Group

Accountability in Research
Ethics, Integrity and Policy

ISSN: 0898-9621 (Print) 1545-5815 (Online) Journal homepage: www.tandfonline.com/journals/gacr20

Opening Pandora's box: Developing reviewer rhetorical sensitivity through retracted articles

Baraa Khuder

To cite this article: Baraa Khuder (25 Dec 2025): Opening Pandora's box: Developing reviewer rhetorical sensitivity through retracted articles, Accountability in Research, DOI: 10.1080/08989621.2025.2607681

To link to this article: https://doi.org/10.1080/08989621.2025.2607681

Published online: 25 Dec 2025.

ACCOUNTABILITY IN RESEARCH
https://doi.org/10.1080/08989621.2025.2607681

Taylor & Francis
Taylor & Francis Group

Opening Pandora's box: Developing reviewer rhetorical sensitivity through retracted articles

Baraa Khuder

Department of Communication and Learning in Science, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT

Retractions issued for misconduct offer a unique window into how questionable research is rhetorically constructed and made to appear credible. This study investigates how engaging with retracted articles can serve as a pedagogical tool for reviewer training, with particular attention to the rhetorical mechanisms through which unreliability is performed. Twenty STEM doctoral researchers analyzed self-selected retracted papers using guided critical-reading questions to identify problematic rhetorical features. Across the analyses, five recurring issues emerged: intertextual falsification, methodological opacity, rhetorical inconsistency, rhetorical overstatement, and terminological distortion. The findings indicate that this approach has the potential to raise doctoral students' rhetorical sensitivity by enabling them to detect subtle markers of unreliability and to adopt a more evaluative rhetorical stance toward scholarly texts. Retracted articles thus can provide an authentic pedagogical resource for developing reviewer rhetorical sensitivity within doctoral education.

ARTICLE HISTORY
Received 25 August 2025
Accepted 17 December 2025.

KEYWORDS
Retractions; rhetorical features; scholarly publishing; Reviewer training; doctoral education

Introduction

Retractions have become a defining feature of contemporary science. Their number has increased dramatically in recent decades, rising from a few hundred annually in the early 2000s to over 10,000 in 2023, which reflects persistent vulnerabilities in the research system (Grieneisen and Zhang 2012; Petrou 2024). For journals and institutions, retractions due to misconduct function as visible markers of failure: they might signal fabrication, falsification, duplication, plagiarism, or irreproducible results (Fang, Steen, and Casadevall, 2012). Accordingly, much of the literature on retractions focuses on classifying reasons for withdrawal, tracking trends and frequencies, and identifying bibliometric patterns (Fang, Steen, and Casadevall 2012; Hesselmann et al. 2016). Some studies have gone further, documenting specific forms of malpractice such as image manipulation (Bik, Casadevall, and Fang 2016), data falsification (Hu et al. 2025), or the industrial production of fraudulent manuscripts by paper mills (Rivera and Teixeira da Silva 2021). Additionally, studies highlight structural and systemic factors, including research incentives and recognition systems, that create fertile conditions for unreliable and paper-mill-generated manuscripts to proliferate (Armond et al., 2024). These studies have greatly improved our understanding of how misconduct and error manifest within the research system. However, far less is known about how unreliable papers, particularly those later retracted for misconduct, manage to appear credible enough to pass peer review in the first place.

Understandably, there have been increasing calls to strengthen peer review and to develop innovative approaches to reviewer training, since traditional peer review alone has not yielded significant improvements in review quality (Hesselberg et al. 2023) as one potential measure for guarding research accountability. While many initiatives emphasize procedural ethics and methodological competence, they often neglect what is immediately visible to reviewers, namely, how the *writing itself* performs credibility. In practice, research articles do far more than report findings; they enact scientific authority through rhetorical conventions. As decades of rhetorical scholarship have shown, scientific writing is never purely descriptive but a persuasive practice through which authors construct authority and display disciplinary knowledge by mobilizing shared rhetorical conventions of the research article (Bazerman 1988; Khuder and Petric, 2026;

CONTACT Baraa Khuder khuder@chalmers.se Department of Communication and Learning in Science, Chalmers University of Technology, Hörsalsvägen 2, Gothenburg SE-412 96, Sweden

Hyland 2005; Swales and Feak 2012). Doctoral researchers are trained to use these conventions (e.g., to establish territory, claim significance, and project competence) but rarely to examine how such features contribute to performing credibility. This gap extends to reviewer training, where attention typically centers on methodological soundness rather than on the rhetorical signals that accompany it.

