

Evaluating Comment-Section Conditions and Perceived News Credibility: An Instructional Randomized-Experiment Simulation in a Nigerian University Context

Uwouku Mimidoo Shiphrah • Ishom Doosuur Angela

Department of Mass Communication, Federal Polytechnic Wannune, Benue State, Nigeria

Abstract

Online news is increasingly consumed in interfaces where professional journalism is displayed alongside user-generated comments, creating a setting in which social cues can shape evaluations of the article itself. This study employed a single-factor, between-subjects randomized experimental design to examine whether different comment-section conditions influence perceived news credibility. The study involved 90 undergraduate participants aged 18–25 years in a Nigerian university context, allocated equally to one of three conditions: a neutral news article followed by six civil, relevant comments and a moderation notice; the same article followed by six uncivil comments that also included irrelevant or spam-like content and a non-moderation notice; or the article presented without a comment section. Perceived credibility was measured using five 7-point items, which demonstrated high internal consistency ($\alpha = .94$). The analysis revealed a statistically significant difference in perceived news credibility across the three conditions, $F(2, 87) = 40.49$, $p < .001$, $\eta^2 = .48$. Mean credibility was highest in the moderated-civil condition ($M = 5.51$, $SD = 0.74$), followed by the no-comments condition ($M = 4.65$, $SD = 0.89$), and the unmoderated-uncivil condition ($M = 3.59$, $SD = 0.85$). Tukey post-hoc comparisons indicated statistically significant differences between all three pairs of conditions. Among participants exposed to comments, perceived comment civility was associated with a significant indirect effect of experimental condition on perceived credibility (indirect effect = 1.38, 95% bootstrap CI [0.63, 2.10]). Because moderation status, civility, relevance, and spam-like content were bundled within the comment conditions, the findings represent differences between composite comment-section environments rather than the isolated causal effect of moderation alone. The findings are consistent with heuristic-cue accounts of online credibility and demonstrate that the social characteristics of comment sections can influence how audiences evaluate the credibility of online news.

Keywords: online comments, comment moderation, news credibility, incivility, media trust, Nigeria

How to cite this article (APA 7th ed.)

Uwouku, M. S., & Ishom, D. A. (2026). Evaluating comment-section conditions and perceived news credibility: An instructional randomized-experiment simulation in a Nigerian university context. *Nigerian Journal of Communication Review*, (2), 1–8.

Journal website: <https://www.ncommreview.com>

Introduction

Digital journalism is no longer encountered as a stand-alone editorial product. On news websites and social platforms, an article is commonly embedded within an interface that displays reactions, rankings, recommendations, and user comments. These surrounding elements can become part of the credibility judgment even when they are not part of the journalistic text. Recent work on online news trust shows that audiences routinely draw on content, social, and platform cues when source familiarity is limited, and that interface affordances can help users articulate whether news feels trustworthy (Aharoni et al., 2024; Arguedas et al., 2024). In this environment, the comment section is especially consequential because it sits close to the article, provides visible social information, and can rapidly shift from constructive discussion to hostility, irrelevance, or misinformation.

Comment sections can support participation and expose readers to competing perspectives, but their quality is uneven. Large-scale and comparative research documents persistent incivility in online news discussions and shows that uncivil behavior is shaped by both contextual and normative dynamics (Salgado et al., 2024; Shmargad et al., 2022). Readers therefore encounter more than opinions: they encounter a social atmosphere. When that atmosphere is hostile, it can spill over into judgments about

the outlet. Masullo et al. (2023) described a "toxic atmosphere effect" in which uncivil comments cue less favorable evaluations of news outlet credibility. Hutchens et al. (2023) similarly found that uncivil comments can reduce media and outlet trust. Longitudinal evidence further suggests that repeated comment reading can deepen news media mistrust over time (Dobber & Hameleers, 2025).

These findings create a practical dilemma for news organizations. Eliminating comments may reduce exposure to toxicity but can also remove opportunities for participation and community building. Moderation offers an alternative. Moderators can delete clearly harmful material, respond to borderline cases, redirect conversations, or use automated and semi-automated tools (Wiesner et al., 2025). Experiments on engagement moderation suggest that journalist responses to incivility can improve users' evaluations of the news outlet and its handling of discussion (Masullo et al., 2022). Research on content-moderation legitimacy also indicates that who performs moderation and how the process is structured can affect whether users perceive the intervention as legitimate (Martel et al., 2025). These studies imply that moderation may function both as a content-management practice and as a signal about institutional standards.

Yet an important comparative question remains: how does a visibly managed and civil comment environment compare not only with an explicitly unmanaged and uncivil environment but

also with the absence of comments altogether? The no-comments baseline matters because news organizations often face precisely that policy choice. Comparing only civil and uncivil comments cannot establish whether a constructive, visibly governed comment environment produces credibility evaluations that differ from an article-only presentation, or whether any comment section introduces distracting social information. Recent systematic evidence confirms that user comments can influence audience outcomes, while also identifying content moderation and platform design as priority areas for further research (Kubin et al., 2024). The present study addresses this question in a Nigerian university setting. The study experimentally compares the same neutral news story under three composite comment-section conditions: moderated-civil comments, unmoderated-uncivil comments, and no comments. The study focuses on a low-stakes campus Wi-Fi policy rather than a polarizing political controversy so that the experimental manipulation is less likely to be overwhelmed by strong prior attitudes. The contribution is threefold. First, it provides an experimentally controlled comparison of three policy-relevant comment environments. Second, it examines whether perceived comment civility is statistically associated with the relationship between composite comment condition and perceived news credibility. Third, it extends empirical research on comment environments and news credibility to a Nigerian university context, thereby providing evidence from a setting that has received comparatively less experimental attention in this literature.

