

Evaluating Visual and Textual Framing in Corporate Integrated Sustainability Reports: A Multimodal Framing Audit Framework

Authors

Abiodun Okunola, Abi Cit, Billy Elly

Date; 24 October 2026

Abstract

Corporate integrated sustainability reports increasingly combine narrative text with photographic imagery to communicate environmental and social commitments, yet stakeholders lack systematic tools to evaluate whether these two modalities tell a coherent story. Prior research has examined textual impression management and visual rhetoric in isolation, but no validated framework exists for auditing cross-modal alignment within integrated reports. This study develops and tests a **Multimodal Framing Audit Framework (MFAF)** to evaluate the congruence between textual claims and visual framing in corporate sustainability disclosures. Using a mixed-methods design, we analyzed 48 integrated reports from publicly listed firms across three sectors (energy, consumer goods, and financial services), coding 1,247 text segments and 892 photographs. The MFAF achieved **89.4% inter-rater reliability** in identifying framing congruence and flagged **23.7% of reports as exhibiting significant text-image dissonance**, particularly in environmental performance sections. Textual analysis revealed that firms used more hedging language when visual imagery depicted positive stakeholder engagement, suggesting compensatory framing. The framework offers practitioners a replicable diagnostic tool for report assurance and provides academics with a validated instrument for studying

multimodal sustainability communication. Implications extend to regulatory oversight of greenwashing and the design of more credible sustainability narratives.

Keywords: sustainability reporting, multimodal framing, impression management, visual

1. Introduction

1.1 Background

Corporate sustainability reporting has evolved from a voluntary supplement to a near-universal disclosure practice, with over 96% of the world's largest 250 companies now publishing integrated or sustainability reports . As these reports have grown in length and complexity, firms have increasingly supplemented narrative text with photographic imagery to convey environmental and social commitments . This multimodal turn reflects a strategic recognition: visual elements can communicate complex sustainability messages more efficiently than text alone, while also shaping stakeholder perceptions in ways that text cannot .

Scholarship on sustainability disclosure has traditionally focused on textual content, examining narrative strategies such as hedging, self-praise, and selective emphasis through the lens of impression management and legitimacy theory . A parallel stream of research has investigated visual elements—photographs, infographics, and design choices—as instruments of legitimacy work . Yet these two literatures have developed largely in isolation. The interaction between what firms *say* and what they *show* remains undertheorized and, critically, unmeasured.

Recent work has begun to bridge this gap. Okeke and Ugbebor propose a performative theory of corporate purpose that treats sustainability discourse as a governance mechanism, arguing that multimodal coherence—the alignment of verbal and visual signals—shapes whether stakeholder inferences stabilize or fracture . Okeke separately introduces a diagnostic approach to environmental misrepresentation, emphasizing the need for tools that can detect discrepancies between claims and evidence . These contributions establish the conceptual groundwork for a systematic audit of cross-modal framing, but they do not provide an operational instrument.

1.2 Problem Statement

Despite growing recognition that sustainability reports are multimodal artifacts, existing evaluation approaches remain modality-specific. Textual analyses employ content coding schemes that ignore accompanying imagery , while visual studies typically treat photographs as standalone rhetorical devices without systematically linking them to adjacent narrative claims . This fragmentation produces a diagnostic blind spot: a report may appear transparent when its

text is analyzed alone, and equally transparent when its images are analyzed alone, yet exhibit subtle but consequential dissonance when the two are read together.

The consequences of this gap are practical as well as theoretical. Regulators and assurance providers lack validated tools for detecting cross-modal impression management .

Stakeholders—particularly non-expert audiences who rely heavily on visual cues—may form judgments based on photographic evidence that contradicts or selectively amplifies textual disclosures . And researchers cannot test hypotheses about multimodal strategy without a reliable measure of alignment.

No validated framework currently exists for auditing the congruence between textual and visual framing in corporate integrated sustainability reports. This study addresses that absence.

1.3 Objectives of the Study

General objective: To develop and validate a framework for evaluating the congruence between textual and visual framing in corporate integrated sustainability reports.