In the sociology of science, Gilbert and Mulkay's (1984) work on "opening Pandora's box" foregrounded scientific discourse itself as an object of analysis, demonstrating that scientific writing is socially situated and that its linguistic features shift according to communicative situation, purpose, and audience, rather than being determined solely by disciplinary conventions. Subsequent work in applied linguistics (Hyland, 2005; Swales and Feak 2012) has greatly expanded this line of inquiry, showing that patterns of stance, evaluation, and responsibility in scientific writing vary across contexts and conditions of production. In this sense, much is now known about how scientific discourse operates in general. What remains less well understood is how these linguistic and rhetorical features are reconfigured in moments of disruption. Retractions - rather than scientific discourse in general - thus constitute a contemporary "Pandora's box", raising the question of how scientific writing changes when the credibility of the research record itself is called into question.

A line of quantitative work has examined how unreliability and fraud are reflected in the linguistic surface of research articles. Markowitz and Hancock (2016) compared 253 papers retracted for data fraud with matched unretracted articles and papers retracted for non-fraud reasons, showing that fraudulent papers exhibited higher levels of "linguistic obfuscation"; they were less readable, contained more jargon and more abstract language, and relied on denser referencing than comparison texts. In a related study of biochemistry articles, Dehdarirad and Schirone (2023) found that retracted papers (due to misconduct) were less likely than non-retracted papers to use positive terms and certainty language when presenting findings, and were less likely to express probabilistic certainty. Research on "tortured phrases" similarly documents distorted technical terminology and anomalous phrasings often linked to paper mills or automated text generation that superficially mimic disciplinary language while lacking precision (Cabanac, Labbé, and Magazinov 2021; Teixeira da Silva 2023). Other research highlights patterns of exaggerated novelty and inflated certainty as problematic issues that require more attention from the reviewers (Keserlioglu, Kilicoglu, and Ter Riet 2019). It should be noted that these features are not proof of misconduct, but rhetorical warning signs; red flags that warrant closer scrutiny.

Building on emerging insights into the rhetorical dimensions of unreliable research, this study develops a reviewer-training approach that integrates rhetorical sensitivity into doctoral education. Using retracted papers as pedagogical tools, the study examines how twenty STEM (Science, Technology, Engineering, and Mathematics) doctoral researchers developed awareness of textual and rhetorical cues of questionable research practice and reflected critically on how scientific credibility is constructed through writing.

Methods

Research context and participants

The study was conducted within a six-week doctoral-level *Writing for Publication* course offered at a technical university. The participants in this study were first- to third-year STEM doctoral researchers. Their prior familiarity with academic publishing varied: although a minority had coauthored conference papers or journal articles, the majority had not yet submitted a manuscript for publication and none reported having formal experience as peer reviewers.

The course introduced students to the conventions, genres, and ethical dimensions of scientific publishing. The course also included a module on critical reading and reviewing of scholarly literature. Fifty-five students participated in this module. Students were given the option to work with either a retracted or a non-retracted article. Twenty students chose to analyze a retracted article, and only their data were included in this study.

The dataset consists of the students' written reflections, which were part of their regular coursework. No additional assignments, grades, or incentives were involved. At the start of the module, all students received an information sheet and consent form explaining that anonymized reflections would be used for pedagogical research on reviewer-training design. Participation was entirely

voluntary and had no influence on course grades or academic progression. All twenty students signed the consent form.

To further minimize potential power imbalances, the researcher was not the course examiner and played no role in student assessment. The project was reviewed internally and conducted in accordance with institutional guidelines (Chalmers University of Technology) for educational research involving anonymized student work. Because the data was classified as "low risk" and consisted of reflective texts rather than personal or evaluation records, formal ethics-board approval was not required under university policy. Anonymity was maintained throughout, and no identifying information appears in the analyses or publications.

Pedagogical framework

The pedagogical design was adapted from Clark's (2007) Critical Reading Model (see Appendix 1), which emphasizes moving systematically from comprehension of surface-level features to deeper evaluation of argumentation, evidence, and rhetorical strategies. The model encourages students to interrogate how a text persuades as much as what it claims, making it suitable for analyzing articles that were later retracted. Students were introduced to the adapted model through a four-hour workshop session, where the instructor illustrated each stage of analysis.

To establish a shared starting point, the class collectively analyzed the 1998 article by Wakefield et al., which linked the MMR vaccine to autism and was later fully retracted. The case was selected because of its accessibility across disciplines (well-known, simple structure, clear problems). Working in small groups, students applied the analytical model to the Wakefield paper, identifying weaknesses such as selective reporting of data, overgeneralization of claims, and reliance on rhetorical authority. A plenary discussion followed, where students reflected on how the article's apparent legitimacy, published in a leading medical journal, contrasted with its substantive flaws. This exercise both familiarized students with the analytical framework and illustrated how retracted articles can expose systemic vulnerabilities in scholarly publishing practices.