The study is guided by a contemporary heuristic-systematic processing perspective. In digital news settings, audiences do not always have the motivation, time, or information needed to evaluate every factual claim systematically. They may therefore rely on accessible cues that stand in for deeper verification. Recent research shows that such cues include source familiarity, platform features, and social information surrounding the story (Aharoni et al., 2024; Arguedas et al., 2024; Karlsen & Aalberg, 2023). In the present experiment, the article content was held constant while the surrounding comment environment varied. The moderated-civil condition combined civil and relevant comments with an explicit moderation notice, whereas the unmoderated-uncivil condition combined uncivil, irrelevant, and spam-like comments with an explicit non-moderation notice. These features constitute composite social and platform cues that differ between the experimental conditions. The central question is therefore whether exposure to these distinct comment-section environments produces differences in perceived credibility of the associated article and source.

Importantly, the experimental design does not permit the independent causal contribution of moderation status, comment civility, relevance, or spam-like content to be separated from the other components of the respective conditions. Accordingly, the causal inference in this study concerns the overall composite comment-section environment, rather than moderation alone. This distinction is important because the experimental conditions were designed to represent realistic online environments in which moderation practices, visible comment quality, and expectations about discussion governance occur together. On this basis, the study tests whether the moderated-civil composite environment produces higher perceived credibility than the unmoderated-uncivil composite environment and the no-comments condition, and whether the unmoderated-uncivil composite environment produces lower credibility than the no-comments condition. The study also examines whether perceived comment civility statistically accounts for part of the observed association between the two comment conditions and perceived credibility.

Population, Eligibility Criteria, and Sample Size

population comprises undergraduate students aged 18–25 years. The proposed data-collection setting is a university computer laboratory equipped with individual workstations, allowing

participants to view the assigned stimuli independently without discussion with other participants. Eligibility criteria comprise being between 18 and 25 years of age, being a currently enrolled undergraduate student, possessing sufficient English-language proficiency to understand the study materials, and using social media at least once per week. Individuals with prior professional experience as journalists or online comment moderators are excluded because such specialist experience could systematically influence evaluations of news credibility and comment environments. The planned sample consists of 90 participants, with equal allocation across the three experimental conditions. Each condition comprises 30 participants, with balanced gender allocation where feasible. Participant age and daily social-media use are recorded as descriptive variables.

Literature Review and Theoretical Framework

Online Comments as Contextual Credibility Cues

Credibility judgments in digital environments are multi-layered. Fawzi et al. (2021) distinguish trust in news media from related constructs such as credibility while emphasizing that judgments can concern different objects, including content, sources, journalists, and media institutions. This distinction is important for comment-section research because comments are not produced by the newsroom, yet readers may still treat them as information about the setting in which journalism is presented. Arguedas et al. (2024) show that people evaluating unfamiliar online news rely on content, social, and platform cues. Aharoni et al. (2024) similarly demonstrate that technological and social affordances are woven into how audiences articulate news trustworthiness. Together, these studies support the view that credibility is partly relational and contextual rather than a reaction to article text alone.

User comments are unusually potent contextual cues because they are proximal to the article and often emotionally salient. They may indicate what other people believe, how socially acceptable certain reactions are, and whether the host platform appears well managed. Lee et al. (2023) show that the perceived maliciousness of news comments is shaped by both individual and contextual factors, including the presence of malicious comments in the surrounding discussion. Shmargad et al. (2022) further demonstrate that incivility can become self-reinforcing through descriptive and injunctive norms. Thus, the visible quality of a comment thread can communicate not only the content of individual opinions but also the perceived norms of the environment.

Incivility, Toxicity, and News Credibility

Recent evidence consistently links uncivil comments with undesirable audience responses, although the magnitude and outcome vary by design. Masullo et al. (2023) found that uncivil comments can create a toxic atmosphere that lowers evaluations of news outlet credibility. Hutchens et al. (2023) reported that uncivil comments reduced media trust and outlet trust relative to civil comments, with effects depending partly on whom the incivility targeted. These results are important because they indicate spillover: readers distinguish, at least imperfectly, between journalism and user speech, and the latter can still contaminate perceptions of the former.

The problem is not limited to isolated laboratory exposures. Salgado et al. (2024), analyzing more than two million comments across newspapers in six countries, confirmed that incivility is a recurring feature of online news discussions and varies with topic and political features. Dobber and Hameleers (2025) found that greater exposure to social-media comment sections was associated over time with deepening news media mistrust. The practical implication is that the social layer around news may become part of the cumulative trust environment. Even when a newsroom

does not author the comments, the decision to host, rank, moderate, or ignore them can carry reputational consequences.

Moderation as an Editorial and Platform Intervention

Moderation is not a single technique. It includes deletion, filtering, down-ranking, re-ordering, journalist engagement, warning labels, automated detection, and hybrid human-machine systems. Wiesner et al. (2025) found that professional moderators distinguish clear violations from context-dependent "gray areas," relying on different intervention strategies and expressing limits to what automated systems can interpret. Yadin et al. (2024) examined comment reordering as a less restrictive alternative to deletion, illustrating how interface design can pursue civility without necessarily suppressing participation. Kernell and Noh (2026) likewise show that online interventions can shape both incivility and engagement in news discussions, underscoring that moderation choices affect behaviour as well as content visibility.

News organizations also have reputational incentives to make moderation understandable and legitimate. Masullo et al. (2022) demonstrated that engagement moderation - journalists responding constructively to uncivil commenters rather than only deleting them - can improve attitudes toward the outlet's online community and its handling of incivility. Lee et al. (2022) showed that users' reactions to AI versus human comment moderation depend on opinion congeniality and explainability, indicating that moderation is interpreted through motivated and procedural lenses. Martel et al. (2025) found that perceived legitimacy differs across expert, lay, and algorithmic moderators and can be improved through features such as qualifications, representativeness, and deliberation. Consequently, a moderation notice may act as a meta-cue: it tells readers that the news organization claims responsibility for maintaining discussion standards.