Specific objectives:

1. To identify the dimensions along which textual and visual framing can diverge in sustainability reports, drawing on impression management and multimodal discourse theories.
2. To design a coding instrument—the Multimodal Framing Audit Framework (MFAF)—that operationalizes cross-modal congruence assessment.
3. To validate the framework through application to a multi-sector sample of integrated reports and assessment of inter-rater reliability.

1.4 Research Questions

1. What dimensions of textual-visual framing divergence are empirically observable in corporate integrated sustainability reports?
2. How reliably can trained coders apply the MFAF to identify instances of cross-modal dissonance?
3. What patterns of text-image alignment and misalignment characterize reports across different sectors and environmental performance contexts?

1.5 Significance of the Study

For practitioners and assurance providers: The MFAF offers a structured protocol for reviewing sustainability reports before publication or during third-party assurance. By flagging sections where textual claims and visual evidence diverge, it enables targeted revisions that reduce greenwashing risk.

For policymakers and regulators: The framework provides a basis for disclosure guidance that explicitly addresses multimodal coherence, extending beyond current requirements that treat text and imagery as separate disclosure categories.

For academic literature: The study contributes a validated measurement instrument to the sustainability communication field, enabling hypothesis testing about the antecedents and consequences of multimodal framing strategies.

For future researchers: The MFAF can be adapted for longitudinal analysis, cross-cultural comparison, and experimental designs testing stakeholder responses to congruent versus dissonant reports.

1.6 Scope and Limitations

This study analyzes integrated sustainability reports published between 2022 and 2024 by 48 publicly listed firms headquartered in Europe (n=24), North America (n=16), and Asia-Pacific (n=8), spanning three sectors: energy, consumer goods, and financial services. Only reports available in English and following the International Integrated Reporting Framework or GRI Standards were included.

Several limitations are acknowledged upfront. First, the sample size, while adequate for instrument validation, limits sector-level generalization. Second, the framework assesses *observable framing congruence* rather than *substantive accuracy*; a report may be internally coherent yet factually misleading. Third, photographic coding necessarily involves interpretive judgment, despite efforts to standardize through codebook development and reliability testing.

2. Literature Review

2.1 Conceptual Review

Integrated sustainability reporting. Integrated reports combine financial and non-financial information to present a holistic view of organizational value creation. Unlike standalone sustainability reports, they explicitly link environmental and social performance to business strategy and financial outcomes. This study focuses on the sustainability sections of integrated reports, where firms disclose environmental targets, performance data, and stakeholder impacts.

Textual framing. Textual framing refers to the linguistic choices through which firms present sustainability information: the use of hedges (e.g., “aim to,” “seek to”) versus boosters (e.g., “achieved,” “delivered”), the attribution of responsibility, and the selective emphasis on favorable outcomes. Prior research has documented systematic patterns in sustainability narratives, including the tendency to obscure negative performance through passive voice and nominalization.

Visual framing. Visual framing encompasses the selection, composition, and juxtaposition of photographs and graphics in reports. Photographs of smiling workers, pristine landscapes, and engaged stakeholders perform what Davison terms “visual rhetoric of objectivity,” naturalizing corporate claims by associating them with seemingly unmediated reality . The choice of *who* appears in photographs—and who does not—constitutes a form of framing that operates below the threshold of textual argument.

Cross-modal congruence. We define cross-modal congruence as the degree to which textual claims and visual representations in a given report section convey consistent implications about the firm’s sustainability performance, priorities, and stakeholder relationships. Congruence is not mere topical correspondence; it requires that the *evaluative stance* implied by text and image align. A photograph of a wind farm alongside text about emissions reduction is topically related; congruence depends on whether the text hedges or asserts, and whether the image depicts the firm’s own infrastructure or generic stock imagery.

2.2 Theoretical Framework

This study integrates two theoretical perspectives.

Impression management theory posits that organizations strategically manage stakeholder perceptions through selective disclosure, concealment, and framing . In sustainability reporting, impression management manifests in both verbal strategies (e.g., emphasizing awards and regulatory compliance while downplaying incidents) and visual strategies (e.g., selecting photographs that signal innocence or competence) . The theory predicts that firms facing legitimacy pressure will deploy multimodal resources in coordinated ways—or, when coordination fails, produce dissonance that attentive stakeholders can detect.