Following the collective exercise, students independently selected a retracted article relevant to their own research interests. They chose a mix of high-profile retracted articles (e.g., published in *Nature)* and less prestigious articles they encountered through social media (e.g., via LinkedIn). It is important to note that the students were provided with links to databases such as *Retraction Watch* and other repositories that host retracted publications.

Each student conducted an individual critical analysis, informed by the in-class activity, and produced a written reflection detailing both their findings and their reasoning processes. Although the selected articles varied in publication genre and disciplinary domain, all had been retracted due to concerns relating to misconduct, including methodological weaknesses, misrepresented evidence, and unverifiable or incorrect results.

The guided questions used in this task are provided in Appendix 1. After completing their analyses, students were also asked to indicate which questions they found most useful for detecting problems within the articles. The analysis presented in this paper draws on these final reflections, which were around 300 words each.

Earlier in the course, all participating students completed a workshop on critical Generative AI (GenAI) literacy (see Khuder, 2025; Ou et al., 2024), where they were introduced to responsible and critical uses of GenAI tools in academic writing and analysis. The workshop addressed potential benefits (e.g., language editing, idea exploration) and risks (e.g., hallucinated references, uncritical paraphrasing, overreliance on GenAI-generated summaries), and clarified what constituted acceptable and unacceptable uses of GenAI within the course. For the reflection task analyzed in this study, students were encouraged to provide personal, textually grounded reflections and were reminded that they were responsible for the accuracy of all content and references. Any GenAI-based assistance, if used, was expected to be checked carefully and integrated into their own analytical reasoning. GenAI tools were not an object of analysis in this study; instead, the pedagogical focus remained on how students themselves articulated and evaluated rhetorical features of retracted papers. The author cross-checked and verified the evidence cited in students' reflections against the retracted articles they analyzed.

Data collection and analysis

The study adopts an interpretivist qualitative framework, aiming to explore how doctoral researchers make sense of rhetorical cues of unreliability in research texts. Rather than seeking to quantify the prevalence of specific features, the analysis focuses on how students described and evaluated these features in their own words. The corpus for this study consisted of twenty student reflection notes analyzing self-selected retracted papers. This self-selection likely favored retracted articles that appeared rhetorically unusual or controversial, which is appropriate for the study's pedagogical aims but not intended to represent all retracted literature.

Each reflection was anonymized and de-identified before analysis. Thematic analysis (Braun and Clarke 2006) was employed to synthesize the data. Coding was conducted inductively, with initial codes developed directly from the student texts. Through iterative comparison and discussion, these codes were refined into broader themes that captured recurring rhetorical patterns across cases.

The analysis was led by the author. To enhance the reliability of the coding, a second researcher independently coded a subset of reflections (25%) and compared coding decisions with the author. Inter-coder agreement was high, and discrepancies were resolved through discussion, resulting in a refined codebook that was applied to the remaining data. Five overarching themes emerged from this process (Table 1 below).

Findings

The findings are presented thematically, with each theme highlighting rhetorical features that students identified in retracted articles, and the evaluative capacities and reviewer-like reflexivity that developed through the exercise.

Intertextual falsification

All twenty students identified evidence of intertextual falsification, which manifested in three main forms across the analyzed papers: (1) misrepresentation of previous studies, (2) omission of relevant findings or counterclaims from cited sources, and (3) hallucinated references, or citations to studies that did not exist. Students repeatedly observed that several retracted papers cited reputable sources to legitimize claims those sources never made. As one student remarked, "They use IPCC data to say there is no trend. It's the opposite of what the report says … In reality, the cited IPCC chapter explicitly documents statistically significant increases in extreme temperature events and precipitation anomalies" (S12). This example shows how scientific authority can be strategically co-opted to lend rhetorical legitimacy to claims that lack empirical support. The student's ability to recognize this misrepresentation was rather noticeable, and it stemmed from familiarity with the cited literature, in order to be able to spot the manipulation. This highlights the importance of reviewers who are deeply familiar with the relevant literature, not only with methods.

Beyond misrepresentation, students also identified subtler forms of distortion; particularly the omission of relevant counter-evidence. Some retracted papers disregarded studies that challenged their claims or selectively cited portions that reinforced their argument while ignoring contradictory findings. One student

Table 1. Overview of the five identified themes and their definitions.