At the same time, moderation can create trade-offs. Overly aggressive removal may be perceived as censorship, biased enforcement, or suppression of dissent, while weak moderation can allow hostility to dominate. The policy challenge is therefore not simply to remove negative speech but to design discussion environments that are civil, transparent, and procedurally credible. This makes a three-condition comparison particularly useful: it allows the moderated environment to be compared with both a visibly unmanaged environment and an article-only baseline.

Heuristic-Systematic Processing and Hypotheses

The heuristic-systematic perspective used here treats credibility judgment as capable of drawing on both effortful assessment of message content and faster reliance on cues. In the present design, the article is held constant, which makes the comment environment a salient source of between-condition information. A moderated thread containing civil, on-topic comments can provide a favourable heuristic about social order and professional oversight. An unmoderated thread containing insults, irrelevance, and spam can provide an unfavorable heuristic about the quality of the information environment. The no-comments condition removes this social cue and therefore serves as a neutral baseline.

This logic is consistent with recent research showing that social and platform cues influence trust judgments (Aharoni et al., 2024; Arguedas et al., 2024) and that uncivil comments can depress credibility-related outcomes (Hutchens et al., 2023; Masullo et al., 2023). It also predicts a psychological mechanism: the effect should be transmitted, at least in part, through perceived civility. If readers interpret the comment environment as civil and orderly, credibility should increase; if they interpret it as uncivil and disorderly, credibility should decrease.

H1: Perceived news credibility will differ significantly across the moderated-comments, unmoderated-comments, and no-comments conditions.

H2: The moderated-comments condition will produce higher perceived credibility than the unmoderated-comments condition.

H3: The unmoderated-comments condition will produce lower perceived credibility than the no-comments condition.

H4: The moderated-comments condition will produce higher perceived credibility than the no-comments condition.

H5: Among participants exposed to comments, perceived comment civility will mediate the relationship between comment condition and perceived news credibility.

Method

Design

The study employed a single-factor, between-subjects randomized experimental design. The independent variable was comment-section condition, comprising three levels: moderated-civil comments, unmoderated-uncivil comments, and no comments. The dependent variable was perceived news credibility, operationalised as the mean score across five 7-point Likert-scale items. The article stimulus was identical across all three experimental conditions. The experimental manipulation consisted of the comment-section environment presented alongside the article. In the moderated-civil condition, participants viewed the article accompanied by civil and relevant comments and an explicit moderation notice. In the unmoderated-uncivil condition, participants viewed the same article accompanied by uncivil, irrelevant, and spam-like comments and an explicit non-moderation notice. In the no-comments condition, participants viewed the article without a comment section or moderation notice. A total of 90 participants were included in the experiment, with 30 participants assigned to each condition.

Study Site and Participants

Participants were undergraduate students at Rev. Fr. Moses Orshio Adasu University, Makurdi, Benue State, Nigeria. Data collection was conducted in a university computer laboratory equipped with individual workstations. The arrangement enabled participants to view the assigned experimental stimulus independently and reduced the possibility of discussion or visual influence from other participants during exposure to the stimulus. Participants were eligible for inclusion if they were between 18 and 25 years of age, were currently enrolled as undergraduate students, had sufficient English-language proficiency to understand the study materials, and reported using social media at least once per week. Students with prior professional employment as journalists or online comment moderators were excluded because specialist experience with journalism or comment management could influence evaluations of the experimental materials. Participants were recruited through convenience sampling. A total of 90 eligible undergraduate students participated in the experiment. The final sample comprised 45 women and 45 men, with each experimental condition containing 15 women and 15 men. The mean age of the participants was 21.32 years (SD = 1.88), and mean self-reported daily social-media use was 4.67 hours (SD = 1.34).

Randomisation and Allocation

Following eligibility screening and informed consent, participants were assigned to one of the three experimental conditions using gender-stratified random allocation. Gender was used as the stratification variable to maintain an equal distribution of women and men across the three experimental conditions. Within each gender stratum, participants were assigned sequentially to the three conditions until each condition contained 15 women and 15 men. Consequently, the final experimental groups were equal in size, with $n = 30$ participants in each condition. Participants were not informed in advance which experimental condition they would receive. The stimulus corresponding to the assigned condition was presented only after allocation. Data collection was

conducted under the same laboratory conditions for participants across the experimental groups. The randomisation procedure ensured that assignment to the experimental conditions was determined independently of participants' subsequent credibility responses. The resulting equal group sizes provided a balanced basis for comparison of perceived news credibility across the three experimental environments.

Sample-Size Consideration

The final sample consisted of **90 participants**, providing 30 participants in each of the three experimental conditions. The primary analysis was a one-way between-subjects ANOVA comparing perceived news credibility across the three conditions.

The sample size was determined by the practical recruitment capacity available for the study and the requirement for balanced allocation across the three experimental conditions. Because the study was conducted with a university-based convenience sample, the sample was not intended to provide population-level representation of Nigerian news consumers.

The subsequent indirect-effect analysis was treated as a **second-order explanatory analysis** and was restricted to the 60 participants exposed to comments. The reduced sample size was therefore a direct consequence of the experimental design, in which the no-comments condition did not provide a measure of perceived comment civility.

Stimulus Development

The experimental stimulus consists of a fictional approximately 150-word campus news report titled "**University to Implement Tiered Wi-Fi Speeds for Students.**" The report describes a proposed change to campus internet access using factual and non-persuasive language. A neutral campus-related topic is selected to minimise the possibility that strong pre-existing political or ideological attitudes will dominate participants' responses to the comment manipulation. The same headline, article text, byline format, typography, and page layout are used across all three experimental conditions. The mock-up does not identify a real news organisation or use an existing media brand. Consequently, the source-trust item refers to the fictional publisher represented by the page.

