Title of the Paper:

Unveiling the Black Box: A Transparency-Based Framework to Evaluate ESG Rating Methodologies and Their Governance Implications

Author:
Dr. Zeina Samy Mohamed
Postdoctoral Researcher
Department of Economics and Management
University of Ferrara, Italy

Email: mhmzsm@unife.it

Award Recognition:

First Place Winner – EFFAS Gasperini Award 2025
The European Federation of Financial Analysts Societies (EFFAS)
Brussels, Belgium

Unveiling the Black Box: A Transparency-Based Framework to Evaluate ESG Rating Methodologies and Their Governance Implications

Abstract

As Environmental, Social, and Governance (ESG) ratings become central to sustainable investment strategies and regulatory frameworks, concerns over their methodological transparency and consistency remain unresolved. This study offers the first systematic and critical comparative analysis of seven major ESG rating providers, Bloomberg, Refinitiv, MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global, focusing on the internal logic, disclosure practices, and evaluative coherence of their rating systems. While these scores are widely used as proxies for corporate sustainability performance, this research reveals sharp methodological divergence across key dimensions: materiality mapping, indicator weighting, score aggregation, and algorithmic disclosure.

To address this opacity, the paper introduces the Black Box Severity Index (BBSI), a novel diagnostic tool that classifies ESG rating frameworks based on four rigorously defined dimensions of transparency: aggregation model disclosure, materiality logic, replicability, and proprietary algorithm use. Grounded in peer-reviewed ESG literature and institutional accountability standards, the BBSI offers a replicable and comparative framework for evaluating the interpretive integrity of ESG scores. Empirical results reveal that no ESG rating provider achieves full transparency. For example, four agencies, MSCI, Sustainalytics, ISS ESG, and S&P Global, are classified as high-opacity, with proprietary, non-replicable architectures. The remaining providers offer partial disclosure but still fall short of auditability. This systemic opacity undermines comparability, invites regulatory arbitrage, and erodes trust in ESG scores as governance tools.

The paper provides policy-relevant solutions. It calls for mandated methodological disclosure, third-party audits of scoring systems, and the development of open-source benchmarks. These measures would enhance rating integrity, support sustainable finance policy coherence, and reduce the risk of greenwashing by proxy. The BBSI also offers practical applications for institutional investors and analysts, enabling them to evaluate ESG data pipelines' reliability and incorporate transparency risk into portfolio construction and stewardship engagement. This research reframes ESG rating divergence not as a symptom of conceptual pluralism but as a governance failure rooted in institutional opacity. Our paper proposes a structured, actionable framework that bridges academic insight with policy design and offers a roadmap for restoring trust in ESG scores as credible sustainability instruments.

Keywords: ESG ratings, black box methodology, transparency, methodological divergence, sustainability governance, investor trust.
JEL Classification: G32, M14, Q56, G18, L51, C81

Word count: 8541

1. Introduction

ESG ratings have become integral to how markets interpret sustainability-related risks and allocate capital. Across investment portfolios, regulatory frameworks, and stewardship mandates, ESG scores now function as authoritative signals of non-financial performance. According to Gibson et al. (2023), ESG data integration has become a global norm among institutional investors, influencing asset selection strategies and long-term risk modelling. Furthermore, the debate over methodological transparency in ESG ratings aligns closely with broader research on sustainability accounting and policy, which emphasizes the importance of disclosure integrity, comparability, and governance accountability (Patten & Shin, 2019; Shen et al., 2020). Prior studies in the *Sustainability Accounting, Management and Policy Journal* have highlighted both the evolution of sustainability reporting practices and the policy frameworks shaping disclosure environments, reinforcing the need for structured and enforceable transparency standards in ESG data ecosystems. However, as the influence of ESG ratings expands, so do concerns over their methodological legitimacy and transparency.

Empirical evidence reveals significant discrepancies among ESG scores provided by different rating agencies. For example, Berg et al. (2022) demonstrate that the ESG scores divergences are not marginal but stem from structural differences in scope, measurement approaches, and weighting schemes. Such inconsistency means a single firm can simultaneously be rated a sustainability leader and laggard, depending solely on the rating's internal framework. Following this critique, Christensen et al. (2022) argue that the absence of transparent scoring logic has produced a "legitimacy gap" in ESG evaluation, whereby stakeholders are expected to trust scores without insight into the logic that generates them. Similarly, Kotsantonis and Serafeim (2019) describe an "epistemic asymmetry" in which rating agencies exert significant evaluative influence while disclosing little about their methodologies. The previous literature critiques underscore that the limitations of ESG ratings are not incidental but structurally embedded.

Moreover, the opacity inherent in ESG ratings carries significant material implications. As elucidated by Berg et al. (2022), the inconsistencies associated with ESG metrics transcend theoretical concerns; they can substantially distort capital allocation processes and mislead investors regarding the sustainability credentials of various entities. At the policy level, regulators who use ESG ratings to determine eligibility for sustainable finance instruments, such as green bonds or ESG-linked indices, may inadvertently rely on classifications derived from unverifiable or conflicting assumptions. Roulet and Bothello (2020) emphasize that ESG rating agencies have become powerful intermediaries of reputational and regulatory credibility, yet they operate outside

formal systems of accountability, raising substantial concerns over the integrity of ESG as a governance mechanism.

Efforts to improve ESG data infrastructure have been initiated at the policy level, particularly within the European Union. The Corporate Sustainability Reporting Directive (CSRD), embedded within the broader European Green Deal and Sustainable Finance Strategy, mandates structured and detailed sustainability disclosures for qualifying firms. According to the European Commission (2021a), the CSRD aims to enhance data comparability, accessibility, and quality across EU member states. However, these regulations focus overwhelmingly on corporate-level reporting, while ESG rating agencies remain largely unregulated.

According to the European Commission (2021b), current regulatory frameworks do not cover the methodologies used by these agencies, particularly regarding the selection, weighting, and aggregation of ESG indicators. Moreover, this regulatory asymmetry creates a critical paradox where companies are subject to rigorous disclosure obligations, while the private entities that interpret these disclosures and translate them into scores do so under minimal oversight. Christensen et al. (2022) argue that this regulatory asymmetry undermines the policy objectives that ESG disclosure regimes seek to achieve: investor confidence, market comparability, and the institutional legitimacy of sustainability metrics. Addressing this regulatory gap is essential if ESG ratings are to serve not merely as market signals but as robust tools of sustainable policy implementation. The implications of this methodological opacity are multi-layered. Investors increasingly cannot discern which ratings reflect substantive sustainability performance versus which are shaped by favourable scoring logic.

Moreover, policymakers risk enacting finance mechanisms based on conflicting ESG assessments, potentially distorting regulatory incentives and compromising the credibility of sustainable finance frameworks. According to Drempetic et al. (2020), such opacity may incentivize firms to pursue "ESG score arbitrage," selectively aligning disclosures to the most favourable rating agency rather than driving authentic sustainability performance. This raises serious concerns about the potential for symbolic compliance and greenwashing, further complicating ESG governance effectiveness.

Our research addresses the previous ESG ratings challenges by offering a systematic and critical comparative evaluation of seven dominant ESG rating frameworks: Bloomberg, Refinitiv, MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global. Rather than focusing exclusively on rating outputs, our paper interrogates the internal methodological logic of these systems,

specifically, how each agency selects, weights, and aggregates ESG indicators and the extent to which such processes are disclosed or concealed.

To operationalize this evaluation, the research introduces the Black Box Severity Index (BBSI), a novel conceptual tool designed to assess the methodological opacity of ESG rating providers. The index is grounded in four core dimensions: aggregation model disclosure, materiality mapping transparency, replicability, and proprietary algorithm use. Following the typological methodology employed by Drempetic et al. (2020), the BBSI offers a structured and reproducible framework for assessing transparency across ESG frameworks and contributes a diagnostic tool for scholars, investors, and regulators alike.

By mapping the dimensions of methodological opacity, this study advances a more critical and institutional understanding of ESG evaluation infrastructure. The objective is not to impose convergence across scoring systems, but to render the evaluative logics that shape ESG assessments visible, making them more transparent, auditable, and accountable. Such transparency is essential for restoring investor trust, enabling policy coherence, and ensuring that ESG ratings fulfil their promise of being instruments of sustainable finance and corporate accountability.