Multimodal discourse analysis, following Kress and van Leeuwen, treats text and image as semiotic modes with distinct affordances that jointly construct meaning . Text is linear, propositional, and amenable to hedging; image is simultaneous, affective, and resistant to qualification. This asymmetry creates the conditions for *compensatory framing*: a firm may rely on imagery to convey confidence that its text, constrained by disclosure regulations or legal caution, cannot assert directly. The MFAF is designed to detect precisely such patterns.

Okeke and Ugbebor’s performative purpose alignment theory extends these insights by specifying that sustainability discourse acquires governance force only when verbal and visual signals jointly stabilize stakeholder inference . Their framework treats multimodal coherence not as an aesthetic preference but as a condition for credibility—a claim that motivates the development of measurement tools.

2.3 Empirical Review

Li and Xie analyzed PetroChina’s ESG reports (2018–2024) using multimodal metadiscourse analysis, finding that visual boosters (positive imagery) systematically amplified linguistic

claims, creating an “immersive, persuasive” narrative that discouraged critical scrutiny . Their study demonstrated the interaction of text and image but did not develop a replicable congruence metric.

Pujiningsih, Utami, and Trisnawati examined environmental disclosures by 119 Green PROPER companies in Indonesia, identifying a strategy they term “narcissistic texts and innocent photos”: firms used self-praise in narrative while selecting photographs of people and pristine environments that conveyed innocence rather than operational reality . The study established that text and image can serve distinct impression management functions but stopped short of quantifying dissonance.

Experimental work by Invernizzi and colleagues found that low-context photographs—generic imagery not reflective of actual operations—significantly penalized reader trust, particularly when paired with positive performance text . This counterintuitive finding suggests that stakeholders may possess a “dissonance detector” that the MFAF could help formalize and refine.

Okeke’s diagnostic framework for environmental misrepresentation provides principles for identifying claims that lack evidentiary grounding, but it is oriented toward textual claims rather than multimodal artifacts . The present study extends that diagnostic logic to the text-image interface.

2.4 Research Gap

No validated, replicable framework exists for auditing cross-modal framing congruence in corporate integrated sustainability reports. Existing studies either analyze text and image separately, examine their interaction qualitatively without operationalization, or focus on experimental effects without providing diagnostic tools for report evaluation. The MFAF fills this gap by specifying coding dimensions, decision rules, and reliability benchmarks for assessing whether textual claims and visual representations tell a coherent or dissonant story.

3. Methodology

3.1 Research Design

This study employs a **mixed-methods instrument development design** with three phases: (1) framework specification through literature synthesis and pilot coding; (2) application to a multi-sector sample; and (3) reliability and validity assessment. This design is appropriate because the objective is to produce and validate a measurement instrument, a task requiring both qualitative codebook development and quantitative reliability testing.

3.2 Study Area / Population

The population consists of integrated sustainability reports published in English between 2022 and 2024 by publicly listed firms in three sectors: energy (n=18 eligible), consumer goods (n=22 eligible), and financial services (n=20 eligible). These sectors were selected to capture variation in environmental exposure and visual disclosure practices.

3.3 Sample Size and Sampling Technique

A stratified random sample of 48 reports (16 per sector) was drawn from the population. Sample size was determined by the need for sufficient cases to test inter-rater reliability across sector strata while remaining feasible for intensive multimodal coding. Stratification ensures representation of sectors with different legitimacy pressures and reporting conventions.

3.4 Data Collection Methods

Reports were downloaded from corporate websites and company investor relations portals. From each report, we extracted the **sustainability performance section** (typically 15–30 pages) for coding. Within these sections, all text paragraphs directly adjacent to photographs were paired with their accompanying images, yielding 1,247 text-image pairs. Photographs were classified by content type (people, nature, infrastructure, abstract) and composition (close-up, wide shot, stock-like). Text segments were coded for hedging, boosting, attribution, and specificity.

No data were simulated; all materials are publicly available corporate disclosures.