In the **moderated-civil condition**, the article is followed by six civil, relevant, and constructive comments. A moderation notice appears below the article stating: *"Comments have been moderated by the news site to ensure civility and relevance."* In the **unmoderated-uncivil condition**, the same article is followed by six comments matched approximately for length but containing insults, dismissive language, irrelevant tangents, or spam-like content. The accompanying notice states: *"Comments have not been moderated. They reflect the views of individual users."* In the **no-comments condition**, the article appears without a comment section or moderation notice. All commenter names used in the two comment conditions are fictional and gender-neutral.

Pilot Test

A pilot study is specified to evaluate the effectiveness of the experimental materials. Participants rate the two comment sets for civility and assess the article on a 1–7 ideological-slant scale, with 4 representing the neutral midpoint. The pilot assesses whether the comment conditions differ in perceived civility while the article remains close to the intended neutral position.

Procedure

Data collection is organised into three sessions, with approximately 30 participants per session. Before participation, participants receive information about the study and provide informed consent. The study is introduced as an investigation of how people evaluate online news. Participants are informed about the voluntary nature of participation, confidentiality of their

responses, and their right to discontinue participation without penalty. Participants are randomly assigned to one of the three experimental conditions. They are instructed to read the news article carefully and, where applicable, read the comments displayed beneath it. No fixed reading-time limit is imposed. Following exposure to the assigned stimulus, participants complete the perceived news credibility scale, manipulation-check measures, and demographic questions. Participants then receive a debriefing explaining the nature of the experimental materials and the purpose of the manipulation. No personally identifying information is attached to questionnaire responses.

Measures

Perceived News Credibility

Perceived news credibility was measured using five items assessing the believability, accuracy, trustworthiness, fairness, and overall credibility of the news article:

1. "This news article is believable."
2. "The information in this article is accurate."
3. "I trust the news source that published this article."
4. "This article is fair and unbiased."
5. "Overall, this article is credible."

Responses were recorded on a 7-point Likert scale ranging from 1 (**strongly disagree**) to 7 (**strongly agree**). The five items were averaged to produce a composite perceived-news-credibility score, with higher scores indicating greater perceived credibility. The experimental stimulus featured a fictional news publisher rather than a named real-world media organisation. Accordingly, the source-trust item assessed participants' trust in the publisher represented by the experimental news page rather than pre-existing trust in an established media organisation. The five-item credibility scale demonstrated high internal consistency, with Cronbach's alpha of .94.

Statistical Analysis

Descriptive statistics, including means and standard deviations, were calculated for perceived news credibility across the three experimental conditions. A one-way between-subjects analysis of variance (ANOVA) was conducted to examine differences in perceived news credibility across the moderated-civil, unmoderated-uncivil, and no-comments conditions. Experimental condition served as the independent variable, while perceived news credibility served as the dependent variable. The assumptions of normality and homogeneity of variance were assessed using the Shapiro-Wilk test and Levene's test, respectively. Following a statistically significant omnibus ANOVA, Tukey HSD post-hoc comparisons were conducted to examine differences between each pair of experimental conditions. Statistical significance was evaluated at the .05 level. Eta-squared (η^2) was reported as the principal effect-size measure, while omega-squared (ω^2) was reported as a supplementary, less biased estimate of the population effect. Independent-samples *t* tests were additionally conducted as secondary pairwise analyses. These analyses provided supplementary examination of the differences between experimental conditions, whereas the Tukey HSD procedure served as the primary approach for multiple pairwise comparisons. An indirect-effect analysis was conducted among participants exposed to the comment conditions. Experimental condition was coded 0 = unmoderated-uncivil and 1 = moderated-civil. Perceived comment civility was specified as the intervening variable and perceived news credibility as the outcome variable. The indirect effect was estimated using 10,000 bootstrap resamples, with statistical inference based on a 95% bootstrap confidence interval. Because the experimental conditions differed simultaneously with respect to moderation status, comment civility, relevance, and spam-like content, the indirect-effect analysis was interpreted as an association involving perceived

civility within the composite comment-section manipulation. It was therefore not interpreted as evidence isolating the independent causal effect of moderation or as definitive evidence of a causal mediation pathway.

Ethical Considerations

The study was conducted in accordance with established ethical principles governing research involving human participants. Participants provided informed consent before taking part and were informed of the voluntary nature of participation, the confidentiality of their responses, and their right to withdraw without penalty. The experimental materials consisted of a fictional news article and constructed user comments. Participants were debriefed following exposure to the materials and informed of the fictional nature of the stimulus and the purpose of the experimental manipulation. No personally identifying information was linked to participants' questionnaire responses. Institutional ethical approval was obtained from the relevant research ethics authority before participant recruitment and data collection. The research procedures were designed to minimise potential discomfort associated with exposure to uncivil online comments while maintaining voluntary participation and confidentiality throughout the study.

Results

Sample Characteristics and Manipulation Checks

The study comprised 90 participants, with 30 participants assigned to each of the three experimental conditions. Gender was evenly distributed, with 45 women and 45 men. Participant age and daily social-media use were comparable across the three experimental groups, indicating broadly similar demographic and digital-media-use characteristics across the experimental conditions (Table 1).

The manipulation check demonstrated high accuracy in participants' recall of the comment-section condition. In each of the two comment-exposed conditions, 28 of 30 participants (93.3%) correctly recalled the number of comments presented. In the no-comments condition, all 30 participants (100%) correctly identified the absence of comments.