Furthermore, this paper draws on the transparency theory, which provides a foundation for interrogating how information is structured, concealed, or revealed in organizational processes. In the context of ESG ratings, transparency refers to data disclosure by firms and the visibility of evaluative procedures by those who assess them. Following Christensen et al. (2022), our research views the absence of publicly available scoring algorithms, weighting schemes, and aggregation models as a deliberate form of opacity that undermines interpretability and trust. The BBSI's dimensions, particularly aggregation model disclosure and materiality mapping visibility, are directly informed by this theoretical lens.

The remainder of the paper is structured as follows: Section 2 provides a comprehensive literature review synthesizing prior findings on ESG rating divergence, methodological opacity, and rating credibility. Section 3 outlines the research aims, questions, and theoretical rationale for the Black Box Severity Index. Further, Section 4 presents a comparative evaluation of the seven major ESG rating frameworks across key methodological dimensions. Section 5 conceptualizes and formalizes the black box problem, while Section 6 details the design and application of the BBSI, including justification grounded in ESG evaluation theory. Section 7 presents the results and interprets the severity of opacity across providers. Section 8 concludes with policy implications, while Section 9 provides actionable recommendations for regulators and practitioners. Section 10 outlines the

study's limitations and future research directions, followed by Section 11 highlights the practical implications for financial analysts, investors, and rating providers.

2. Literature Review

The rise of ESG ratings has prompted extensive scholarly inquiry into their reliability, interpretability, and institutional legitimacy (Krueger et al., 2023). Further, ESG metrics have shifted from being peripheral to becoming integral to investment screening, corporate sustainability disclosures, and regulatory classifications. According to Gibson (2023), ESG has become a standardized tool for evaluating non-financial risk and performance across global investment portfolios. However, as ESG scores increasingly shape capital allocation and regulatory incentives, concerns regarding their methodological coherence have intensified (Berg et al., 2022).

A central theme in the academic literature is the pronounced divergence of ESG ratings across providers. For instance, Berg et al. (2022) analysed that inter-provider correlations are frequently below 0.5, even when assessing the same firms during the same period. This "aggregate confusion" is not merely a by-product of data differences but a manifestation of deeper epistemological and methodological fragmentation (Berg et al., 2022). ESG, as a construct, is interpreted through multiple philosophical lenses: some providers emphasize financial materiality, others ethical impact, and still others reputational or regulatory risk (La Torre et al., 2020; Kotsantonis & Serafeim, 2019). These conceptual divergences are operationalised through differing materiality mappings, indicator definitions, weighting schemes, and normalisation procedures. Consistent with the policy-oriented trajectory of sustainability accounting research (Patten & Shin, 2019), our study frames ESG rating opacity not merely as a measurement issue but as a governance challenge requiring policy intervention. In the context of emerging sustainability management practices, recent evidence from China underscores the interplay between national policy priorities, disclosure practices, and institutional accountability mechanisms (Shen et al., 2020), further supporting the cross-jurisdictional relevance of our proposed transparency framework.

Moreover, the technical architecture underlying ESG scoring systems is rarely detailed. According to Christensen et al. (2022), most providers do not reveal the logic of score aggregation, indicator weighting, or data transformation, effectively rendering their models opaque or "black-boxed." Investors, analysts, and regulators are thus expected to interpret ratings without access to the evaluative logic that drives them. This opacity is often justified based on commercial

confidentiality, as proprietary intellectual property (Kotsantonis & Serafeim, 2019). However, this rationale raises critical concerns about ESG accountability, particularly as these ratings influence regulatory eligibility and investor decision-making.

The governance implications of this opacity have been the subject of growing academic critique. Roulet and Bothello (2020) argue that ESG rating agencies have evolved into powerful institutional actors—mediating reputational risk, influencing regulatory compliance, and shaping market norms—yet remain exempt from the transparency standards imposed on the firms they evaluate. In this view, ESG scores are not neutral metrics, but socially constructed assessments shaped by proprietary algorithms and unexamined normative assumptions. The result is a credibility gap in the ESG ecosystem, where private evaluators exercise outsized influence without corresponding accountability mechanisms.

Several studies have attempted to diagnose these problems using score dispersion analysis, inter-agency correlation metrics, or descriptive mapping of indicator structures. For example, Drempetic et al. (2020) find that firm size influences ESG ratings, but do not examine how rating systems account for or distort such variables. Similarly, La Torre et al. (2020) assess how ESG indices affect stock performance but do not scrutinize the methodologies that generate those indices. As a result, existing literature offers important descriptive insights, yet lacks a comparative, structural framework for evaluating where and how opacity occurs within ESG scoring processes.

However, what remains underdeveloped is a theoretically informed and dimension-based methodology to assess methodological transparency across rating frameworks. Specifically, there is limited critical engagement with the computational invisibility of ESG scores, the absence of replicable weighting schemes, and the proprietary nature of score aggregation logic. Moreover, few studies explicitly explore how this opacity compromises ESG's broader function as a governance mechanism, nor how it obstructs the standardization of sustainability performance across firms and sectors.

Our paper contributes to the literature on ESG rating divergence, black-box opacity, and institutional accountability by developing a novel, operationalized framework to evaluate methodological transparency across ESG providers. Building on the work of Christensen et al. (2022), Kotsantonis and Serafeim (2019), and Roulet and Bothello (2020), this study advances the discourse by shifting the analytical focus from correlation-based diagnostics to the structural

architecture of scoring systems. It thus offers both a conceptual and applied contribution to ESG studies, bridging gaps between sustainability theory, financial evaluation, and algorithmic governance.

To this end, our research introduces the Black Box Severity Index (BBSI), a multidimensional evaluative tool grounded in the transparency literature and critiques of ESG methodology. The BBSI assesses opacity across four dimensions: aggregation model disclosure, materiality mapping visibility, score replicability, and reliance on proprietary algorithmic processes. In doing so, this study provides a systematic basis for comparing ESG frameworks and enables more precise identification of where methodological opacity erodes rating credibility.

Drawing from these theoretical and empirical insights, the paper interrogates the evaluative infrastructure of seven leading ESG rating frameworks: Bloomberg, Refinitiv, MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global. Based on the identified gaps in transparency, conceptual inconsistency, and score construction logic, the following hypothesis is proposed:

Hypothesis: The divergence in ESG ratings across major frameworks is driven by a lack of clarity in methodological approaches to issue selection, weighting, and aggregation. These factors significantly undermine the interpretability, replicability, and trustworthiness of ESG scores.

This hypothesis aims to guide the subsequent comparative analysis and provides a foundation for assessing the extent to which ESG rating providers disclose or conceal their evaluative logic.

3. Research Aims and Objectives

As ESG ratings become increasingly embedded within investment strategies, corporate disclosures, and regulatory taxonomies, their methodological integrity has emerged as both an academic concern and a policy imperative. The literature has consistently documented that ESG scores diverge substantially across rating agencies (Berg et al., 2022), and this divergence is now recognized as a systemic feature of the ESG landscape. However, what remains underexplored is a critical interrogation of the evaluative infrastructures underpinning ESG ratings—specifically, the extent to which scoring models are transparent, reproducible, and methodologically coherent.

According to Christensen et al. (2022), this divergence cannot be addressed through better data alone; it requires scrutiny of the scoring architectures and the computational logic that transforms ESG indicators into final ratings. Nevertheless, as highlighted in previous sections, ESG rating providers often operate through proprietary, opaque models, limiting external scrutiny and compromising stakeholder confidence (Kotsantonis & Serafeim, 2019; Roulet & Bothello, 2020). These challenges raise important questions about ESG's function as a governance instrument: How can ESG ratings support capital allocation, regulatory design, or stakeholder accountability when their construction remains concealed?

In response, our study seeks to critically evaluate the methodological transparency of ESG rating frameworks by constructing a novel diagnostic framework—the Black Box Severity Index (BBSI). This index assesses the severity of methodological opacity across four rigorously defined dimensions: (1) aggregation model disclosure, (2) materiality mapping transparency, (3) score replicability, and (4) proprietary algorithm use. The BBSI thus enables a systematic, replicable, and comparative evaluation of ESG frameworks, grounded in both transparency theory and ESG evaluation literature.