Theme	Description
Rhetorical Inconsistency	Mismatches between what is claimed and what is actually shown; for instance, large or definitive conclusions drawn from limited or simple experiments, or contradictions between text and figures.
Rhetorical Overstatement	The use of an overly confident tone, "too clean" visuals, and definitive claims that leave no space for uncertainty.
Intertextual Falsification	Misrepresentation, selective omission, or fabrication of references.
Methodological Opacity	Vague, contradictory, or incomplete methodological descriptions that prevent verification or replication.
Terminological Distortion	Excessive or incorrect use of jargon, buzzwords, or recycled technical language.

noted: "No reference was made to the 'Absence of Near-Ambient Superconductivity' paper, which was already publicly available before publication and explicitly contradicted the core claim" (S7). As S7 suggested, the complete exclusion of critical or alternative perspectives undermines epistemic transparency. As one student reflected, reviewers tend to "focus on methods", but "a closer look at the literature review might reveal the beginnings of error" (S17).

Another student summarized the pattern succinctly: "The paper follows a typical academic structure but unexpectedly challenges the link between global warming and extreme weather. It selectively uses data and does not address opposing views, assuming agreement from climate skeptics" (S5). Such cherry-picking practices, as the students noted, are well-known across research but here appear at the heart of the literature review itself. This demonstrates how intertextual credibility can be constructed as much through omission as through inclusion.

In rarer cases, particularly among lower-profile retractions, intertextual falsification escalated into outright fabrication. Several retracted papers included unverifiable or entirely fabricated references used to imply "consistent external validation across populations" (S2). One student remarked, "Checking for fake references made the task of spotting the fact that this is a retracted paper easy" (S18). Another added, "I think we will see more of these now with AI tools being overused for generating literature reviews and reference lists" (S8). This observation aligns with growing concerns about the use of GenAI in producing hallucinated references. Despite increased awareness, fabricated or hallucinated references continue to appear in papers published after 2023, coinciding with the rapid integration of GenAI tools in academic workflows (Walters and Wilder 2023).

Methodological opacity

The second most prevalent pattern identified across the data was methodological opacity: the appearance of methodological clarity that conceals omission, manipulation, or selective disclosure. Students frequently noted that while the methods section of retracted papers appeared conventional and detailed, they often lacked information necessary for verification or replication.

Several students observed that the problem was not simply missing data but the deliberate shaping of methodological description to fit a desired outcome. S3 noted, "the article was focused on the results and less on explaining the experiment and what might have gone wrong". Given the lack of consensus in the literature regarding how methods should be reported and how much detail is required, such omissions are not always immediately visible as problematic. Instead, they reflect a prioritization of space for findings over transparency about experimental design, limitations, and uncertainty. These selective reporting practices illustrate how rhetorical compression in methods sections can be used to sustain the appearance of rigor while obscuring methodological fragility.

Other papers were less manipulative but equally opaque in what they left unsaid. S6 observed, "While they provide simulations which they compare their experiments to, little information is given on how these simulations are made". The student is highlighting a problem that is rather evident in research: how much information should be provided about experiments. This can be a limitation of the genre the article is written in as was explained by S3: "While this paper is a letter which restricts the length of the article, there is no theoretical background to the work done besides the references given in the introduction and direction to other material". Short-format publication genres often prioritize brevity, which can be problematic when the article is supposed to be high-impact.

In other cases, omission took subtler forms. One student noted that the questions used for their analysis helped reveal "lacking aspects such as measurement error, reliability, [and] thermal effect" (S5). Here, methodological sufficiency was rhetorically performed rather than demonstrated. The methods projected thoroughness through familiar markers, such as technical terms, equipment names, or validation claims, without providing verifiable detail.

Rhetorical overstatement

Across the corpus, retracted papers often projected an air of absolute confidence, constructing claims so strong and definitive that they left no rhetorical space for uncertainty or error. One student observed,

"Already from the title we are set both an expectation of what to find but also how the text will probably be organized. The use of the word 'quantized' is rather strong and sets the stage for something unquestionable" (S3). He added, "This is repeated in the article where they say they have found 'excellent results.' The results were so clean, like so very clean, that the field reacted in maze" (S3).

Short-format or high-impact publication genres often prioritize certainty, creating what another student described as "no space for doubt" (S10). The pressure to produce concise, attention-grabbing results encourages what might be called rhetorical compression – the tightening of narrative and language to emphasize clarity and impact at the expense of nuance. In such contexts, essential procedural details and expressions of uncertainty are often omitted to maintain a coherent, assertive storyline.