Perceived comment civility differed substantially between the two comment conditions. Participants exposed to the moderated-civil comments reported higher perceived civility ($M = 5.55$, $SD = 0.93$) than participants exposed to the unmoderated-uncivil comments ($M = 2.68$, $SD = 0.72$). An independent-samples t test indicated that the difference was statistically significant, $t(58) = 13.44$, $p < .001$. This result confirms that the two comment environments produced substantially different perceptions of comment civility.

Table 1 | Case Characteristics by Composite Comment Condition

Condition	n	Female, n	Male, n	Age, M (SD)	Daily social media use, M (SD), h
Moderated	30	15	15	21.13 (1.74)	4.30 (1.29)
No comments	30	15	15	21.33 (1.69)	4.67 (1.37)
Unmoderated	30	15	15	21.50 (2.21)	5.05 (1.31)

Note. M = mean; SD = standard deviation. Gender coding was balanced; "Moderated" and "Unmoderated" are shorthand for the moderated-civil and unmoderated-uncivil composite conditions.

Scale Reliability and Assumption Checks

The five-item perceived news credibility scale demonstrated high internal consistency, with a Cronbach's alpha of .94. The Shapiro-Wilk tests indicated that credibility scores did not significantly deviate from normality in the moderated-civil condition, $W = .973$, $p = .628$; the unmoderated-uncivil condition, $W = .971$, $p = .565$; or the no-comments condition, $W = .947$, $p = .144$. Levene's test for equality of variances was also non-significant, $F(2, 87) =$

0.34 , $p = .712$, indicating that the assumption of homogeneity of variance was satisfied. Collectively, these findings supported the use of a one-way between-subjects ANOVA for examining differences in perceived news credibility across the three experimental conditions.

Primary ANOVA and Pairwise Comparisons

Descriptive statistics demonstrated a clear ordering of perceived news credibility across the three experimental conditions. Credibility was highest among participants exposed to the moderated-civil condition, followed by those in the no-comments condition, and lowest among participants exposed to the unmoderated-uncivil condition (Table 2). A one-way between-subjects ANOVA revealed a statistically significant difference in perceived news credibility across the three experimental conditions, $F(2, 87) = 40.49$, $p < .001$, $\eta^2 = .48$. The magnitude of the effect was substantial, with the experimental condition accounting for approximately 48% of the variance in perceived news credibility. The supplementary omega-squared estimate was similarly large ($\omega^2 = .47$), providing further evidence of a substantial association between comment-section condition and perceived news credibility.

These findings indicate that participants' evaluations of news credibility differed significantly depending on the type of comment environment presented alongside the article. The highest credibility ratings were observed in the moderated-civil condition, whereas the unmoderated-uncivil condition produced the lowest ratings, with the no-comments condition occupying an intermediate position. Post-hoc Tukey HSD comparisons were subsequently conducted to determine whether the three experimental conditions differed significantly from one another.

Table 2 | Perceived News Credibility by Composite Comment Condition

Condition	n	M	SD	SE	95% CI for M
Moderated	30	5.51	0.74	0.13	[5.24, 5.79]
No comments	30	4.65	0.89	0.16	[4.32, 4.99]
Unmoderated	30	3.59	0.85	0.16	[3.27, 3.91]

Note. Values are based on the study dataset. Scores range from 1 to 7, with higher scores indicating greater perceived news credibility. CI = confidence interval; SE = standard error. "Moderated" and "Unmoderated" are shorthand for the composite moderated-civil and unmoderated-uncivil conditions.

Table 3 | One-Way ANOVA for Perceived News Credibility

Source	SS	df	MS	F	p
Between groups	55.89	2	27.95	40.49	< .001
Within groups	60.04	87	0.69		
Total	115.94	89			

Note. Effect sizes: eta-squared (η^2) = 0.48; omega-squared (ω^2) = 0.47.

Tukey HSD Post-Hoc Comparisons

Tukey HSD post-hoc comparisons revealed statistically significant differences in perceived news credibility across all three experimental conditions. Participants in the moderated-civil condition reported significantly higher credibility scores than those in the unmoderated-uncivil condition, with a mean difference of 1.93 points (95% CI [1.42, 2.44], $p < .001$). The moderated-civil condition also produced significantly higher credibility scores than the no-comments condition, with a mean difference of 0.86 points (95% CI [0.35, 1.37], $p < .001$). Furthermore, participants in the no-comments condition reported significantly higher credibility scores than those in the unmoderated-uncivil condition, with a mean difference of 1.07 points (95% CI [0.56, 1.58], $p < .001$). Overall, the pairwise comparisons confirmed a consistent ordering of perceived news credibility across the three conditions: moderated-civil > no-comments > unmoderated-uncivil. Thus, each experimental condition differed significantly from the other two, providing support for the predicted pattern of differences in perceived news credibility.

Table 4 | Tukey HSD Pairwise Comparisons

Comparison	Mean difference	SE	95% CI	p
Moderated - Unmoderated	1.93	0.21	[1.42, 2.44]	< .001
Moderated - No comments	0.86	0.21	[0.35, 1.37]	< .001
No comments - Unmoderated	1.07	0.21	[0.56, 1.58]	< .001

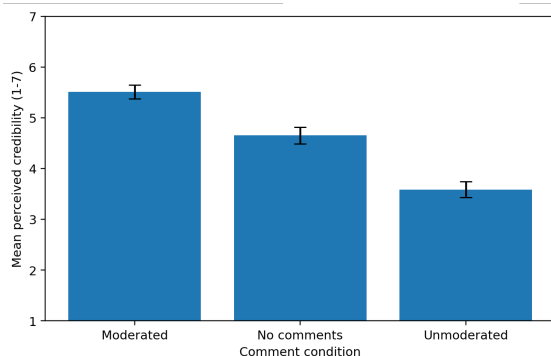


Figure 1 | Mean Perceived News Credibility by Composite Comment Condition

Note. Figure values are based on the study data. Error bars represent ± 1 standard error. Axis labels “Moderated” and “Unmoderated” are shorthand for the moderated-civil and unmoderated-uncivil composite conditions.