Our research aims to advance a structural understanding of how ESG scores are constructed, where opacity resides, and how this opacity influences rating divergence and stakeholder trust. Rather than treating ESG scores as outputs to be statistically correlated or back-tested, the study adopts a conceptual lens grounded in epistemic governance—an approach that views ESG ratings as institutional artefacts shaped by hidden choices, competing logics, and accountability gaps (Christensen et al., 2022; La Torre et al., 2020).
Accordingly, the study is guided by the following research questions:

1. How do major ESG rating providers differ in their methodological treatment of materiality mapping, indicator selection, issue weighting, and score aggregation?
2. Where does methodological opacity manifest within ESG rating frameworks, and how does this affect the transparency and interpretability of ESG scores?
3. To what extent does methodological opacity contribute to ESG score divergence, and what are the implications for investor confidence, regulatory reliability, and the legitimacy of ESG as a governance tool?

The previous research questions extend the existing literature in several critical ways. First, they shift the analytical focus from outcome-level dispersion to the internal mechanics of rating construction. Moreover, the questions develop a dimension-based typology, embodied in the BBSI, which allows opacity to be measured, compared, and problematized across frameworks. Third, they create space for actionable policy and institutional reform, offering concrete criteria by which regulators (e.g., ESMA, EFRAG, and IOSCO) can evaluate whether ESG rating methodologies are fit for integration into green taxonomies, public investment eligibility, or fiduciary governance practices.

In treating ESG ratings not as neutral metrics but as contested instruments of institutional power, this research aligns with emerging calls for transparency, auditability, and methodological pluralism within sustainable finance. Ultimately, it seeks to reframe ESG rating legitimacy not around brand reputation or adoption rate, but around methodological integrity, offering a replicable, theoretically grounded, and policy-relevant approach to restoring trust in ESG as a credible evaluative regime.

4. Comparative Review of ESG Rating Frameworks

Building on the research objectives in section 3, this section applies a diagnostic, comparative analysis of seven leading ESG rating frameworks—Bloomberg, Refinitiv, MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global. The analysis focuses on evaluating methodological transparency, not score outcomes. This distinction is critical: while many studies examine the correlation between ESG ratings (Berg et al., 2022), few interrogate how those scores are constructed or where opacity is systematically embedded.

To operationalize this analysis, the paper applies the Black Box Severity Index (BBSI), a diagnostic framework based on four evaluative dimensions: Aggregation model disclosure, materiality mapping transparency, replicability, and use of proprietary algorithms. Each of these dimensions is grounded in established literature and reflects core institutional and policy concerns regarding the credibility of ESG evaluations (see Tables 1 and 2). For instance, non-disclosure of score aggregation mechanisms compromises interpretability (Christensen et al., 2022); opaque materiality logic limits cross-sector comparability (La Torre et al., 2020); lack of replicability reinforces epistemic asymmetry (Kotsantonis & Serafeim, 2019); and proprietary algorithm use undermines institutional accountability (Roulet & Bothello, 2020).

The frameworks analysed below are evaluated using BBSI criteria, based on publicly available methodological reports, academic reviews, and regulatory disclosures. Table 1 synthesizes the comparative results and supports the categorization of each provider's black box severity as low, moderate, or high.

4.1 Bloomberg ESG Scores

Bloomberg's ESG framework is designed around a rules-based scoring system that emphasizes financial materiality. According to Bloomberg (2022), ESG scores are disaggregated into three pillars, Environmental (E), Social (S), and Governance (G), which are weighted according to sector-specific risk relevance. Each pillar is scored on a scale of 1 to 5, with the final ESG score derived using a power mean aggregation method, allowing pillar scores to be weighted non-linearly. Bloomberg's methodology is grounded in industry-specific issue prioritization based on internal fundamental research and materiality heat maps (Bloomberg, 2022). Nevertheless, while Bloomberg discloses it's scoring pillars and industry guides, it does not reveal the complete set of indicators, weightings, or the computational formula used in score aggregation (La Torre et al., 2020). Moreover, the Bloomberg partial disclosure undermines replicability and situates Bloomberg within a moderate black box severity category (Bloomberg, 2022; La Torre et al., 2020).

4.2 Refinitiv (LSEG)

Refinitiv's ESG scoring model is among the most data-intensive frameworks, covering over 630 data points across 10 categories. Its emphasis lies on objective and publicly reported data, resulting in lower reliance on subjective analysis. The ESG score is split into three components (E, S, and G), with an additional controversy overlay score that adjusts the base score in response to adverse events (Refinitiv, 2023). Unlike Bloomberg, Refinitiv ranks companies relative to industry peers on a percentile basis. Further, the methodological architecture is more transparent, as category weights and data point coverage are disclosed. However, the scoring algorithms, adjustment thresholds, and logic for integrating controversy remain undisclosed (Christensen et al., 2022). Therefore, while the framework is highly granular, it remains partially opaque in how qualitative and quantitative data are reconciled. Therefore, Tables 1 and 2 show that LSEG is moderately opaque.

4.3 MSCI ESG Ratings

MSCI's ESG methodology is predicated on a relative risk model that evaluates how well a firm manages ESG risks relative to peers within the same industry. Its ratings span from AAA (leader) to CCC (laggard). MSCI uses a forward-looking approach, emphasizing exposure to financially relevant ESG risks and the firm's ability to manage them. Sectoral materiality maps guide the weighting of issues, which vary by industry (MSCI, 2023). The transparency of MSCI's framework is limited; while its pillars and rating scale are publicly available, the quantitative thresholds, issue-specific weights, and aggregation models are not (Berg et al., 2022).

Moreover, MSCI leverages proprietary scores based on reported and estimated data, making replicating and validating its assessments difficult. As such, MSCI's model is widely seen as one of the most black-boxed among major ESG providers (Christensen et al., 2022). Nevertheless, the MSCI approach is highly proprietary, as materiality mappings, score thresholds, and aggregation weights are not disclosed. Further, MSCI incorporates modelled and estimated data, which cannot be externally verified. These features contribute to MSCI's classification as high black box severity (MSCI, 2023; Berg et al., 2022).

4.4 Sustainalytics

Sustainalytics, a Morningstar company, adopts a quantified risk exposure model that calculates an "ESG Risk Rating" for companies based on unmanaged risk, the portion of ESG exposure that remains unmitigated by management practices. This model allows for a sector-agnostic comparison across firms, unlike MSCI's peer-relative approach. Sustainalytics publishes methodology guides and indicates how risk categories are defined, but the criteria for assigning severity levels, adjustment mechanisms, and thresholds for risk classification remain undisclosed. Moreover, according to Drempetic et al. (2020), the score often includes controversies and governance issues that are not weighted or explained, limiting transparency. Although Sustainalytics positions itself as a methodologically rigorous entity, its black-box components significantly limit score reproducibility.

Sustainalytics assesses unmanaged ESG risk, defined as exposure not mitigated by a firm's management systems. While its approach enables cross-sector comparisons, severity classification, scoring logic, and aggregation models remain concealed. The opacity of the risk quantification and issue weighting mechanisms restricts transparency and score auditability, resulting in a high severity rating (Sustainalytics, 2022; Drempetic et al., 2020).

4.5 ISS ESG

ISS ESG adopts a dual evaluation model, combining performance indicators with normative screening criteria. Its Corporate Rating assesses ESG performance and alignment with international standards such as the UN Global Compact. The scoring is structured around sector-specific key performance indicators (KPIs) and employs an "absolute best-in-class" rating logic. While ISS ESG does disclose issue areas and high-level rating criteria, the specific scoring rubrics, internal scoring algorithms, and weighting methodologies are not publicly available. Their methodology exhibits high opacity and limited replicability (ISS ESG, 2022; Roulet & Bothello, 2020). Furthermore, its integration of qualitative assessments and client-specific custom scoring introduces subjective elements that are not systematically disclosed. The previous critical analysis makes ISS ESG's ratings difficult to audit or interpret externally.

4.6 FTSE Russell ESG Ratings

FTSE Russell uses a themes-based model comprising 14 ESG themes and over 300 indicators. Scores are generated at the pillar and thematic levels, with companies evaluated relative to sector peers. The framework is updated biannually and emphasizes alignment with global standards such as SASB and TCFD. Compared to peers, FTSE Russell is moderately transparent—it publishes theme-level weightings and offers details on data sources. However, the process of translating raw data into final numeric scores remains proprietary, and adjustments for missing data or controversies are not disclosed. Thus, while the thematic breadth is vast, the transformation logic of data into ratings remains obscured (FTSE Russell, 2023; La Torre et al., 2020). The framework demonstrates moderate transparency and black box severity but retains key proprietary elements (FTSE Russell, 2023; La Torre et al., 2020).