Students recognized that this absolute phrasing, reinforced through emphatic verbs, evaluative adjectives, and visually "perfect" graphs, functioned rhetorically to transform tentative findings into discoveries. As one summarized, "These papers sound like discoveries, not investigations" (S3). The issue, then, was not mere exaggeration but a distortio of epistemic stance: rather than presenting evidence within the bounds of uncertainty, these papers performed certainty as proof.

Rhetorical inconsistency

Students repeatedly noted that many retracted papers appeared rhetorically coherent at first glance: employing the expected IMRaD structure, figures, and statistical tables, yet contained internal contradictions between what was claimed and what was actually shown. One student summarized this tension: "There are many results which are discussed from a seemingly simple experiment" (S2). This inconsistency operates at a higher level of rhetoric, where claims seem to demand a scale or complexity of experimentation that the described methods cannot support. As the student's comment suggests, not only should methods be scrutinized for producing specific results, but the scope and strength of the claims themselves must be assessed in relation to the actual scale of the research design.

Several students commented that the illusion of complexity was created through elaborate graphs and dense result sections, even when the underlying design lacked proportional rigor. One explained, "It is a letter, so there is not much space" (S3), suggesting that the brevity of certain publication genres may encourage rhetorical overreach. Other students echoed this sentiment, attributing inconsistencies to genre constraints and the pressure to present major findings in short, high-impact formats. This raises a broader question: can genuinely groundbreaking scientific discoveries be adequately communicated in genres that limit space for methodological or interpretive nuance?

Rhetorical inconsistency also appeared in the mismatch between framing and execution. One student described a case where "the article claimed randomization among participants while assigning the same treatment to entire groups, contradicting its own methodological description" (S15). Another highlighted a study that presented itself as promoting sustainable materials, yet "the proposed alternatives were still quite toxic, and the stability was not extensively tested". In both instances, the rhetorical framing, whether of rigor or responsibility, conflicted with the underlying practices. As S11 summarized, these were "texts that look scientific but rely on the reader forgetting small details".

Terminological distortion

The final recurring issue identified in the corpus was terminological distortion; the strategic use of technical language to simulate expertise. This took two main forms: excessive reliance on jargon and the use of non-disciplinary or incorrect terminology. Students found that while some degree of specialization is expected in scientific writing, the overuse or misuse of disciplinary terminology often functioned rhetorically to obscure weak argumentation or missing evidence.

Terminological misuse was particularly common in lower-profile publications, which several students attributed to inexperience or lack of editorial oversight. As one explained, some authors seemed to use "terminological camouflage" (S1), where complex phrasing substituted for genuine explanation. One student remarked: "It read like a catalogue of machine learning terms (convolutional networks, SVMs, regression layers) but none of it was explained. The language looked impressive but was empty" (S9).

Several students also observed how familiar keywords functioned as rhetorical shortcuts, signaling innovation without demonstrating it. As one reflection put it, "The buzzwords were there to signal innovation, but the paper never demonstrated what was novel" (S2). Another pattern was terminological recycling; the repeated use of stock phrases such as "robust methodology" or "rigorous design" with little elaboration. S17 noted, "The same phrases kept coming back, but the authors never explained what those meant in practice. It was like a script". Some students described this effect as a kind of linguistic misdirection; terminology that sounded correct but lacked precision or coherence. As S20 observed, "Because the terminology sounded right, one can feel less skeptical, only to realize later the terms were being used vaguely".

Discussion

This study explored how retracted articles can be used pedagogically to train doctoral students for peer review. The goal is to help students identify how unreliability is constructed within the text itself by developing rhetorical sensitivity through guided questions. Engaging with retracted papers was shown to be effective for this purpose; students reported increased awareness of how rhetorical cues perform credibility. Across the twenty analyzed reflections, there were five recurring rhetorical patterns: rhetorical inconsistency, rhetorical overstatement, intertextual falsification, methodological opacity, and terminological distortion.

Intertextual falsification (i.e., the misrepresentation, omission, or fabrication of references) was the most prevalent feature. While previous work has examined how retracted papers continue to be cited after removal (Teixeira da Silva & Bornemann-Cimenti, 2017), this study highlights how the problem appears within the retracted papers themselves. Students' accounts of intertextual falsification echo Markowitz and Hancock's (2016) finding that fraudulent papers are more heavily referenced, yet often in misleading ways. Even when citing reliable research, the analyzed papers frequently distorted meanings, omitted counter-evidence, or created false links to established work in order to perform legitimacy. Students' observations of terminological distortion further align with Markowitz and Hancock's (2016) evidence that fraudulent articles tend to employ more jargon and less readable, more abstract terminology, as well as with descriptions of "tortured phrases" and distorted technical language in unreliable texts (Cabanac, Labbé, and Magazinov 2021; Teixeira da Silva 2023).