3.5 Research Instruments

Coding was conducted using **NVivo 14** for qualitative analysis and **R (Version 4.3.1)** for reliability statistics. The MFAF codebook specifies four congruence dimensions:

1. **Evaluative alignment:** Does the text assert confidence (boosters) while the image depicts ambiguity, or vice versa?
2. **Specificity correspondence:** Does specific, verifiable text accompany generic, stock-like imagery (a dissonance pattern)?
3. **Actor representation:** Do the people depicted in photographs correspond to the stakeholders discussed in adjacent text (e.g., workers, community members, executives)?
4. **Outcome visualization:** Does the imagery depict outcomes claimed in text (e.g., emissions reduction infrastructure) or unrelated scenes?

Each text-image pair was coded as *congruent*, *dissonant*, or *ambiguous* on each dimension. The framework draws on Okeke's diagnostic logic for identifying unsubstantiated claims and Okeke and Ugbebor's emphasis on multimodal coherence as a condition for discursive credibility .

3.6 Validity and Reliability

Content validity was established through expert review by three scholars in sustainability accounting and communication (two external, one internal), who assessed whether the four dimensions adequately captured the construct of cross-modal congruence. Revisions were made to clarify decision rules for ambiguous cases.

Inter-rater reliability was assessed by having two trained coders independently code 240 randomly selected text-image pairs (approximately 20% of the sample). Cohen’s kappa was calculated for each dimension. The overall agreement rate for congruence classification was **89.4%**, exceeding the conventional 80% threshold for acceptable reliability. Kappa values ranged from 0.74 (outcome visualization) to 0.86 (evaluative alignment), indicating substantial to almost perfect agreement.

Predictive validity was explored by examining whether reports flagged as high-dissonance in environmental sections also scored lower on an independent third-party environmental transparency rating (CDP disclosure score). A negative correlation ($r = -0.34$, $p < 0.05$) provided preliminary evidence that MFAF dissonance flags capture something consequential.

3.7 Data Analysis Techniques

Descriptive statistics summarize the prevalence of dissonance across sectors and dimensions. Chi-square tests examine sector differences in dissonance frequency. Qualitative content analysis of flagged dissonant pairs identifies recurrent patterns (e.g., “specific text + generic image” combinations). All analyses were conducted in R.

3.8 Ethical Considerations

This study uses only publicly available, de-identified corporate documents. No human subjects were involved, and no proprietary or confidential information was accessed. Institutional review board exemption was granted under the category of secondary analysis of public documents.

4. Results

4.1 Data Presentation

Table 1. Sample Characteristics by Sector

Sector	Reports (n)	Pages coded (mean)	Text-image pairs (n)
Energy	16	24.3	412

Sector	Reports (n)	Pages coded (mean)	Text-image pairs (n)
Consumer goods	16	21.7	398
Financial services	16	19.2	437
Total	48	21.7	1,247

Table 1 shows that energy reports contained the most text-image pairs per report (mean = 25.8), consistent with the sector’s extensive use of operational imagery.

Table 2. Dissonance Prevalence by Dimension and Sector

Dimension	Energy	Consumer Goods	Financial	Total
Evaluative alignment	18.4%	14.6%	11.2%	14.7%
Specificity correspondence	29.1%	22.4%	16.5%	22.7%
Actor representation	12.6%	19.3%	8.9%	13.6%
Outcome visualization	21.8%	15.1%	13.3%	16.7%
Any dissonance flag	27.4%	22.9%	20.6%	23.7%

Table 2 presents the proportion of text-image pairs flagged as dissonant on each dimension. Specificity correspondence was the most frequently flagged dimension overall (22.7%), indicating that generic stock imagery paired with specific textual claims is the dominant pattern of cross-modal misalignment.

4.2 Analysis of Results

Overall reliability. The MFAF achieved 89.4% inter-rater agreement on the primary congruence classification, with a mean kappa of 0.79 across dimensions. This exceeds the threshold for instrument validation and supports the framework’s use in future research and assurance contexts.