4.7 S&P Global ESG

&P Global ESG framework is derived from its Corporate Sustainability Assessment (CSA), a detailed survey-based approach that includes over 1,000 data points for eligible companies. Moreover, the framework emphasizes direct corporate engagement and uses company-reported, publicly available, and analyst-assessed data (S&P Global, 2022). Furthermore, S&P's ESG scores are comprehensive and considered among the most robust, yet the score aggregation model, weighting schemes, and risk normalization techniques are not publicly disclosed. Although the CSA questionnaire is extensive, its conversion into scoring remains partially opaque. Moreover, access to full methodology details often requires commercial agreements, limiting public scrutiny,

hence placing the framework in the high-opacity category according to Figure 1 (S&P Global, 2022; Christensen et al., 2022).

4.8 Summary and Link to the BBSI Framework

The comparative analysis confirms that none of the major ESG rating frameworks fully meet the transparency standards expected of institutions with public governance roles. While Refinitiv, FTSE Russell, and Bloomberg offer partial transparency, all seven frameworks employ proprietary elements that restrict replicability and limit scrutiny. In line with Table 2, this systematic opacity undermines ESG's function as a governance tool and creates policy asymmetries, where firms are subject to prescriptive disclosure regimes (e.g., CSRD), while their evaluators remain unregulated.

Table 1 presents a synthesized classification of each provider based on the BBSI dimensions, critically supported by the transparency theory and the literature review. These findings provide the empirical foundation for Section 5, which builds on this framework to formally define the Black Box Severity Index and critically analyse its implementation as a practical and policy-relevant tool for ESG accountability.

Table 1 is about here.

Table 2 is about here.

5. Operationalizing ESG Methodological Opacity: The Case for the BBSI Framework

The comparative review presented in Section 4 reveals a pervasive pattern: none of the seven leading ESG rating providers demonstrate complete methodological transparency. Despite their centrality in corporate sustainability governance, these providers maintain evaluative systems that are either partially or wholly opaque. This structural opacity, called the ESG "black box" problem, limits sustainability assessments' interpretability, replicability, and accountability. The resulting governance asymmetry, wherein regulated firms are subject to increasingly stringent reporting requirements while raters themselves operate without equivalent oversight, has been highlighted by Christensen et al. (2022) and Kotsantonis and Serafeim (2019) as a core institutional risk.

Indeed, opacity is not a by-product of conceptual pluralism but a systemic feature of ESG infrastructure design. The undisclosed nature of aggregation models, materiality logic, and proprietary scoring algorithms generates what Kotsantonis and Serafeim (2019) term "epistemic asymmetry", a condition wherein private rating agencies function as de facto regulators without

transparent methodologies or accountability mechanisms. The previous analysis creates a structural paradox where firms are obliged to comply with robust sustainability regulations such as the CSRD, SFDR, or the EU Taxonomy, yet the metrics used to evaluate compliance remain unregulated and non-replicable (European Commission, 2021a; Roulet & Bothello, 2020).

In response to this critical institutional gap, the present study introduces the Black Box Severity Index (BBSI)—a conceptual and operational tool designed to diagnose the degree of methodological opacity across ESG rating providers. Building on literature from transparency theory and algorithmic governance (Berg et al., 2022; Christensen et al., 2022), the BBSI evaluates each provider using four core dimensions: aggregation model disclosure, materiality mapping transparency, score replicability, and proprietary algorithm use. These dimensions are theoretically justified and empirically derived, as detailed in Table 2.

Each dimension is scored on a 1–3 ordinal scale and averaged to compute a Black Box Severity Score (BSS). This score enables a comparative, reproducible classification of methodological opacity: a score below 2.0 reflects low opacity, between 2.0 and 2.74 indicates moderate opacity, and 2.75 or above denotes high opacity. This scoring logic reflects typological methodologies used in financial evaluation and institutional accountability research (Drempetic et al., 2020; Roulet & Bothello, 2020). Thus, the BBSI enables stakeholders to distinguish between frameworks that exhibit transparent evaluative logic and those whose scores derive from concealed, non-auditable architectures.

Moreover, the BBSI offers institutional utility beyond academic critique. For policymakers, BBSI provides a diagnostic benchmark for evaluating whether ESG scores meet baseline standards of transparency before being embedded in taxonomies, stewardship codes, or sustainable finance legislation. Moreover, for practitioners, BBSI offers a means to assess whether the ESG scores used in investment decisions or risk models are methodologically sound and accountable. Consequently, the BBSI bridges conceptual literature with applied governance tools and contributes to an emerging policy discourse that demands greater transparency in ESG evaluation (IOSCO, 2021; ESMA, 2022).

6. Critical Methodological Divergence across Frameworks

6.1. Methodological Justification for Black Box Severity Index Design

This study introduces the Black Box Severity Index (BBSI), a composite evaluative tool designed to classify the methodological opacity of ESG rating providers. Consistent with definitions of evaluative indices in financial and sustainability analysis (La Torre et al., 2020; Berg et al., 2022), the BBSI integrates four rigorously selected dimensions: aggregation model disclosure, materiality mapping transparency, replicability, and use of proprietary algorithms. These dimensions are drawn from transparency theory, sustainability accountability literature, and prior literature critiques of ESG methodological opacity (Christensen et al., 2022; Roulet & Bothello, 2020).

Each dimension is evaluated on a three-tier ordinal scale (Low = 1, Moderate = 2, High = 3), allowing for provider-level comparison while ensuring reproducibility and policy applicability. Thus, the BBSI framework fills a key methodological gap by offering a standardized yet flexible diagnostic tool that supports academic critique and institutional reform.

The design of the *Black Box Severity Index* is methodologically anchored in both empirical scholarship and conceptual models that address the opacity and inconsistency of ESG rating systems. While prior literature has examined divergence in ESG scores across providers (Berg et al., 2022; Christensen et al., 2022), few studies have proposed structured evaluative typologies that pinpoint the sources and severity of methodological opacity. Our paper addresses that gap by developing a composite index based on four core dimensions: aggregation model disclosure, materiality mapping, replicability, and use of proprietary algorithms. Each criterion is selected based on its presence in academic frameworks assessing transparency and credibility in sustainability ratings.

6.1.1. Aggregation Model Disclosure

The decision to include aggregation disclosure as a primary dimension is grounded in methodological concerns highlighted by Christensen et al. (2022), who argue that “the lack of visibility into how ESG inputs are mathematically aggregated is a principal source of institutional distrust.” ESG rating models rely on proprietary weighting and normalization schemes, which significantly influence final scores yet remain unobservable to external stakeholders. Furthermore, including this criterion directly responds to the call for greater visibility into how qualitative and quantitative indicators are synthesized (Kotsantonis & Serafeim, 2019).

6.1.2. Materiality Mapping Disclosure

While Materiality mapping, often cited as a core strength of ESG frameworks, is inconsistently implemented and disclosed. Gibson et al. (2023) and La Torre et al. (2020) show that some providers define materiality based on financial risk, while others use broader ethical or sectoral lenses. This lack of harmonization drives score divergence and raises critical concerns about interpretive validity. Therefore, the index includes materiality disclosure as a criterion to evaluate whether users can understand which issues are deemed relevant and why.

6.1.3 Replicability of ESG Ratings

Replicability is a foundational principle of scientific and financial evaluation, yet it is rarely achieved in ESG ratings due to opaque scoring architectures. As Berg et al. (2022) demonstrate, ESG scores are typically non-replicable even when raw data is available, because the transformation logic remains proprietary. Drawing on Roulet and Bothello (2020), our paper considers replicability essential for rating credibility, especially when such ratings inform regulatory classifications, capital allocations, and fiduciary investment decisions.

6.1.4 Use of Proprietary Algorithms

Using proprietary algorithms represents the most significant epistemic barrier to transparency in ESG evaluations. According to Eccles and Klimenko (2019), the commercialization of ESG methodologies has created an "intellectual property arms race" that prioritizes competitive advantage over public accountability. This criterion evaluates whether the core scoring mechanisms, beyond raw indicator lists, are accessible or locked behind paywalls and commercial licenses, a concern also raised by Drempetic et al. (2020).

To structure this assessment systematically, each provider is evaluated against these four dimensions using a three-tier classification: Low (1), Moderate (2), and High (3) opacity. This scoring logic aligns with typological methodologies commonly used in institutional finance and policy assessment (Berg et al., 2022; La Torre et al., 2020). The framework enables cross-provider comparison and a reproducible audit of methodological transparency, an aspect that existing ESG research has largely overlooked.