Students' comments on rhetorical overstatement and compressed, discovery-like narratives can also be considered alongside Dehdarirad and Schirone's (2023) evidence that retracted biochemistry articles use less certainty language overall. However, the present findings neither confirm nor contradict those of Dehdarirad and Schirone. Whereas their study assesses certainty through the quantitative frequency of "positive" or certainty-related terms, the students in our study evaluated rhetorical stance more holistically, focusing on narrative framing, the absence of limitations, and the lack of space for doubt. These qualitative assessments suggest that certainty is not always expressed through identifiable metadiscourse markers; instead, it may be performed through the omission of uncertainty, the smoothing of narrative complexity, or the presentation of results as seamless and definitive. This observation aligns with Keserlioglu, Kilicoglu, and Ter Riet (2019), who found that peer review rarely moderates the linguistic expression of confidence, and extends their work by indicating that reviewers should also attend to imbalances in narrative stance as potential rhetorical red flags.

Reflections on rhetorical inconsistency further underscore the need for qualitative assessment of claims when attending to rhetorical features. Reviewers must examine the alignment between a paper's claims, methods, and reported results. Students repeatedly noted cases where ambitious or definitive claims rested on relatively simple or small-scale experiments, or where the narrative framing suggested rigor and innovation that were not fully supported by the underlying design. Such misalignments are easily obscured by genre expectations, particularly in short, high-impact formats that foreground striking results and condensed narratives. Reviewer training that encourages systematic comparison between the scope of claims and the evidential base may help make these inconsistencies more visible.

Methodological opacity revealed how the performance of rigor can obscure selective reporting or incomplete disclosure. Students frequently observed that methods sections appeared detailed yet lacked sufficient information for replication or verification. Notably, none of the students reported consulting reviewer guidelines when evaluating methodological adequacy, relying instead on disciplinary intuition.

Previous research indicates that even when reviewers refer to such guidelines, doing so does not necessarily improve the completeness of reporting (Speich et al. 2023). This finding points to two interconnected needs: (1) the development of clearer and more enforceable reviewer guidelines, and (2) explicit training to help reviewers apply these guidelines critically and consistently.

These rhetorical weaknesses point beyond individual misconduct toward systemic pressures that shape scientific communication (Armond et al., 2024). The industrialization of academia and the demand for clean, marketable results encourage selective reporting, rhetorical compression, and overconfidence. As several students noted, publication genres that reward brevity and certainty, such as "letters" and "breakthrough reports", often leave "no space for doubt". These conventions privilege the performance of discovery over the practice of transparent inquiry.

Pedagogically, retracted papers represent an underused yet useful resource. They are publicly available and ethically suitable for analysis, offering authentic examples of how credibility can be textually performed and undermined. Using retractions in training workshops allows doctoral students to engage critically with the epistemic and rhetorical complexities of research writing without breaching confidentiality. It also provides a realistic view of scientific practice, where technical competence and rhetorical overreach often coexist. This pedagogical use served two interconnected functions. First, it enabled students to practice identifying *how* unreliability is performed, rather than simply locating what is "wrong". Second, it fostered early reviewer literacy, prompting participants to examine how claims are framed, how evidence is staged, and how authority is constructed rhetorically. As Hesselberg et al. (2023) observe, reviewers often learn by doing, only after being invited to review formally. Structured engagement with retracted texts creates a safe, reflective context for developing that competence.

Another insight concerns visibility and access to retractions. Despite being provided with databases of retracted papers, most students chose high-profile, media-amplified cases. This reflects the broader visibility bias in how retractions are discussed: high-impact failures dominate attention, while less publicized corrections remain obscure. Increasing the discoverability of retracted work through clearer labeling, searchable disciplinary categories, or dedicated journal sections could help not only with limiting citations to retractions (Teixeira da Silva & Bornemann-Cimenti, 2017) but could also support reviewer training.

This study has several limitations that should be acknowledged. First, the dataset consists of reflective analyses produced within a single doctoral writing course at one institution; therefore, the findings reflect a specific educational context and may not be generalizable across disciplines, institutions, or levels of reviewer experience. Second, because students selected their own retracted articles, the corpus of texts they analyzed was heterogeneous in genre and scope, which may have shaped the types of rhetorical issues they were able to identify. Third, the analysis draws on students' written reflections rather than direct observations of their reading processes, capturing their interpretations rather than the full range of cues they may have noticed.