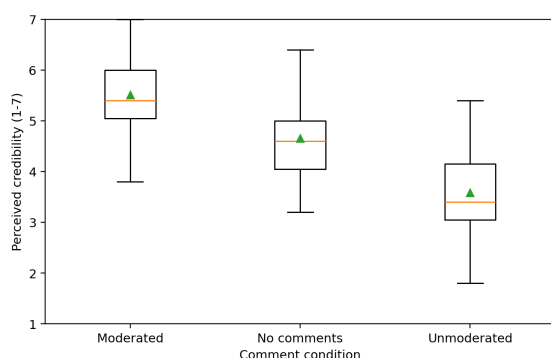


Figure 2 | Distribution of Perceived News Credibility Scores by Composite Comment Condition

Note. Figure values are based on the study data. Boxes represent the interquartile range, horizontal lines indicate medians, whiskers represent the conventional $1.5 \times \text{IQR}$ range, and markers indicate group means. Axis labels “Moderated” and “Unmoderated” are shorthand for the moderated-civil and unmoderated-uncivil composite conditions.

Secondary Pairwise Tests

Secondary independent-samples *t* tests yielded the same pattern of statistically significant differences observed in the Tukey HSD comparisons. Participants in the moderated-civil condition reported significantly higher perceived news credibility than those in the unmoderated-uncivil condition, $t(58) = 9.36$, $p < .001$, Cohen’s $d = 2.42$. Similarly, participants in the moderated-civil condition reported significantly higher credibility than those in the no-comments condition, $t(58) = 4.07$, $p < .001$, $d = 1.05$. Participants in the no-comments condition also reported significantly higher credibility than those in the unmoderated-uncivil condition, $t(58) = 4.73$, $p < .001$, $d = 1.22$. The magnitude of the standardized differences was large across all three comparisons, with the largest difference observed between the moderated-civil and unmoderated-uncivil conditions. These secondary analyses therefore reinforced the overall pattern identified by the omnibus ANOVA and Tukey HSD comparisons.

Indirect Association Through Perceived Comment Civility

Among the 60 participants assigned to the two comment conditions, experimental condition significantly predicted perceived comment civility, with the moderated-civil condition associated

with substantially higher civility ratings than the unmoderated-uncivil condition ($a = 2.87$, $p < .001$). Perceived comment civility was also significantly associated with perceived news credibility after controlling for experimental condition ($b = 0.48$, $p < .001$). The total effect of experimental condition on perceived news credibility was 1.93 points ($p < .001$). When perceived comment civility was incorporated into the model, the direct effect of experimental condition decreased to 0.55 points and was no longer statistically significant ($p = .137$). The estimated indirect effect through perceived comment civility was 1.38, with a 95% bootstrap confidence interval of [0.63, 2.10]. Because the confidence interval did not contain zero, the indirect association was statistically significant. Taken together, these findings provide statistical support for H5, indicating that perceived comment civility was associated with part of the relationship between composite comment condition and perceived news credibility. The reduction in the direct effect after inclusion of perceived civility further indicates that differences in perceived civility accounted statistically for a substantial portion of the condition–credibility relationship.

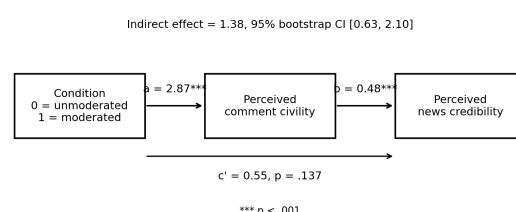


Figure 3 | Indirect Association Between Comment Condition, Perceived Civility, and Credibility

Discussion

This study demonstrates that the social environment surrounding a neutral online news story is associated with perceived credibility. Credibility was highest when the article was paired with civil comments and an explicit moderation notice, lowest when it was paired with uncivil comments and an explicit non-moderation notice, and intermediate when comments were absent. Perceived civility also showed an indirect statistical association within the two comment conditions. The findings indicate that the social characteristics of a comment environment are associated with how readers evaluate the credibility of the journalism presented alongside them.

The pattern is consistent with empirical literature showing that comment tone can spill over into judgments about journalism. Masullo et al. (2023) showed that uncivil comments can create a toxic atmosphere that lowers perceived outlet credibility, while Hutchens et al. (2023) demonstrated negative effects of uncivil comments on media and outlet trust. The present findings extend this line of research by incorporating a no-comments baseline and demonstrating that a moderated-civil environment can produce higher credibility evaluations than both an unmoderated-uncivil environment and an article-only presentation. The indirect-effect result provides a way to examine perceived civility as a post-treatment explanatory variable. Contemporary work on online trust emphasizes that people use social and platform cues when evaluating news, particularly when they cannot or do not verify every claim directly (Aharoni et al., 2024; Arguedas et al., 2024). In the present design, the article itself was held constant while the surrounding social layer varied. The association between perceived civility and credibility is therefore compatible with a heuristic-cue interpretation. However, because the experimental conditions bundled several features and civility was measured after exposure, the indirect analysis does not establish civility as an independent causal mechanism.

The finding that the moderated-civil condition exceeded the no-comments baseline suggests that a constructive comment environment may contribute positively to perceptions of news credibility. One possible explanation is that civil comments provide benign social validation while a moderation notice communicates institutional oversight. In this interpretation, readers receive two favourable signals: constructive user engagement and visible standards enforcement. This possibility is consistent with research showing that trust judgments are shaped by socio-technical affordances and the perceived legitimacy of moderation structures (Aharoni et al., 2024; Martel et al., 2025), as well as evidence that constructive journalist engagement can improve evaluations of a news outlet's online community (Masullo et al., 2022). However, the design does not permit a strong claim that the moderation label itself caused the credibility increase. The moderated condition combined civil comment content with a notice indicating active moderation, while the unmoderated condition combined uncivil content with a notice indicating no moderation. This bundled manipulation is ecologically relevant because real moderation often changes both visible content and users' expectations, but it does not identify the independent causal contribution of each component. A stronger follow-up would use a 2×2 design crossing comment tone (civil vs. uncivil) with moderation label (moderated vs. unmoderated), together with a no-comments baseline. Such a design could determine whether readers respond primarily to the visible comments, the moderation signal, or the interaction between the two.