In developing this index, our research combines critical content analysis of providers' methodological documentation with an evaluation framework grounded in transparency theory, accountability literature, and best practices in sustainability reporting evaluation. The result is a

structured, literature-validated metric highlighting where and how ESG ratings obscure their evaluative logic.

6.2. Black Box Severity Assessment Criteria

Building on the conceptual and methodological foundation established in Section 6.1, this section applies the Black Box Severity Index (BBSI) to seven leading ESG rating providers. The BBSI offers a replicable framework through which each provider's degree of methodological opacity is classified along four rigorously defined dimensions: aggregation model disclosure, materiality mapping transparency, replicability, and use of proprietary algorithms. Each dimension is scored on a 1–3 ordinal scale, with lower values indicating greater methodological transparency and replicability. The average of these four scores yields the Black Box Severity Score (BSS), used to categorize each provider as exhibiting Low, Moderate, or High opacity.

This scoring method is grounded in the literature on rating system design, transparency theory, and sustainability governance (Berg et al., 2022; Christensen et al., 2022; Drempetic et al., 2020). It responds to the methodological vacuum identified in earlier comparative ESG research, which has often focused on divergence without evaluating its structural causes. By contrast, the BBSI enables intra- and inter-provider comparison, while offering a theoretically and empirically justified basis for evaluating rating credibility and institutional accountability.

Following Drempetic et al. (2020) and Christensen et al. (2022), we calculated each provider's Black Box Severity Score (BSS) using the following composite formula:

$$BSS_i = \frac{1}{n} + \sum_{j=1}^{n} D_{ij} \quad (1)$$

Where BSS_i represents the average opacity score for provider i, D_{ij} is the assigned score for dimension j, and n reflects the number of evaluative dimensions. Each dimension is scored as 1 (transparent), 2 (partially transparent), or 3 (opaque). This quantitative structure enables cross-provider comparison and aligns with typological methodologies used in institutional governance assessment (Roulet & Bothello, 2020).

Severity classifications are assigned according to the following criteria:

$$Severity\ Classification = \begin{cases} Low, & if\ BSS_i < 2.0 \\ Moderate, & if\ 2.0 \leq BSS_i < 2.75 \\ High, & if\ BSS_i \geq 2.75 \end{cases} \quad (2)$$

This formal calculation in equation 2 structure enhances the investigation's methodological transparency and supports the reproducibility of BBSI classifications across future applications.

The BSS scoring mirrors and follows the practices in typological methodology (Drempetic, Klein, & Zwergel, 2020), transparency evaluation (Christensen et al., 2022), and institutional accountability frameworks (Roulet & Bothello, 2020).

6.3 Black Box Severity Application

Building on Section 6.2, the BBSI is applied to seven leading ESG rating agencies: Bloomberg, Refinitiv (LSEG), MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global. Each agency is assessed across the four dimensions, scored on a scale of 1 (transparent) to 3 (opaque). The arithmetic mean of these scores yields the Black Box Severity Score (BSS), which is then used to assign the following classifications:

The classification criteria used in this analysis are summarized as follows:

High Black Box Severity (Score $\geq$ 2.75): Providers fall into this category if most of their methodological components are undisclosed, proprietary, or non-replicable. These frameworks typically withhold score aggregation models, employ undisclosed weighting mechanisms, and rely heavily on algorithmic opacity. This group includes MSCI, Sustainalytics, ISS ESG, and S&P Global—all of which present minimal transparency on how ESG indicators are selected, transformed, or synthesized.

Moderate Black Box Severity (Score between 2.0 and 2.74): These providers demonstrate partial transparency in certain areas (e.g., disclosing issue themes or some weighting structures) but stop short of publishing aggregation formulas or adjustment thresholds. Bloomberg, Refinitiv, and FTSE Russell fall into this classification, offering more granular methodological information than their high-opacity counterparts but maintaining proprietary scoring architectures that limit replicability and auditability.

Low Black Box Severity (Score < 2.0): No provider in the current sample achieves this threshold. This absence is analytically significant, underscoring a systemic lack of complete transparency in the ESG rating industry and reinforcing prior critiques that ESG ratings operate within a regime of asymmetrical accountability (Christensen et al., 2022; Roulet & Bothello, 2020).

These results are summarised in Table 3. Notably, no provider demonstrates comprehensive transparency, indicating that opacity is structurally embedded in the ESG rating industry. MSCI,

Sustainalytics, ISS ESG, and S&P Global fall into the High severity group, exhibiting complete opacity in aggregation, replicability, and score logic. Bloomberg, Refinitiv, and FTSE Russell demonstrate moderate transparency, sharing partial documentation while maintaining proprietary models. These findings confirm the study's core hypothesis: that ESG score divergence is not merely a function of philosophical pluralism, but of institutional opacity and unregulated methodological discretion (Christensen et al., 2022; Berg et al., 2022).

The results of this classification are presented in Table 2, which integrates provider-level BSS scores and severity rankings. These findings form the evidentiary foundation for the critical analysis in Section 7, where the implications of methodological opacity regarding investor trust, market signal distortion, and regulatory risk are examined.

7. Results

This section presents the empirical results of applying the Black Box Severity Index (BBSI) to seven leading ESG rating frameworks. The BBSI evaluates each provider across four diagnostic dimensions: aggregation model disclosure, materiality mapping transparency, replicability, and use of proprietary algorithms, yielding a composite Black Box Severity Score (BSS). This composite classification places each provider into one of three categories: Low, Moderate, or High opacity.

Table 3 summarises the results. Crucially, no provider achieved a Low-opacity classification (BSS < 2.0). This finding supports the central hypothesis of this study: that methodological opacity is structurally embedded across the ESG ratings industry. Four providers, MSCI, Sustainalytics, ISS ESG, and S&P Global, fall within the High opacity category (BSS ≥ 2.75), reflecting consistently low transparency across all four evaluative dimensions. These providers offer high-level thematic or issue coverage but fail to disclose their computational scoring architecture, making their ESG evaluations non-replicable, unauditable, and effectively inscrutable (MSCI, 2023; Sustainalytics, 2022; Roulet & Bothello, 2020).

In contrast, Bloomberg, Refinitiv (LSEG), and FTSE Russell demonstrate Moderate opacity. These providers exhibit partial methodological disclosure, such as sector-specific materiality maps or controversy themes, but continue withholding score aggregation formulas, weighting schemes, or adjustment thresholds. For example, Bloomberg publishes scoring dimensions and sectoral guidance but does not disclose its power mean aggregation process, rendering its overall score construction opaque (Bloomberg, 2022; La Torre et al., 2020). Similarly, Refinitiv discloses data

categories and peer benchmarking procedures while keeping controversy scoring logic confidential (Refinitiv, 2023).

The absence of any fully transparent provider is analytically significant. It validates the claim that opacity is not a result of isolated shortcomings but a systemic and institutionalized feature of ESG scoring infrastructures. These findings echo and extend the critiques advanced by Christensen et al. (2022) and Berg et al. (2022), who argue that ESG rating divergence is driven not merely by differing philosophical frameworks but by concealed evaluative mechanisms that hinder interpretability, comparability, and regulatory accountability.
Table 3 is about here.

7.1 Visualizing Opacity: The Black Box Radar Chart

Figure 1 presents a radar chart comparing the seven providers across the four BBSI diagnostic dimensions and an overall transparency indicator to substantiate and visually consolidate these findings. The chart enables a multidimensional diagnostic lens through which institutional opacity is made visually and comparatively explicit.
Figure 1 is about here.

Each axis of the radar chart represents a core dimension of opacity: aggregation model disclosure, materiality mapping transparency, replicability, proprietary algorithm use, and an aggregate transparency level. The size and shape of each polygon reflect each provider's relative opacity. Polygons closer to the centre denote greater transparency, while expanded shapes reaching outer zones indicate systemic black-box severity.

Three critical patterns emerge:
First, the High-opacity providers—MSCI, Sustainalytics, ISS ESG, and S&P Global—display expansive shapes, reflecting opacity across all evaluative dimensions. These firms disclose little information about their aggregation logic, materiality mappings, or algorithmic mechanisms. Their methodologies operate as closed epistemic systems, inaccessible to investors, regulators, and rated entities (Berg et al., 2022; Roulet & Bothello, 2020).