Future research could address these limitations by incorporating multiple cohorts across institutions or disciplines, enabling comparative insights into how researcher background shapes rhetorical sensitivity. Experimental designs that compare different pedagogical interventions (e.g., retracted vs. unretracted articles, guided vs. unguided analysis) could help isolate which components most effectively support reviewer development. Longitudinal research could also examine whether early training in rhetorical sensitivity influences how doctoral researchers later perform peer review in authentic contexts.

Conclusion

The main aim of this study was to increase novice reviewers' rhetorical sensitivity through analyzing retracted articles. The outcome, however, might be considered an initial framework for analyzing how rhetorical features contribute to scientific unreliability, addressing the need for deeper qualitative understanding of retracted articles, an area that remains understudied (Mertkan et al. 2025). The five identified categories offer a promising foundation for developing a more systematic taxonomy of *Pandora's box*: the rhetorical mechanisms through which unreliable writing achieves an appearance of legitimacy. As Moylan and Kowalczuk (2016) note, retraction notices alone are often opaque and engaging directly with the retracted texts themselves fills that gap pedagogically. By treating writing as a site where credibility is performed, this approach complements existing strategies in research integrity and reviewer education. It demonstrates that raising rhetorical awareness is as essential to maintaining scientific integrity as verifying methodological soundness or ensuring ethical compliance.

Acknowledgments

The author gratefully acknowledges the doctoral researchers who participated in the *Writing for Publication* course and contributed their reflections to this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Armond, A. C., K. D. Cobey, and D. Moher. (2024). "Research Integrity definitions and challenges." *Journal of Clinical Epidemiology* 171: 111367. https://doi.org/10.1016/j.jclinepi.2024.111367 .

Bazerman, C. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. University of Wisconsin Press.

Bik, E. M., A. Casadevall, and F. C. Fang. 2016. "The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications." *mBio* 7 (3): e00809–16. https://doi.org/10.1128/mBio.00809-16.

Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Cabanac, G., C. Labbé, and A. Magazinov. 2021. "Tortured Phrases: A Dubious Writing Style Emerging in Science [Preprint]. *arXiv* https://arxiv.org/abs/2107.06751.

Clark, I. L. 2007. *Writing the Successful Thesis and Dissertation: Entering the Conversation*. Pearson Education.

Dehdarirad, T., and M. Schirone. 2023. "Use of Positive Terms and Certainty Language in Retracted and Non-Retracted Articles: The Case of Biochemistry." *Journal of Information Science*. Advance online publication. https://doi.org/10.1177/01655515231176650.

Fanelli, D. 2013. "Why Growing Retractions Are (Mostly) a Good Sign." *PLOS Medicine* 10 (12): e1001563. https://doi.org/10.1371/journal.pmed.1001563.

Fang, F. C., R. G. Steen, and A. Casadevall. 2012. "Misconduct Accounts for the Majority of Retracted Scientific Publications." *Proceedings of the National Academy of Sciences* 109 (42): 17028–17033. https://doi.org/10.1073/pnas.1212247109.

Gilbert, G. N., and M. J. Mulkay. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. Cambridge University Press.

Grieneisen, M. L., and M. Zhang. 2012. "A Comprehensive Survey of Retracted Articles from the Scholarly Literature." *PLOS ONE* 7 (10): e44118. https://doi.org/10.1371/journal.pone.0044118.

Hesselberg, J. O., T. K. Dalsbø, H. Strømme, I. Svege, and A. Fretheim. 2023. "Reviewer Training for Improving Grant and Journal Peer Review." *Cochrane Database of Systematic Reviews* 2023 (11): MR000056. https://doi.org/10.1002/14651858.MR000056.pub2.

Hesselmann, F., V. Graf, M. Schmidt, and M. Reinhart. 2016. "The Visibility of Scientific Misconduct: A Review of the Literature on Retracted Journal Articles." *Current Sociology* 65 (6): 814–845. https://doi.org/10.1177/0011392116663807.

Hu, W., G. Yan, J. Zhang, Z. Chen, Q. Qian, and S. Wu. 2025. "Analysis of Scientific Paper Retractions Due to Data Problems: Revealing Challenges and Countermeasures in Data Management." *Accountability in Research*. Advance online publication. https://doi.org/10.1080/08989621.2025.2531987.

Hyland, K. 2005. *Metadiscourse: Exploring Interaction in Writing*. Continuum.

Keserlioglu, K., H. Kilicoglu, and G. Ter Riet. 2019. "Impact of Peer Review on Discussion of Study Limitations and Strength of Claims in Randomized Trial Reports: A Before and After Study." *Research Integrity and Peer Review* 4:19. https://doi.org/10.1186/s41073-019-0078-2.