Sector patterns. Energy reports exhibited the highest dissonance rate (27.4%), driven primarily by specificity correspondence failures—generic images of wind turbines or solar panels accompanying text about company-specific emissions reductions. Financial services reports showed the lowest overall dissonance (20.6%), possibly reflecting more conservative visual strategies in a sector less reliant on operational imagery.

Compensatory framing. Qualitative analysis of dissonant pairs revealed a recurrent pattern: firms that used hedging language in text (“we aim to,” “we are working toward”) were more likely to accompany such hedges with confident, positive imagery (smiling workers, pristine environments). This pattern, observed in 41% of evaluative alignment dissonances, suggests that visual confidence compensates for textual caution—a finding consistent with the multimodal asymmetry thesis.

Predictive validity. Reports with three or more dissonance flags in environmental sections had significantly lower CDP disclosure scores (mean = 4.2, SD = 1.1) than reports with zero or one flag (mean = 5.8, SD = 0.9), $t(46) = 2.31$, $p = 0.026$. This provides preliminary evidence that MFAF flags are associated with broader transparency deficits.

5. Discussion

5.1 Interpretation

The findings support three principal conclusions. First, cross-modal dissonance in sustainability reporting is measurable and prevalent: nearly one in four text-image pairs in our sample exhibited framing misalignment on at least one dimension. This prevalence suggests that multimodal coherence is not routinely monitored by report preparers, despite its potential reputational consequences .

Second, the specificity correspondence dimension emerged as the most diagnostic. Generic imagery paired with specific claims is a signature of what Okeke terms environmental misrepresentation : the visual provides affective warrant for claims that the image cannot, in fact, substantiate. This pattern is distinct from outright fabrication; it operates through the selective juxtaposition of unverifiable visual evidence with verifiable textual assertions.

Third, the compensatory framing pattern—hedged text accompanied by confident imagery—illuminates a previously untheorized dynamic. Kress and van Leeuwen’s multimodal framework anticipated that different modes have different affordances, but the specific compensatory use of imagery to offset textual caution has not been documented at scale. This finding extends Okeke and Ugbebor’s performative purpose alignment theory by showing that the *failure* of alignment may itself be strategic: firms may deploy visual confidence precisely where textual confidence would invite regulatory or legal scrutiny .

5.2 Implications

Academic implications. The MFAF introduces a validated construct—cross-modal framing congruence—into the sustainability disclosure literature. Future research can use the instrument to test hypotheses about the antecedents of multimodal strategy (e.g., does legitimacy pressure increase dissonance or decrease it?) and its consequences (do stakeholders detect dissonance, and does it affect trust?). The compensatory framing finding suggests that impression management theory should be expanded to account for *modal substitution*, not merely selective disclosure within a single mode.

Practical implications. For report preparers and assurance providers, the MFAF offers a structured protocol for pre-publication review. We recommend that firms audit text-image pairs in environmental performance sections at minimum, using the four dimensions as a checklist. The 89.4% reliability rate indicates that trained reviewers can apply the framework consistently. For regulators, the findings support guidance that explicitly addresses visual-text alignment, moving beyond current requirements that treat imagery as decorative rather than evidential.

5.3 Limitations

1. **Sample scope.** The 48-report sample, while stratified, limits sector-level generalization. Financial services, in particular, may warrant larger samples given its distinct visual conventions.
2. **Construct boundary.** The MFAF measures observable congruence, not substantive accuracy. A report can be highly congruent yet factually misleading if both text and image consistently distort performance.
3. **Photographic interpretation.** Despite standardized coding rules, photographic meaning is inherently polysemic. The 89.4% agreement rate indicates acceptable reliability but does not eliminate interpretive variance.
4. **Temporal snapshot.** The 2022–2024 window captures current practice but cannot assess whether multimodal strategies are evolving in response to regulatory pressure.

5.4 Future Research Directions

- **Extension to assurance contexts:** Test whether MFAF-informed review reduces the incidence of dissonance in subsequent reporting cycles.
- **Stakeholder perception experiments:** Examine whether non-expert readers detect dissonance flagged by the MFAF and whether detection affects trust and investment intentions.
- **Longitudinal analysis:** Apply the framework to reports from 2015–2024 to assess whether multimodal congruence is improving or deteriorating.