Moreover, the Moderate-opacity providers—Bloomberg, Refinitiv, and FTSE Russell—show more compact but irregular polygons. Bloomberg, for instance, reveals some materiality logic but retains opacity in score computation. These providers offer partial transparency that, while improved, remains insufficient for auditability or regulatory use (Refinitiv, 2023; FTSE Russell, 2023). In

addition, no provider approaches the chart's inner zone, reinforcing the conclusion that opacity is a structural feature rather than a circumstantial limitation. The visual confirms that even the most transparent providers fall short of disclosing sufficient methodological detail to meet minimum governance or audit thresholds.

7.2 Implications of the Radar Analysis

The radar chart serves as a visual aid and diagnostic tool with concrete policy and practice implications. It translates abstract methodological critiques into a clear, comparative framework that regulators, investors, and institutional data users can utilize to benchmark transparency and accountability (Saltelli & Funtowicz, 2017; IOSCO, 2021; ESMA, 2022). For regulators, the radar chart offers a baseline to evaluate which ESG rating frameworks are suitable for incorporation into sustainable finance taxonomies, ESG-linked instruments, or stewardship strategies (IOSCO, 2021; ESMA, 2022). For investors, it helps identify which providers introduce the highest uncertainty into ESG-integrated financial models.

Furthermore, the visualization reaffirms that divergence in ESG scores arises not solely from conceptual pluralism, such as risk-based versus values-based approaches, but from unacknowledged methodological discretion. When left unregulated, this discretion threatens the integrity of ESG data flows within financial and policy ecosystems.

In conclusion, the BBSI and its visual representation expose the evaluative asymmetries that underpin ESG score construction. This section's findings substantiate the paper's broader argument: that opacity in ESG ratings is not a peripheral challenge but a core governance risk. The following section translates these insights into actionable recommendations for regulatory intervention, institutional reform, and investor engagement.

8. Discussion

The findings of this study confirm and extend concerns raised in the literature regarding the opaque evaluative role of ESG rating providers. As argued by Roulet and Bothello (2020), these agencies increasingly function as unregulated epistemic authorities, issuing scores that influence capital flows, reputational outcomes, and regulatory classifications, while remaining exempt from the transparency and auditability requirements imposed on the firms they evaluate. The Black Box Severity Index (BBSI) exposes a structural asymmetry: firms must comply with prescriptive disclosure regimes like CSRD and SFDR, while rating agencies operate without equivalent

methodological oversight or auditability. As ESG ratings increasingly inform instruments such as the EU Taxonomy, green bond eligibility, and ESG-linked credit assessments, this opacity threatens the integrity of sustainable finance architecture (European Commission, 2021a; Christensen et al., 2022).

Critically, the BBSI clarifies the distinction between conceptual divergence—differences in ESG philosophy—and procedural opacity, i.e., the concealment of scoring mechanisms and algorithmic weighting. Moreover, the radar chart in Figure 1 and the classifications in Table 3 demonstrate that the dominant driver of ESG score divergence is not conceptual disagreement but hidden methodologies that prevent score replication and undermine interpretability. This dynamic reinforces what Kotsantonis and Serafeim (2019) term "epistemic asymmetry," where users must rely on ESG assessments they cannot verify or deconstruct. This dynamic reinforces what Kotsantonis and Serafeim (2019) term "epistemic asymmetry," where users must rely on ESG assessments they cannot verify or deconstruct.

While the former may reflect pluralistic governance values, the latter undermines investor trust, comparability, and regulatory reliability (Kotsantonis & Serafeim, 2019). This diagnostic framework helps explain the low correlations observed across ESG scores, a concern cited by institutional investors as a barrier to effective ESG integration (Gibson et al., 2023), and offers methodological remedies rooted in transparency theory.

In response to the three research questions posed in Section 3, this study offers the following evidence-based conclusions. First, the comparative review documents pronounced heterogeneity in how materiality is defined, indicator weighting, and aggregation across rating providers, from rules-based systems like Bloomberg to risk-weighted peer models like MSCI (La Torre et al., 2020; Berg et al., 2022). Second, the BBSI scoring reveals that opacity is especially prevalent in disclosing aggregation methods and proprietary algorithm use, confirming prior claims of systematic non-replicability (Christensen et al., 2022). Third, and most importantly, this research shows that opacity is a structural cause of ESG score divergence, with significant implications for investor trust, regulatory reliability, and market signal integrity (Eccles & Klimenko, 2019).

From a policy and market perspective, the BBSI is a replicable tool for regulators (e.g., ESMA, IOSCO, EFRAG) to determine whether ESG rating methodologies meet minimum transparency thresholds for sustainable finance regimes. It also enables investors, asset managers, and analysts to evaluate the robustness and reliability of ESG scores before using them in capital allocation or stewardship decisions. In this respect, the BBSI bridges the gap between academic critique and

institutional implementation to offer a policy-relevant, literature-validated instrument for strengthening the legitimacy of ESG assessments as tools of market governance.

9. Conclusion

This research has undertaken a comprehensive, literature-grounded, and methodologically structured evaluation of seven leading ESG rating frameworks—Bloomberg, Refinitiv, MSCI, Sustainalytics, ISS ESG, FTSE Russell, and S&P Global. While these providers differ in issue emphasis, industry focus, and terminological framing, they converge in one critical respect: all operate with varying degrees of methodological opacity. ESG scores, which are widely treated as proxies for corporate sustainability, are frequently constructed using undisclosed aggregation models, inaccessible weighting systems, and algorithmic processes shielded by commercial confidentiality.

To interrogate this opacity systematically, the paper introduces the Black Box Severity Index (BBSI), a novel evaluative framework that classifies ESG providers based on four transparency-related dimensions: aggregation model disclosure, materiality mapping visibility, replicability, and algorithmic openness. This index enables both comparative assessment and conceptual clarity by transforming critiques of opacity into a structured, measurable framework grounded in the literature on financial transparency, sustainability reporting, and epistemic governance (Berg et al., 2022; Christensen et al., 2022; Kotsantonis & Serafeim, 2019).

The results are unequivocal. None of the seven ESG providers achieved a low-opacity rating. Four—MSCI, Sustainalytics, ISS ESG, and S&P Global—fall within the *high* severity category, revealing deeply concealed evaluative architectures. Bloomberg, Refinitiv, and FTSE Russell *exhibit* moderate opacity, disclosing some elements of their methodology while retaining core scoring mechanisms as proprietary. These findings confirm that methodological opacity is not an aberration but a structural condition of the ESG rating landscape, with serious consequences for regulatory integrity, capital market comparability, and institutional trust.

Furthermore, this paper contributes to the literature on ESG rating divergence and black-box opacity by operationalizing a novel transparency framework that shifts the analytical lens from outcome-level correlation to the structural logic of scoring systems (Christensen et al., 2022;

Kotsantonis & Serafeim, 2019; Roulet & Bothello, 2020). In doing so, it bridges theoretical debates on sustainability governance with applied diagnostic tools for institutional accountability.

Critically, this research reframes the ESG divergence debate. Score discrepancies do not simply result from differing normative values or philosophical perspectives. Instead, they stem from an institutionalized absence of transparency, a black box problem that impedes meaningful evaluation, stakeholder interpretation, and regulatory scrutiny. Furthermore, the consequences are far-reaching: ESG scores, in their current form, are used to construct portfolios, define green finance eligibility, influence stock valuations, and guide public policy, often without clear accountability for how these scores are derived.

By making the internal logic of ESG scoring systems visible, this study contributes to a more rigorous and transparent understanding of ESG as a form of private sustainability governance. It advances the debate beyond descriptive comparisons and toward a replicable, academically grounded diagnostic framework that informs academic research and regulatory oversight. The BBSI thus serves not merely as a classificatory device but as a foundational step toward the standardization and democratization of ESG methodologies.

This paper's findings carry urgent practical implications in a regulatory environment increasingly attentive to greenwashing, ESG manipulation, and algorithmic opacity. Moreover, our research offers a pathway for policymakers and regulators. For ESG ratings to be considered legitimate governance and capital allocation tools, they must be auditable, interpretable, and accountable. This research outlines a way to achieve those objectives.

10. Policy Recommendations

The findings of this study offer urgent implications for regulatory bodies seeking to ensure that ESG scores function as reliable instruments of sustainable finance. The Black Box Severity Index (BBSI) reveals a structural opacity compromising transparency, comparability, and interpretability across ESG rating providers. In light of this, several regulatory actions are necessary. First, institutions such as the European Securities and Markets Authority (ESMA), the U.S. Securities and Exchange Commission (SEC), and the International Organization of Securities Commissions (IOSCO) should implement mandatory transparency standards that require rating providers to publicly disclose their aggregation models, indicator weighting, and materiality mapping

methodologies (IOSCO, 2021; ESMA, 2022). Such reforms must go beyond superficial disclosure, mandating the release of computational logic and adjustment thresholds, thereby ensuring auditability and minimizing greenwashing risk (Christensen et al., 2022; Kotsantonis & Serafeim, 2019).