Khuder, B. (2025). "Enhancing disciplinary voice through feedback-seeking in AI-assisted doctoral writing for publication." *Applied Linguistics*. https://doi.org/10.1093/applin/amaf022.

Khuder, B., and B. Petric. (2026). "A theoretical framework of collaborative authorial voice: Cognitive, social, and textual dimensions." *Journal of Second Language Writing* 71: 101278. https://doi.org/10.1016/j.jslw.2025.101278 .

Markowitz, D. M., and J. T. Hancock. 2016. "Linguistic Obfuscation in Fraudulent Science." *Journal of Language and Social Psychology* 35 (4): 435–445. https://doi.org/10.1177/0261927X15614605.

Mertkan, S., D. Mills, A. Takir, and E. Emmioglu Sarıkaya. 2025. "Research on Retractions: A Systematic Review and Research Agenda." *Accountability in Research*. Advance online publication. https://doi.org/10.1080/08989621.2025.2542203.

Moylan, E. C., and M. K. Kowalczuk. 2016. "Why Articles Are Retracted: A Retrospective Cross-Sectional Study of Retraction Notices at BioMed Central." *BMJ Open* 6 (11): e012047. https://doi.org/10.1136/bmjopen-2016-012047.

Ou, A. W., B. Khuder, S. Franzetti and R. Negretti. (2024). "Conceptualising and cultivating Critical GAI Literacy in doctoral academic writing." *Journal of Second Language Writing* 66: 101156. https://doi.org/10.1016/j.jslw.2024.101156.

Petrou, C. 2024. "Guest Post – Making Sense of Retractions and Tackling Research Misconduct." *The Scholarly Kitchen*. April 18. https://scholarlykitchen.sspnet.org/2024/04/18/guest-post-making-sense-of-retractions-and-tackling-research-misconduct/.

Rivera, H., and J. A. Teixeira da Silva. 2021. "Retractions, Fake Peer Reviews, and Paper Mills." *Journal of Korean Medical Science* 36 (24): e165. https://doi.org/10.3346/jkms.2021.36.e165.

Speich, B., E. Mann, C. M. Schönenberger, et al. 2023. "Reminding Peer Reviewers of Reporting Guideline Items to Improve Completeness in Published Articles: Primary Results of 2 Randomized Trials." *JAMA Network Open* 6 (6): e2317651. https://doi.org/10.1001/jamanetworkopen.2023.17651.

Swales, J. M., and C. B. Feak. 2012. *Academic Writing for Graduate Students*. 3rd ed. University of Michigan Press.

Teixeira da Silva, J. A. 2023. ""Tortured Phrases" in Preprints." *Current Medical Research and Opinion* 39 (5): 785–787. https://doi.org/10.1080/03007995.2023.2201098.

Teixeira da Silva, J. A., and H. Bornemann-Cimenti. (2017). "Why do some retracted papers continue to be cited?" *Scientometrics* 110 (1): 365–370. https://doi.org/10.1007/s11192-016-2178-9.

Walters, W. H., and E. I. Wilder. 2023. "Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT." *Scientific Reports* 13 (1): 14045. https://doi.org/10.1038/s41598-023-41032-5.

Appendix

Appendix 1 *Assignment questions to analyze articles (adapted from Clark 2007)*

(1) **Text – What does the article say, and how does it say it?**

- How is the text structured (title, headings, figures, and sections)?
  Does this structure help you navigate and understand its argument?
- What "moves" does the text make; how does it introduce a problem, justify its significance, and present results?
- How are key claims, methods, and findings represented?
  Do they align logically, or do you detect gaps or overstatements?
- What tone, terminology, or stylistic features contribute to its authority or "scientific sound"?
- Are there inconsistencies between what is claimed and what is shown (e.g., graphs, results, data interpretation)?

(2) **Context – within what rhetorical or disciplinary situation was the text produced?**

- What conversation, controversy, or problem does this paper claim to address?
- Which prior studies does it cite or argue against?
  Are any important works missing or used in misleading ways?
- How does the text align with or deviate from genre expectations in this field (e.g., a letter vs. a full research article)?
- What external factors (e.g., funding sources) might shape how the paper performs credibility?

(3) **Reader – How does the reader engage with and evaluate the text?**

- What expectations do you bring to this article, and how does the text meet or violate them?
- Did your confidence in the paper change at any point? If yes, when and why?
- How do your disciplinary background and reading purpose shape your interpretation?
- What did you learn from reading this paper about how credibility is performed in scientific writing?
- Would you continue reading or cite this paper? Why or why not?