

Construct Validation for Vendor Environmental Compliance Scoring: A Design-Based Research Approach to Measure Corporate Environmental Misrepresentation

Authors

Abi Cit, Billy Elly, Timileyin Olawale, Jimmy Grace, Kate Bright

Date; October 13, 2026

Abstract

Vendor environmental compliance scoring systems have proliferated across supply chain management, yet their construct validity remains largely unexamined. Existing research identifies persistent methodological fragmentation in ESG and climate rating systems, with poor correlation among scores ostensibly measuring similar constructs . This study addresses the research gap by developing and validating a vendor environmental compliance scoring framework that integrates the tripartite construct of compliance, performance, and misrepresentation detection. Using a design-based research approach combining retrospective analysis of publicly available environmental disclosure data (n = 412 vendor-year observations) with prospective simulation, the study develops and validates a hybrid scoring model. Findings demonstrate an overall classification accuracy of 89.4% (AUC = 0.873, $p < .001$), with peer-relative disclosure variance and regulatory violation history emerging as the strongest predictors. The validated framework provides procurement practitioners and sustainability officers with a replicable, evidence-based tool for vendor assessment, while contributing to academic literature a construct-validated instrument that explicitly distinguishes genuine compliance from misrepresentation. The study concludes that vendor environmental compliance scoring requires multidimensional construct definition to avoid the greenwashing detection failures documented in existing rating systems .

Keywords: construct validation, vendor environmental compliance, greenwashing detection, supplier scoring, ESG ratings

1. Introduction

1.1 Background

Corporate environmental compliance assessment has become a central component of supply chain risk management, regulatory oversight, and sustainable procurement practices.

Organizations increasingly rely on vendor scoring systems to evaluate suppliers' environmental performance, with such assessments informing contract awards, preferred supplier designations, and supply chain risk mitigation strategies . The global ESG rating market has expanded substantially, with platforms such as EcoVadis, Cority, and various proprietary systems offering environmental compliance scoring as core functionality .

This expansion, however, has occurred without commensurate attention to the validity of the constructs these systems purport to measure. Kohler and colleagues' comprehensive review of corporate climate rating systems identified significant disparities in assessment scope and methodology, finding that on average only 35% of scoring systems aligned with best practices for reliability, accuracy, and fairness . The authors documented that while "overall transparency of the systems was found to be satisfactory, most systems are still hazy on the indicators and weighting functions they use" . This opacity raises fundamental questions about what vendor environmental scores actually measure.

The theoretical foundation for understanding environmental disclosure behavior draws on multiple perspectives. Okeke's work on measuring greenwashing established that construct validation for corporate environmental misrepresentation requires distinguishing between disclosure quantity and disclosure quality, and between aspirational claims and verified performance . Okeke and Rahim subsequently developed a triangulated greenwashing index demonstrating that single-source indicators systematically fail to identify misrepresentation when firms strategically disclose favorable information while omitting adverse data . These contributions establish that environmental compliance is not a unidimensional construct but rather encompasses discrete dimensions of regulatory adherence, performance outcomes, and disclosure integrity.

1.2 Problem Statement

Despite the proliferation of vendor environmental scoring systems, there exists no validated framework that explicitly models the distinction between genuine environmental compliance and environmental misrepresentation in vendor assessment contexts. This gap manifests in several critical ways.

First, existing vendor scoring systems typically conflate environmental performance with environmental compliance. Gunaratne's two-dimensional supplier sustainability rating model acknowledged this limitation, noting that "in an industry setting uni-dimensional system is insufficient to give an overall picture" and that "while it is important to understand how the supplier performs currently, it is equally important to know whether they can maintain the performance in the long run" . The compliance dimension—reflecting the durability and verifiability of reported performance—requires distinct operationalization from performance magnitude.

Second, the literature on greenwashing detection has developed independently from the literature on vendor scoring. Okeke and Rahim's tripartite greenwashing index demonstrated that firms employ heterogeneous misrepresentation strategies, including selective disclosure, symbolic management, and decoupling, which require differentiated detection approaches . Yet vendor assessment frameworks have not integrated these insights, leaving procurement decisions vulnerable to the same detection failures documented in sustainability rating contexts .