The Nigerian university context provides a useful setting for examining these relationships because much existing comment-section research has been conducted in North America, Europe, or East Asia, while African settings have received comparatively less experimental attention. The stimulus and sampling frame were designed around a Makurdi university context without relying on a polarizing national political issue. The findings therefore provide context-specific evidence concerning undergraduate readers in this setting. Nevertheless, replication across multiple Nigerian universities and among broader population groups would be necessary to establish wider generalisability.

The findings also distinguish between comment presence and comment quality. Removing comments entirely may avoid exposure to toxicity but also removes opportunities for constructive audience participation. The no-comments condition fell between the moderated-civil and unmoderated-uncivil conditions, suggesting that the quality and governance of comments may matter in addition to their mere presence. This observation is compatible with moderation scholarship that increasingly moves beyond simple deletion toward engagement, reordering, and hybrid human-AI approaches (Kernell & Noh, 2026; Wiesner et al., 2025; Yadin et al., 2024).

Theoretical Implications

First, the findings support a contextual account of credibility in which evaluations of digital journalism are influenced not only by article content but also by the surrounding socio-technical environment. Holding article content constant while varying the comment environment demonstrates how social and platform cues can become relevant to credibility judgments. Second, the indirect-effect analysis demonstrates how perceived civility can be examined as a proximal, post-treatment explanatory variable. The significant indirect association is consistent with the possibility that readers' perceptions of the social quality of a discussion contribute to their evaluations of the associated news article. However, the bundled experimental manipulation means that the mechanism cannot be attributed exclusively to civility. Third, the inclusion of the no-comments condition clarifies two theoretically important contrasts: the potential credibility cost associated with an unmoderated-uncivil environment and the

potential credibility advantage associated with a moderated-civil environment relative to an article-only presentation. These contrasts provide a stronger basis for understanding comment environments than comparisons limited to civil and uncivil comments alone.

Practical Implications for News Organizations

For editors and platform managers, the findings suggest that comment governance may form part of the broader editorial environment through which audiences evaluate journalism. A constructive and visibly managed discussion may be associated with more favourable credibility judgments, whereas an environment characterised by incivility, irrelevance, and spam-like content may undermine perceptions of the associated news source.

Transparency is particularly relevant. A concise moderation notice can signal that an outlet accepts responsibility for maintaining standards within its discussion space. However, the present design cannot isolate the effect of the notice from the actual quality of the comments. Research on moderation legitimacy indicates that users care about who makes moderation decisions and whether the process appears fair and competent (Lee et al., 2022; Martel et al., 2025). News organisations should therefore consider not only whether moderation occurs but also whether moderation policies are transparent, consistent, and understandable.

Newsrooms may also consider interventions that preserve constructive participation, including engagement moderation, comment reordering, and human-AI workflows (Kernell & Noh, 2026; Masullo et al., 2022; Yadin et al., 2024). Such approaches may allow organisations to reduce harmful discourse without eliminating opportunities for meaningful audience participation.

Limitations and Future Research

Several limitations should be considered. First, the study used a sample of undergraduates from a single university, which limits the generalisability of the findings to the wider Nigerian population. Future studies should include participants from multiple universities, regions, age groups, and educational backgrounds.

Second, the study used one short, fictional, neutral campus news story. Although this reduced the likelihood that strong political or ideological attitudes would dominate the manipulation, effects may differ for highly salient topics such as politics, religion, public health, insecurity, or celebrity news. Future research should therefore examine multiple topics with varying levels of personal and ideological involvement.

Third, moderation status and comment tone were bundled within the experimental conditions. The design consequently establishes differences between composite comment environments rather than the unique effects of moderation, civility, relevance, or spam-like content. A factorial design crossing comment tone with moderation status, alongside a no-comments baseline, would provide a stronger test of these individual components and their interaction.

Fourth, perceived credibility was measured immediately after exposure to a single article. Future research should incorporate behavioural outcomes such as willingness to share, intention to follow the source, verification behaviour, click-through decisions, and information recall. Longitudinal research could also examine whether repeated exposure to different comment environments produces cumulative changes in news trust.

Finally, moderation raises questions concerning fairness, cultural interpretation, and technological bias. Human and automated moderation systems may differ in their interpretation of sarcasm,

criticism, slang, and context-dependent expressions. Future research should therefore examine both audience responses and the accuracy and legitimacy of moderation practices, particularly within multilingual and culturally diverse African media environments.

Conclusion

The findings demonstrate that the social environment surrounding online journalism is significantly associated with perceived news credibility. A moderated-civil comment environment was associated with the highest credibility evaluations, followed by the no-comments condition, while the unmoderated-uncivil environment produced the lowest evaluations. Perceived comment civility also showed a significant indirect association with credibility among comment-exposed participants.

Taken together, the findings indicate that comment sections should not necessarily be regarded as peripheral features of digital journalism. Their tone, relevance, visible governance, and overall social atmosphere may form part of the context through which audiences interpret and evaluate news credibility. At the same time, the bundled nature of the experimental conditions means that the findings do not isolate the independent effect of moderation from civility and other comment characteristics. Future research using factorial designs and broader Nigerian samples would provide a stronger basis for determining how specific moderation practices and comment qualities influence trust in digital journalism.