Second, third-party verification of rating methodologies should be institutionalized. As financial statements are subject to external assurance, ESG ratings should be subject to standardized audits that verify methodological consistency and replicability (Eccles & Klimenko, 2019; Roulet & Bothello, 2020). Moreover, global standard-setting bodies—including the IFRS Foundation and EFRAG—should coordinate to establish a minimum baseline of ESG scoring transparency. These standards could draw from emerging frameworks such as the EU's Corporate Sustainability Reporting Directive (CSRD) and the SEC's proposed ESG disclosure rules. Without such intervention, the asymmetry between rated entities and rating providers will continue to undermine the credibility of sustainable finance instruments and ESG-aligned regulatory mechanisms (European Commission, 2021a; Berg et al., 2022).

11. Practical Implications for Financial Markets

Beyond policy design, this study provides actionable insights for institutional investors, asset managers, and financial analysts tasked with integrating ESG considerations into capital allocation, stewardship, and risk modelling. The BBSI reveals that ESG ratings are not interchangeable tools but institutionally constructed scores that embed unobservable evaluative logic. Investors must therefore avoid blind reliance on headline ESG scores and interrogate the methodological provenance of each rating. Specifically, analysts should examine whether scoring frameworks are peer-relative or absolute, whether controversy adjustments are transparent, and whether issue weighting reflects sector-specific risk (La Torre et al., 2020; Drempetic, Klein, & Zwergel, 2020).

The BBSI can also be operationalized as a risk-adjustment tool in ESG-integrated portfolio management. Ratings issued by high-opacity providers—such as MSCI, Sustainalytics, ISS ESG, and S&P Global—may carry greater uncertainty in regulatory auditability and performance attribution, particularly in ESG-linked finance products like green bonds or sustainability-linked loans (Christensen et al., 2022; Roulet & Bothello, 2020). As such, transparency filters should be incorporated into ESG due diligence protocols. Furthermore, institutional investors should engage with providers not merely as data consumers but as governance actors—pressuring rating agencies

to disclose evaluative mechanisms, participate in multi-stakeholder advisory boards, and adhere to public transparency benchmarks (Gibson et al., 2023; Eccles & Klimenko, 2019). Ultimately, ESG integration cannot be reduced to numerical aggregation. It must be grounded in critically evaluating how sustainability is measured, by whom, and under what assumptions.

12. Limitations

This section will present the study's limitations. First, the analysis is based on publicly available documentation and academic critiques of rating agencies. While this mirrors the position of most ESG data users, it inherently reflects the very opacity problem the study critiques. Greater access to internal rating processes would allow for a more granular analysis of scoring mechanics.

Second, the study focuses on seven major providers. While these represent a substantial share of global ESG ratings, smaller regional agencies and niche frameworks (e.g., specialized climate risk scores or human rights indices) are omitted. Their methodologies may offer either greater transparency or additional forms of opacity deserving of future analysis.

Third, while this research adopts a qualitative comparative design, future research could combine this with empirical validation, testing how score divergence influences capital flows, investment decision-making, or firm performance across sectors and jurisdictions. Future research is needed to explore how users interpret ESG ratings despite their opacity, including how institutional investors reconcile conflicting scores and how firms strategically respond to rating agencies' expectations. A deeper sociological understanding of ESG as a form of private governance would help clarify what ESG scores mean and what they do in practice.

13. References

Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance, 26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033.

Bloomberg. (2022). *Bloomberg ESG disclosure scores methodology*. https://www.bloomberg.com/professional/blog/esg-disclosure-scores-methodology/.

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review, 97*(1), 147–175. https://doi.org/10.2308/TAR-2019-0506.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics, 167*(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1.

Eccles, R. G., & Klimenko, S. (2019). The investor revolution. *Harvard Business Review, 97*(3), 106–116. https://hbr.org/2019/05/the-investor-revolution.

EFRAG. (2023). *European Sustainability Reporting Standards Exposure Drafts*. European Financial Reporting Advisory Group. https://www.efrag.org/.

ESMA. (2022). *Call for evidence on ESG ratings providers in the EU*. European Securities and Markets Authority. https://www.esma.europa.eu/.

European Commission. (2021a). *Regulation (EU) 2020/852 of the European Parliament and of the Council of 18 June 2020 on the establishment of a framework to facilitate sustainable investment (EU Taxonomy Regulation). Official Journal of the European Union, L 198*, 13–43. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32020R0852.

European Commission. (2021b). *Proposal for a regulation of the European Parliament and of the Council on the transparency and integrity of environmental, social and governance (ESG) rating*

activities (COM(2023) 314 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2023%3A314%3AFIN.

FTSE Russell. (2023). *ESG ratings methodology*. https://www.ftserussell.com/products/data/esg.

Gibson, R., Glossner, S., Krueger, P., Matos, P., & Steffen, T. (2023). Responsible institutional investing around the world. *Review of Financial Studies, 36*(1), 146–189. https://doi.org/10.1093/rfs/hhac028.

IOSCO. (2021). *Environmental, social and governance (ESG) ratings and data products providers*. International Organization of Securities Commissions. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD690.pdf.

ISS ESG. (2022). *ESG corporate rating methodology*. https://www.issgovernance.com/esg/ratings/.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance, 31*(2), 50–58. https://doi.org/10.1111/jacf.12346.

La Torre, M., Mango, F., Cafaro, A., & Leo, S. (2020). Does the ESG index affect stock return? Evidence from the Eurostoxx50. *Sustainability, 12*(16), 6823. https://doi.org/10.3390/su12166823.

MSCI. (2023). *ESG ratings methodology*. https://www.msci.com/our-solutions/esg-investing/esg-ratings.

Patten, D. M., & Shin, H. (2019). Sustainability accounting, management and policy journal's contributions to corporate social responsibility disclosure research: a review and assessment. *Sustainability Accounting, Management and Policy Journal*, *10*(1), 26-40. https://doi.org/10.1108/SAMPJ-01-2018-0017.

Refinitiv. (2023). *ESG scores methodology*. https://www.refinitiv.com/en/sustainable-finance/esg-scores.

Roulet, T. J., & Bothello, J. (2020). The ESG rating industry as mediator of risk and reputation. *Academy of Management Review, 45*(4), 792–816. https://doi.org/10.5465/amr.2017.0483.

S&P Global. (2022). *Corporate sustainability assessment methodology*. https://www.spglobal.com/esg/csa/.

Shen, H., Ng, A. W., Zhang, J., & Wang, L. (2020). Sustainability accounting, management and policy in China: recent developments and future avenues. *Sustainability Accounting, Management and Policy Journal*, *11*(5), 825-839. https://doi.org/10.1108/SAMPJ-03-2020-0077

Sustainalytics. (2022). *ESG risk ratings methodology*. https://www.sustainalytics.com/esg-rating.