Third, and most fundamentally, no study has subjected vendor environmental compliance scoring to rigorous construct validation. Content validity, predictive validity, and reliability evidence are essential for establishing that a scoring instrument measures what it claims to measure . The current absence of such evidence means that procurement decisions based on vendor environmental scores rest on instruments of unknown validity.

The unsolved issue is therefore: what is the construct validity of existing approaches to vendor environmental compliance scoring, and how can a validated framework be designed that distinguishes compliance from misrepresentation?

1.3 Objectives of the Study

General objective: To develop and validate a vendor environmental compliance scoring framework that accurately distinguishes genuine environmental compliance from misrepresentation.

Specific objectives:

1. To identify key predictors of environmental misrepresentation in vendor disclosure data through synthesis of greenwashing detection and supplier assessment literatures.
2. To design a hybrid scoring model integrating content-validated indicators of environmental compliance, performance, and disclosure integrity.
3. To validate the framework using retrospective analysis of publicly available environmental disclosure data and prospective simulation.

1.4 Research Questions

1. What combination of disclosure, performance, and regulatory variables most accurately identifies environmental misrepresentation in vendor populations?
2. How does the proposed construct-validated framework compare to baseline scoring approaches in classification accuracy and feature parsimony?
3. What content validity evidence supports the inclusion of specific indicators in vendor environmental compliance scoring?

1.5 Significance of the Study

For practitioners and administrators: The validated framework provides procurement officers and supply chain managers with an evidence-based instrument for vendor assessment that explicitly accounts for misrepresentation risk. The identification of key predictors enables efficient screening without requiring exhaustive data collection for every vendor.

For policymakers: The study contributes construct validity evidence that can inform regulatory standards for environmental disclosure verification and vendor assessment transparency, addressing the concerns raised by scholars and regulators about the integrity of sustainability rating activities .

For academic literature: The study extends Okeke and Rahim's tripartite misrepresentation framework to the vendor assessment context, providing empirical validation of construct dimensions and demonstrating the application of content validity methodology to environmental scoring instruments.

For future researchers: The methodology provides a replicable template for construct validation of environmental assessment instruments, with transparent reporting of content validity indices and predictive performance metrics that can be adapted to diverse organizational and regulatory contexts.

1.6 Scope and Limitations

This study examines vendor environmental compliance scoring within organizational procurement contexts, using publicly available environmental disclosure data from vendor populations. The time period for retrospective analysis spans three years of disclosure data, with prospective simulation used to evaluate framework performance on unseen data.

The study excludes: (a) primary data collection from vendor organizations, relying instead on publicly available disclosures; (b) social and governance dimensions of ESG assessment, focusing exclusively on environmental constructs; (c) sector-specific regulatory variations beyond those captured in available violation data.

Key limitations acknowledged in advance include the reliance on disclosure data that may itself be subject to manipulation, the assumption of historical pattern stability in predictive modeling, and the use of simulated data for specific variables where direct observation was not feasible.

2. Literature Review

2.1 Conceptual Review

Vendor Environmental Compliance. Environmental compliance in vendor contexts refers to the degree to which supplier organizations adhere to applicable environmental regulations, standards, and contractual requirements. This encompasses regulatory adherence (meeting legal requirements), performance achievement (meeting stated environmental targets), and disclosure integrity (accurately representing environmental conditions) . The compliance construct differs from mere environmental performance in that it incorporates the sustainability and verifiability of performance claims over time .

Environmental Misrepresentation (Greenwashing). Okeke defines greenwashing as "corporate environmental misrepresentation" encompassing strategic disclosure practices that create favorable impressions without corresponding environmental action . Okeke and Rahim operationalize greenwashing through a tripartite framework distinguishing selective disclosure (omitting adverse information), symbolic management (adopting environmental rhetoric without substantive change), and decoupling (separating environmental policy from operational practice) . This multidimensional conceptualization is essential for vendor scoring because different misrepresentation types require different detection indicators.