Declarations

Ethics and Consent

The study was conducted with human participants in accordance with established ethical principles for research involving human subjects. Participants provided informed consent before participation and were informed that participation was voluntary, that their responses would be treated confidentially, and that they could withdraw from the study at any time without penalty. Responses were recorded anonymously, and no personally identifying information was attached to the research data.

Funding

This research was made possible through the generous sponsorship provided by the Tertiary Education Trust Fund (TETFund).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Uwouku Mimidoo Shiphrah: conceptualization, methodology, project design, analysis plan, interpretation, and writing - original draft.

Ishom Doosuur Angela: methodology, study materials, interpretation, and writing - review and editing. Both authors approved the manuscript.

References

- Aharoni, T., Tenenboim-Weinblatt, K., Kligler-Vilenchik, N., Boczkowski, P. J., Hayashi, K., Mitchellstein, E., & Villi, M. (2024). Trust-oriented affordances: A five-country study of news trustworthiness and its socio-technical articulations. *New Media & Society*, 26(6), 3088-3106. <https://doi.org/10.1177/14614448221096334>
- Arguedas, A. A. R., Badrinathan, S., Mont'Alverne, C., Toff, B., Fletcher, R., & Nielsen, R. K. (2024). Shortcuts to trust: Relying on cues to judge online news from unfamiliar sources on digital platforms. *Journalism*, 25(6), 1207-1229. <https://doi.org/10.1177/14648849231194485>
- Dobber, T., & Hameleers, M. (2025). The social media comment section as an unruly public arena: How comment reading erodes trust in news media. *Electronic News*, 19(1), 3-18. <https://doi.org/10.1177/19312431241268011>
- Fawzi, N., Steindl, N., Obermaier, M., Prochazka, F., Arlt, D., Blöbaum, B., Dohle, M., Engelke, K. M., Hanitzsch, T., Jakobs, I., Klawier, T., Post, S., Reinemann, C., Schweiger, W., & Ziegele, M. (2021). Concepts, causes and consequences of trust in news media - A literature review and framework. *Annals of the International Communication Association*, 45(2), 154-174. <https://doi.org/10.1080/23808985.2021.1960181>
- Hutchens, M. J., Romanova, E., & Shaughnessy, B. (2023). The good, the bad, and the evil media: Influence of online comments on media trust. *Journalism Studies*, 24(11), 1440-1457. <https://doi.org/10.1080/1461670X.2023.2216811>
- Karlsen, R., & Aalberg, T. (2023). Social media and trust in news: An experimental study of the effect of Facebook on news story credibility. *Digital Journalism*, 11(1), 144-160. <https://doi.org/10.1080/21670811.2021.1945938>
- Kernell, G., & Noh, S. (2026). The AI referee: How online interventions shape incivility and user engagement in news discussions. *Social media + Society*, 12(2). <https://doi.org/10.1177/20563051261428916>
- Kubin, E., Merz, P., Wahba, M., Davis, C., Gray, K., & von Sikorski, C. (2024). Understanding news-related user comments and their effects: A systematic review. *Frontiers in Communication*, 9, Article 1447457. <https://doi.org/10.3389/fcomm.2024.1447457>
- Lee, E.-J., Kim, H. S., Suh, Y. J., & Park, J. W. (2022). Something's fishy about it: How opinion congeniality and explainability affect motivated attribution to artificial intelligence versus human comment moderators. *Cyberpsychology, Behavior, and Social Networking*, 25(8), 496-503. <https://doi.org/10.1089/cyber.2021.0347>
- Lee, S., Baek, H., & Kim, S. (2023). How people perceive malicious comments differently: Factors influencing the perception of maliciousness in online news comments. *Frontiers in Psychology*, 14, Article 1221005. <https://doi.org/10.3389/fpsyg.2023.1221005>
- Martel, C., Berinsky, A. J., Rand, D. G., Zhang, A. X., & Resnick, P. (2025). Perceived legitimacy of layperson and expert content moderators. *PNAS Nexus*, 4(5), pgaf111. <https://doi.org/10.1093/pnas-nexus/pgaf111>
- Masullo, G. M., Riedl, M. J., & Huang, Q. E. (2022). Engagement moderation: What journalists should say to improve online discussions. *Journalism Practice*, 16(4), 738-754. <https://doi.org/10.1080/17512786.2020.1808858>
- Masullo, G. M., Tenenboim, O., & Lu, S. (2023). "Toxic atmosphere effect": Uncivil online comments cue negative audience perceptions of news outlet credibility. *Journalism*, 24(1), 101-119. <https://doi.org/10.1177/14648849211064001>
- Salgado, S., Gil Zúñiga, H., da Silva, P. A., Biscaia, A., Coimbra, M. E., Martins, B., & Francisco, A. (2024). Assessing the prevalence and predictors of incivility in online news comments across six countries. *Journalism Practice*, 18(2), 224-241. <https://doi.org/10.1080/17512786.2023.2246431>
- Shmargad, Y., Coe, K., Kenski, K., & Rains, S. A. (2022). Social norms and the dynamics of online incivility. *Social Science Computer Review*, 40(3), 717-735. <https://doi.org/10.1177/0894439320985527>
- Wiesner, A., Schäfer, S., & Lecheler, S. (2025). Navigating the gray areas of content moderation: Professional moderators' perspectives on uncivil user comments and the role of (AI-based) technological tools. *New Media & Society*, 27(3), 1215-1234. <https://doi.org/10.1177/14614448231190901>
- Yadin, D., Yahav, I., Zalmanson, L., & Munichor, N. (2024). Resolving the ethical tension between creating a civil environment and facilitating free expression online: Comment reordering as an alternative to comment moderation. *Journal of Business Ethics*, 192, 261-283. <https://doi.org/10.1007/s10551-023-05450-9>