- **Cross-cultural comparison:** Investigate whether dissonance patterns vary across reporting regimes with different regulatory oversight of visual disclosure.

6. Conclusion

This study developed and validated the Multimodal Framing Audit Framework (MFAF), achieving 89.4% inter-rater reliability in identifying cross-modal framing dissonance across 1,247 text-image pairs in 48 integrated sustainability reports. The framework revealed that nearly one-quarter of text-image pairs exhibit misalignment, with specificity correspondence—generic imagery accompanying specific claims—the most prevalent pattern. The central contribution is a replicable, validated instrument that transforms a previously impressionistic concern (do text and image tell the same story?) into a measurable and auditable property. For administrators, the MFAF provides a practical checklist for pre-publication review, particularly in environmental sections where dissonance risk is highest. As sustainability reporting continues to evolve toward greater visual sophistication, the capacity to evaluate multimodal coherence will become not merely an academic exercise but a condition for credible communication.

References

1. Okeke, A., & Ugbebor, I. (2026). Discursive governance and development goals: A performative theory of corporate purpose in sustainability discourse. *Business Strategy and the Environment*. <https://doi.org/10.1002/bse.71242>
2. Okeke, A. (2026). Green claims and grey areas: A diagnostic tool for assessing corporate environmental misrepresentation. *International Journal of Disclosure and Governance*, 1–17. <https://doi.org/10.1057/s41310-026-00258-2>
3. Li, X., & Xie, H. (2026). Investigating environmental, social, and governance (ESG) reports by PetroChina: A multimodal interactional metadiscourse analysis. *Critical Arts: A South-North Journal of Cultural & Media Studies*, *40*(1), 155–174. <https://doi.org/10.1080/02560046.2025.2588344>
4. Pujiningsih, S., Utami, H., & Trisnawati, N. (2026). Narcissistic texts and innocent photos in environmental disclosures by Green PROPER companies. *Oblik i Finans*, *111*(1), 45–56. [https://doi.org/10.33146/2518-1181-2026-1\(111\)-45-56](https://doi.org/10.33146/2518-1181-2026-1(111)-45-56)
5. Davison, J. (2015). Visualising accounting: An interdisciplinary review and synthesis. *Accounting and Business Research*, *45*(2), 121–165.
6. Cho, C. H., Laine, M., Roberts, R. W., & Rodrigue, M. (2015). Organized hypocrisy, organizational façades, and sustainability reporting. *Accounting, Organizations and Society*, *40*, 78–94.
7. Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
8. Invernizzi, A. C., Bellucci, M., & Acuti, D. (2025). The power and pitfalls of photographs in corporate sustainability reporting: Experimental insights on trust and environmental responsibility perceptions. *Corporate Social Responsibility and Environmental Management*. <https://doi.org/10.1002/csr.70081>
9. Hrasky, S. (2012). Visual disclosure strategies adopted by more and less sustainability-driven companies. *Accounting Forum*, *36*(3), 154–165.
10. Davison, J. (2021). Value creation attempts via photographs in sustainability reporting: A legitimacy theory perspective. *Meditari Accountancy Research*, *29*(2), 247–263.
11. Bini, L., Bellucci, M., & Giunta, F. (2020). Sustainability reporting after the Costa Concordia disaster: A multi-theory study on legitimacy, impression management and image restoration. *Accounting, Auditing & Accountability Journal*, *33*(8), 1909–1941.

12. Hahn, R., & Lülfs, R. (2014). Legitimizing negative aspects in GRI-oriented sustainability reporting: A qualitative analysis of corporate disclosure strategies. *Journal of Business Ethics*, *123*(3), 401–420.
13. Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*, *33*, 59–78.
14. Beattie, V., & Jones, M. J. (2008). A comparative study of the use of financial graphs in the corporate annual reports of major U.S. and U.K. companies. *Journal of International Financial Management & Accounting*, *8*(1), 33–68.
15. Dumay, J., Bernardi, C., Guthrie, J., & Demartini, P. (2016). Integrated reporting: A structured literature review. *Accounting Forum*, *40*(3), 166–185.