Construct Validation. Content validity refers to "the degree to which elements of an assessment instrument are relevant to and representativeness of the targeted construct for a particular assessment purpose" . Predictive validity concerns the instrument's ability to forecast outcomes—in this context, identifying vendors with actual compliance deficiencies or misrepresentation risk. Reliability encompasses internal consistency and inter-rater agreement in scoring judgments.

2.2 Theoretical Framework

Prospect Theory and Disclosure Behavior. Prospect theory explains why firms facing environmental compliance challenges may engage in misrepresentation . When decision-makers frame environmental violations or performance shortfalls as potential losses, they exhibit risk-seeking behavior in the disclosure domain—preferring the uncertain outcomes of misrepresentation over the certain penalties of accurate disclosure. This theoretical lens supports the inclusion of financial distress and performance gap indicators as predictors of misrepresentation risk.

Theory of Planned Behavior in Vendor Assessment. The theory of planned behavior illuminates how procurement decision-makers utilize environmental compliance scores . Attitudes toward vendor environmental assessment, subjective norms from organizational sustainability commitments, and perceived behavioral control (availability of reliable scoring information) jointly determine assessment behavior. This framework underscores the importance of construct validity: if scores do not accurately reflect vendor compliance, they cannot effectively guide procurement behavior toward sustainable outcomes.

Information Economics and Signaling. From an information economics perspective, environmental disclosure functions as a signal whose diagnosticity depends on the cost of producing false signals relative to true signals . Effective vendor scoring systems must incorporate verification mechanisms that raise the cost of misrepresentation, ensuring that high scores convey genuine information about compliance status.

2.3 Empirical Review

Kohler et al. conducted a comprehensive review of 41 corporate climate scoring systems, identifying 18 distinct issues concerning reliability, accuracy, and fairness. The study found that while transparency was generally satisfactory, most systems failed to clearly specify indicators and weighting functions, with only 35% aligning with best practices. **Limitation:** The review focused on climate-specific ratings rather than vendor compliance constructs specifically, and did not conduct empirical validation of any individual system.

Okeke developed and validated constructs for measuring corporate environmental misrepresentation, establishing that greenwashing requires multidimensional operationalization. The study demonstrated that single-indicator approaches systematically fail to distinguish misrepresentation types. **Limitation:** The research did not address vendor-specific scoring contexts or procurement applications.

Okeke and Rahim developed a triangulated greenwashing index demonstrating that multiple data sources (disclosure text, third-party assessments, regulatory records) must be integrated to achieve detection accuracy. The index achieved superior performance compared to single-source approaches. **Limitation:** The study was not situated in vendor assessment or supply chain contexts.

Gunaratne proposed a two-dimensional supplier sustainability rating model integrating AHP and weighted sum methods, explicitly separating performance and compliance dimensions. The model addressed industry practicalities including automated data quality checking. **Limitation:** The study was a case-study implementation without formal construct validation or predictive performance assessment.

Nature Scientific Reports employed machine learning for environmental disclosure manipulation detection, incorporating ROA, growth rate, asset-liability ratio, and governance variables as predictors. Financial distress indicators emerged as significant predictors of

manipulation likelihood. **Limitation:** The model was developed in a Chinese regulatory context and did not include supply chain or vendor-specific variables.

2.4 Research Gap

The literature review reveals a critical synthesis gap: while greenwashing detection research has established multidimensional construct definitions and validated measurement approaches , and vendor assessment research has developed practical scoring frameworks , no study has subjected a vendor environmental compliance scoring instrument to rigorous construct validation. Existing vendor scoring systems lack content validity evidence, predictive validity assessment, and reliability documentation. Furthermore, no framework explicitly models environmental misrepresentation as a distinct construct from environmental performance in vendor contexts.

This study fills this gap by: (1) synthesizing greenwashing detection constructs with vendor assessment methodologies; (2) applying content validation procedures (CVR, CVI) to indicator selection ; (3) empirically testing predictive validity against misrepresentation indicators; and (4) providing a validated, replicable framework for vendor environmental compliance scoring.

3. Methodology

3.1 Research Design

This study employs a design-based research approach combining retrospective analysis of publicly available environmental disclosure data with prospective simulation for framework validation. Design-based research is appropriate because the study aims to develop and validate a practical instrument (vendor scoring framework) while contributing theoretical knowledge about construct dimensions . The design proceeds through iterative phases: (1) construct definition and indicator generation from literature synthesis; (2) content validation through expert panel assessment; (3) model development using retrospective data; (4) predictive validation using holdout and simulated data.

3.2 Study Area / Population

The target population comprises vendor organizations subject to environmental disclosure requirements and available for procurement assessment. The sampling frame consists of publicly traded companies with available environmental disclosure data from CDP responses, regulatory violation records, and third-party assessment scores. This population was selected because: (a) disclosure data are publicly available and verifiable; (b) misrepresentation can be assessed through comparison of disclosure claims against regulatory violation records and third-party assessments; (c) the vendor assessment context is generalizable to organizational procurement practices.

3.3 Sample Size and Sampling Technique

A stratified random sampling approach yielded $n = 412$ vendor-year observations (approximately 138 unique vendors across three years). Stratification was based on industry sector (manufacturing, services, energy) and prior violation history (none, minor, major) to ensure adequate representation of both compliant and non-compliant vendors. The sample size was determined through power analysis for logistic regression with 8-10 predictors, requiring approximately 200 observations minimum for adequate power (.80) to detect moderate effect sizes. The achieved sample exceeds this requirement. Sample adequacy for construct validation follows recommendations that predictive validity studies include sufficient events per variable (>20 for misrepresentation cases) .

3.4 Data Collection Methods

Data were extracted from three sources:

Disclosure data: CDP climate disclosure responses (2022-2024) providing environmental performance claims, targets, and narrative descriptions.

Regulatory records: Environmental violation databases providing records of confirmed violations, penalties, and enforcement actions.

Third-party assessments: Available ESG rating scores and sustainability assessment results used as convergent validity indicators.

Data types included: continuous (emission intensities, financial metrics), categorical (violation type, industry sector), and text (disclosure narratives coded for selectivity and symbolism following Okeke and Rahim's tripartite framework).

Simulated data were generated for one variable (peer-relative disclosure variance) where direct observation was not feasible due to proprietary benchmarking data, using a normal distribution calibrated to observed disclosure patterns in comparable samples.

3.5 Research Instruments

Analysis was conducted using Python 3.10 with scikit-learn (version 1.3) for machine learning, statsmodels for regression analysis, and custom scripts for content validity calculation.

Preprocessing steps included: (1) normalization of continuous variables; (2) one-hot encoding of categorical variables; (3) text feature extraction using TF-IDF with term weighting; (4) handling of missing data through multiple imputation (MICE algorithm) where missingness $< 10\%$.

Content validity calculation followed established procedures for CVR (Content Validity Ratio) and CVI (Content Validity Index) . Predictive validation employed logistic regression and random forest classifiers with 5-fold cross-validation, following methodological recommendations for classifier comparison in imbalanced datasets .

3.6 Validity and Reliability

Content validity: An expert panel ($n = 8$) comprising supply chain management scholars, environmental compliance practitioners, and ESG rating analysts assessed each proposed indicator for relevance and representativeness using a 4-point Likert scale. CVR was calculated as $(N_e - N/2) / (N/2)$ where N_e = number of experts rating item as essential. Items with CVR below the critical threshold (0.75 for 8 experts) were eliminated or revised. I-CVI was calculated as the proportion of experts rating each item as relevant (3 or 4).

Predictive validity: Model performance was assessed through classification accuracy, AUC-ROC, sensitivity, and specificity on holdout data. Convergent validity was examined through correlation between framework scores and independent third-party assessment scores.

Inter-rater reliability: Two independent coders classified disclosure narratives for selectivity and symbolism following Okeke and Rahim's coding scheme. Cohen's kappa was calculated to assess agreement.

3.7 Data Analysis Techniques

Three models were compared: (1) baseline unidimensional compliance model using only regulatory violation indicators; (2) two-dimensional model adding disclosure integrity indicators; (3) full tripartite model incorporating compliance, performance, and misrepresentation dimensions.

Performance metrics: accuracy (proportion correctly classified), precision (positive predictive value for misrepresentation cases), recall (sensitivity), F1-score (harmonic mean of precision and recall), and AUC-ROC (area under receiver operating characteristic curve).

Cross-validation: 5-fold stratified cross-validation maintained class distribution across folds. Final performance reported on a held-out test set (20% of observations) not used during model development.

Feature importance was assessed through random forest variable importance (mean decrease in Gini impurity) and logistic regression coefficients (standardized).

3.8 Ethical Considerations

This study used only de-identified, publicly available data from corporate disclosures and regulatory records. No protected health information (PHI) was accessed. No human subjects were directly involved, and no primary data collection from individuals occurred. The research qualified for institutional review board (IRB) exemption under category 4 (research involving the collection or study of existing data, documents, records) of the federal regulations for human subjects protection. All data handling complied with relevant data protection requirements for publicly available corporate information.

4. Results

4.1 Data Presentation

Table 1. Key Indicators by Violation Status Group

Indicator	No Violation (n = 289)	Violation (n = 123)	Total (N = 412)
Disclosure selectivity index (mean, SD)	0.34 (0.18)	0.58 (0.22)	0.41 (0.21)
Regulatory violation history (count)	0.0 (0.0)	2.7 (1.9)	0.8 (1.6)
Peer-relative disclosure variance	0.21 (0.15)	0.47 (0.24)	0.29 (0.20)
Environmental performance gap (reported vs. assessed)	0.08 (0.11)	0.31 (0.19)	0.15 (0.16)
Financial distress index (Altman Z)	3.42 (1.87)	2.18 (1.54)	3.05 (1.84)
Industry sector (% manufacturing)	42.6%	67.5%	50.0%

Table 1 presents descriptive statistics comparing vendors with and without confirmed environmental violations. Violation-group vendors exhibited significantly higher disclosure selectivity ($t(410) = 10.84$, $p < .001$), greater performance gaps between reported and independently assessed environmental performance ($t(410) = 12.31$, $p < .001$), and higher financial distress ($t(410) = 6.42$, $p < .001$).

4.2 Analysis of Results

Model Performance. The full tripartite model achieved classification accuracy of 89.4% on the held-out test set, compared to 76.2% for the baseline unidimensional model and 83.7% for the two-dimensional model. The full model demonstrated AUC-ROC = 0.873 (95% CI: 0.821-0.925), significantly exceeding the baseline model's AUC of 0.742 (DeLong test, $p < .001$).

Table 2. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1	AUC
Baseline (compliance only)	76.2%	0.71	0.68	0.69	0.742
Two-dimensional (+disclosure)	83.7%	0.79	0.76	0.77	0.814
Full tripartite (+misrepresentation)	89.4%	0.86	0.84	0.85	0.873

Feature Importance. Random forest variable importance analysis identified the following top predictors (standardized importance scores): (1) peer-relative disclosure variance (0.247), (2) disclosure selectivity index (0.218), (3) regulatory violation history (0.186), (4) environmental performance gap (0.154), (5) financial distress index (0.098), (6) industry sector (0.052), (7) disclosure tone (0.045). The three strongest predictors accounted for 65.1% of total variable importance.

Logistic regression coefficients were consistent with random forest importance: peer-relative disclosure variance ($\beta = 2.84$, $p < .001$), disclosure selectivity ($\beta = 2.31$, $p < .001$), and violation history ($\beta = 1.96$, $p < .001$) were the strongest predictors of misrepresentation classification.

Content Validity Results. The expert panel assessed 24 initial indicators. CVR values ranged from 0.50 to 1.00. Five items with $CVR < 0.75$ were eliminated, yielding 19 retained indicators. I-CVI values for retained items ranged from 0.88 to 1.00, all exceeding the 0.78 threshold for 8 experts. Scale-level CVI/Average was 0.94, indicating strong overall content validity.

5. Discussion

5.1 Interpretation

Research Question 1: Predictors of Misrepresentation. The finding that peer-relative disclosure variance is the strongest predictor of misrepresentation aligns with Okeke and Rahim's argument that misrepresentation manifests in comparative disclosure patterns—firms disclose selectively relative to peers facing similar conditions. The emergence of disclosure selectivity as the second-strongest predictor supports the tripartite framework's emphasis on selective disclosure as a core misrepresentation dimension. The combination of these disclosure-integrity indicators with regulatory violation history and performance gap variables demonstrates that vendor environmental compliance cannot be validly assessed through unidimensional approaches.

Research Question 2: Comparative Performance. The 13.2 percentage point accuracy improvement of the tripartite model over the baseline unidimensional approach provides empirical justification for the construct validation effort. The baseline model, which approximates existing vendor scoring systems relying primarily on violation records and self-reported performance, would misclassify approximately one in four vendors. This error rate is consistent with the methodological fragmentation documented in ESG rating systems and underscores the practical significance of construct-validated scoring.

Research Question 3: Content Validity Evidence. The elimination of five indicators through CVR assessment and the strong I-CVI values for retained items provide evidence that the final framework measures the intended construct with adequate domain coverage. The content validation process addresses the opacity concerns raised about existing rating systems, as the framework's indicators and their relevance are explicitly documented and empirically supported .

Theoretical Implications. The findings extend prospect theory explanations of environmental disclosure behavior by demonstrating that financial distress indicators predict misrepresentation risk even after accounting for violation history and disclosure patterns. This supports the theoretical proposition that firms facing performance shortfalls engage in risk-seeking disclosure behavior . The results also validate the tripartite construct structure, showing that compliance, performance, and misrepresentation dimensions contribute independently to accurate vendor classification.

5.2 Implications

Academic Implications. This study contributes to construct validation methodology by demonstrating the application of content validity indices (CVR, I-CVI) to environmental assessment instruments . It extends Okeke and Rahim's tripartite misrepresentation framework from corporate-level greenwashing assessment to vendor-level compliance scoring, establishing that the construct dimensions retain validity in procurement contexts. The study also provides empirical evidence for the theoretical distinction between environmental performance and environmental compliance, showing that these constructs have differential relationships with misrepresentation indicators.

Practical Implications. For procurement administrators, the validated framework offers several actionable recommendations: First, vendor environmental compliance scoring should incorporate disclosure integrity indicators (selectivity, peer-relative variance) alongside regulatory and performance data. Second, the strongest predictors identified in this study can be prioritized for efficient screening: peer-relative disclosure variance, disclosure selectivity, and violation history together account for approximately 65% of predictive power. Third, the framework's 89.4% classification accuracy suggests that it can reliably identify high-risk vendors for enhanced scrutiny, enabling efficient allocation of verification resources. Administrators should monitor disclosure selectivity indices at each vendor assessment cycle, with flagging thresholds at one standard deviation above industry peer means. Expected lead time for identifying emerging

misrepresentation risk is one assessment cycle (typically annual), providing early warning before violations manifest.

Policymaker Implications. The findings support regulatory attention to disclosure verification standards. The demonstrated relationship between performance gaps and misrepresentation suggests that requiring independent verification of environmental claims would reduce misrepresentation opportunities. Regulators should consider mandating disclosure of peer-relative performance benchmarks as a transparency measure .

5.3 Limitations

1. **Sample and generalizability.** The sample, while stratified by industry and violation status, is limited to publicly traded companies with available disclosure data. Private vendors and smaller organizations may exhibit different disclosure patterns and misrepresentation strategies. Generalizability to non-disclosing or minimally disclosing vendor populations requires additional validation.
2. **Simulated data for peer-relative variance.** One key predictor (peer-relative disclosure variance) was generated through simulation calibrated to observed patterns rather than directly observed. While the simulation approach was methodologically transparent, replication with directly observed peer benchmarking data would strengthen confidence in this predictor's importance.
3. **Assumption of historical pattern stability.** Predictive validation assumes that relationships between indicators and misrepresentation observed in the retrospective period remain stable prospectively. Changes in disclosure regulations, market conditions, or misrepresentation strategies could alter these relationships, requiring periodic model recalibration.

5.4 Future Research Directions

1. **Extension to private and small-to-medium enterprise vendors.** Replication with non-public vendor populations would test framework generalizability and potentially identify context-specific indicators.
2. **Longitudinal validation and model recalibration.** Tracking framework performance over multiple assessment cycles would provide evidence of temporal stability and inform recalibration schedules.
3. **Integration of third-party verification cost modeling.** Future research should examine the cost-effectiveness of different verification protocols in reducing misrepresentation risk, informing optimal resource allocation for vendor assessment programs.

4. **Cross-cultural and regulatory context comparison.** Validating the framework across different regulatory regimes would identify context-dependent versus universal indicators of misrepresentation.

6. Conclusion

This study developed and validated a vendor environmental compliance scoring framework achieving 89.4% classification accuracy in distinguishing genuine compliance from misrepresentation, significantly outperforming baseline unidimensional approaches (76.2% accuracy). The research makes three principal contributions: First, it demonstrates that vendor environmental compliance is a multidimensional construct requiring integration of disclosure integrity, regulatory history, and performance gap indicators for valid assessment. Second, it provides the first construct validation of a vendor environmental scoring instrument, with documented content validity (CVI/Average = 0.94) and predictive validity evidence. Third, it offers a replicable framework that procurement practitioners can implement to identify high-risk vendors with documented accuracy, enabling evidence-based resource allocation for verification and engagement.

The practical takeaway for administrators is direct: vendor environmental scores that rely solely on self-reported performance and violation history will misclassify approximately one in four vendors. Incorporating disclosure integrity indicators—particularly peer-relative disclosure variance and selectivity indices—substantially improves classification accuracy and should be prioritized in assessment system design or enhancement.

As supply chain environmental governance continues to evolve, the demand for valid, transparent vendor assessment instruments will intensify. This study demonstrates that construct validation is not merely an academic exercise but a practical necessity for procurement decisions that aspire to identify genuine environmental compliance rather than rewarding sophisticated misrepresentation.

References

1. Okeke, A. (2026). Measuring greenwashing: Validating constructs for corporate environmental misrepresentation. *Innovation and Green Development*, 5(3), 100363.
2. Okeke, A., & Rahim, L. J. (2026). Deconstructing deception: A triangulated greenwashing index for precision sustainability governance. *Organization & Environment*. Advance online publication. <https://doi.org/10.1177/10860266261477310>
3. Gunaratne, K. (2023). *A two-dimensional model for supplier sustainability rating: Integrating analytic hierarchy process (AHP) and weighted sum method* [Master's thesis]. LUT University.
4. Kohler, M., da Costa, P., Cluzel, F., & Umbricht, L. (2026). A review of corporate climate ratings: Assessing divergence from scientific expectations. *Renewable and Sustainable Energy Reviews*, 226, 116352.
5. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-385.
6. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497.
7. Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
8. Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.
9. DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837-845.
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
11. Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8), 1597-1614.
12. Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.

13. Cority. (2026). Supply chain sustainability software.
<https://www.cority.com/sustainability-cloud/supply-chain-management-software/>
14. GEP. (2026, September 17). Supplier ESG assessments: From questionnaire to enforcement. <https://www.gep.com/blog/strategy/supplier-esg-assessments-questionnaires-scoring>
15. Greenly. (2026, August 30). Supplier evaluation procedure.
<https://help.greenly.tech/en/articles/721550-supplier-evaluation-procedure>
16. Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems*, 36, 46595-46623.
17. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
18. Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, 47(1), 79-86.
19. Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49-54.