Tables

Table 1. Comparative ESG Rating Providers: Methodological Structures, Transparency, and Supporting Sources

Provider	Approach	Transparency Level	Use of Proprietary Algorithms	Black Box Components	Supporting Citations
Bloomberg	Rules-based, sector-specific materiality and power-mean aggregation	Partial (pillar weights shared, formula not disclosed)	Yes	Aggregation model and issue weighting	Bloomberg (2022); La Torre et al. (2020)
Refinitiv (LSEG)	Data-driven, objective disclosure with controversy adjustment	Moderate (category weights available, aggregation opaque)	Yes	Controversy overlay logic and scoring thresholds	Refinitiv (2023); Christensen et al. (2022)

MSCI	Peer-relative risk model emphasizing financial materiality	Low (aggregation logic, weights proprietary)	Yes	Score normalization and weight allocations	MSCI (2023); Berg et al. (2022)
Sustainalytics	Sector-agnostic ESG risk exposure model	Low (risk severity logic and aggregation not disclosed)	Yes	Risk thresholds and scoring mechanisms	Sustainalytics (2022); Drempetic et al. (2020)
ISS ESG	Normative screening + KPI-based performance scoring	Low (scoring rubrics and algorithm undisclosed)	Yes	Weighting model and normative score adjustment	ISS ESG (2022); Roulet & Bothello (2020)
FTSE Russell	Thematic scoring model with global standards alignment	Moderate (themes and sources disclosed; scoring opaque)	Yes	Data transformation and aggregation procedure	FTSE Russell (2023); La Torre et al. (2020)
S&P Global	Survey-based Corporate Sustainability Assessment (CSA)	Low (CSA structure visible, scoring access restricted)	Yes	CSA weighting and algorithmic score derivation	S&P Global (2022); Christensen et al. (2022)

This table presents a comparative evaluation of seven major ESG rating agencies based on the four diagnostic dimensions of the Black Box Severity Index (BBSI). It combines methodological descriptions, transparency levels, and proprietary components, with academic citations to support the classification

Table 2. BBSI Dimensions and Conceptual Grounding

BBSI Dimension	Description	Theoretical Justification
Aggregation Model Disclosure	Refers to whether the ESG rating provider discloses the mathematical or computational method used to combine ESG indicators into a final score.	Non-disclosure of aggregation logic contributes to a 'black box' effect, undermining interpretability and investor trust (Christensen et al., 2022).
Materiality Mapping Transparency	Assesses whether the provider clearly defines which ESG issues are deemed material and how this varies across sectors or regions.	Opaque materiality frameworks hinder comparability across sectors and reduce confidence in score consistency (La Torre et al., 2020).
Replicability	Evaluates whether an external party could replicate the ESG score using publicly available information and methods.	Lack of replicability exacerbates epistemic asymmetries between providers and users, reducing methodological accountability (Kotsantonis & Serafeim, 2019).
Proprietary Algorithm Use	Identifies whether the scoring logic is based on proprietary or undisclosed algorithms that prevent external validation or audit.	Proprietary algorithms concentrate evaluative authority in private hands, undermining governance legitimacy and regulatory reliability (Roulet & Bothello, 2020).

This table defines the four core dimensions used in the Black Box Severity Index (BBSI), explaining their relevance and grounding each in the academic literature. These dimensions form the analytical foundation for assessing methodological opacity in ESG rating systems.

Table 3. Black Box Severity Scores across ESG Rating Providers

Provider	Aggregation Disclosure	Materiality Transparency	Replicability	Proprietary Algorithm Use	BBS Score	Severity Classification	Supporting Citations
Bloomberg	2	2	2	3	2.25	Moderate	Bloomberg (2022); La Torre et al. (2020)
Refinitiv (LSEG)	2	2	2	3	2.25	Moderate	Refinitiv (2023); Christensen et al. (2022)
MSCI	3	3	3	3	3	High	MSCI (2023); Berg et al. (2022)
Sustainalytics	3	3	3	3	3	High	Sustainalytics (2022); Drempetic et al. (2020)
ISS ESG	3	3	3	3	3	High	ISS ESG (2022); Roulet & Bothello (2020)
FTSE Russell	2	2	2	3	2.25	Moderate	FTSE Russell (2023); La Torre et al. (2020)
S&P Global	3	3	3	3	3	High	S&P Global (2022); Christensen et al. (2022)

This table present each dimension scored from 1 (Low Opacity) to 3 (High Opacity). BBS Score is the mean across four dimensions. Furthermore, each ESG rating provider is evaluated across four core dimensions of methodological transparency: (1) Aggregation Model Disclosure, (2) Materiality Mapping Transparency, (3) Replicability, and (4) Use of Proprietary Algorithms. For each dimension, a score is assigned on a three-point ordinal scale, where 1 indicates high transparency and low opacity**,** 2 reflects partial or conditional disclosure**,** and 3 signifies full opacity and lack of public accessibility**.** The Black Box Severity Score (BBS) represents the unweighted arithmetic mean of the four dimension scores. This composite score provides a diagnostic measure of overall methodological opacity, enabling structured comparison across providers and serving as an empirical proxy for the interpretability and auditability of ESG evaluations.

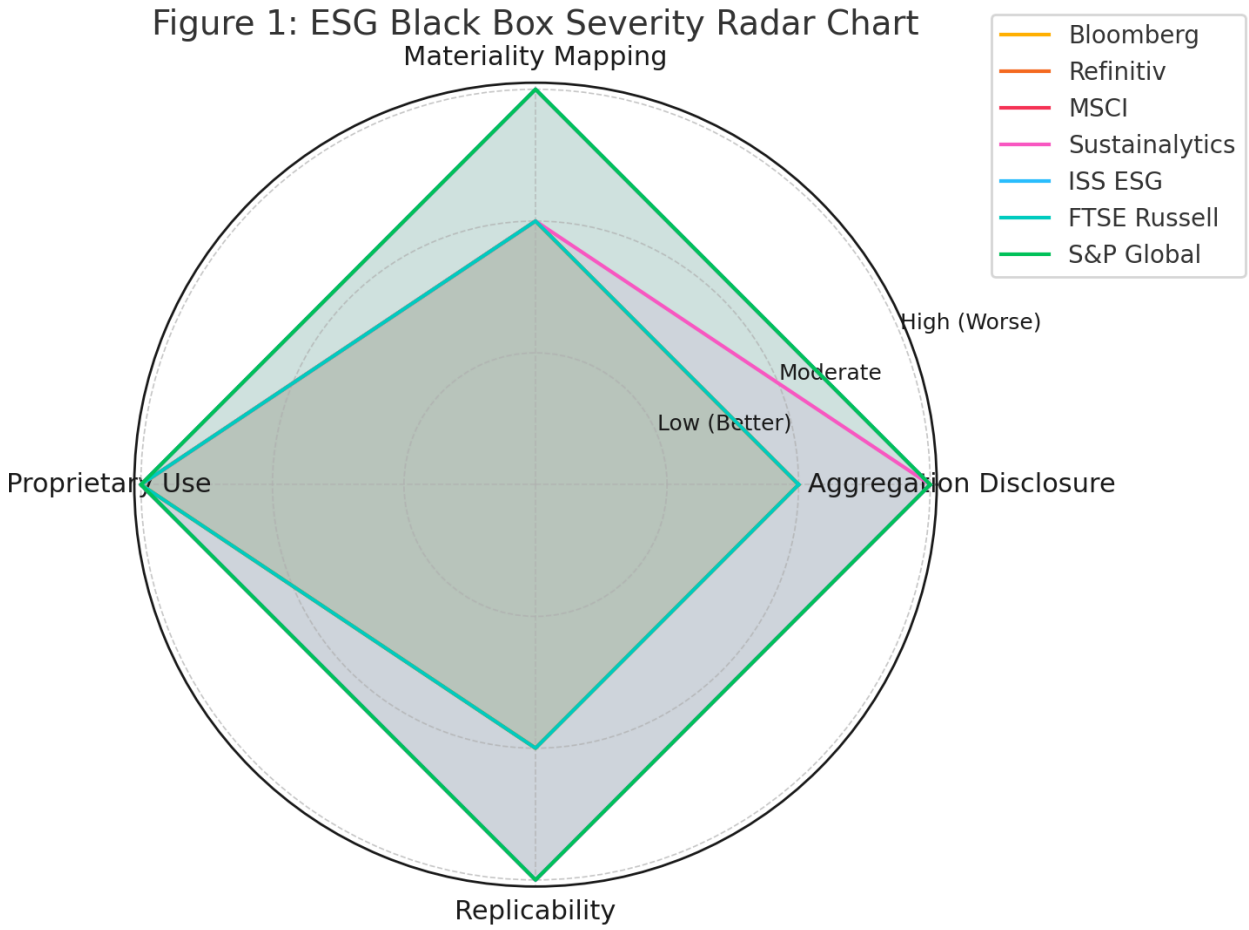

Figure 1. ESG Black Box Severity Radar Chart.

This radar chart provides a comparative visualization of methodological opacity across seven major ESG rating providers based on the Black Box Severity Index (BBSI). Each axis represents one of the four core BBSI dimensions—aggregation model disclosure, materiality mapping transparency, replicability, and proprietary algorithm use—scored from 1 (low opacity/high transparency) to 3 (high opacity/low transparency). The composite transparency level summarizes overall severity. Larger polygons extending toward the chart's perimeter indicate higher opacity, while smaller, more centralized shapes denote greater transparency. As shown, MSCI, ISS ESG, S&P Global, and Sustainalytics exhibit uniformly high opacity across all dimensions, justifying their classification as "High Severity" providers. Bloomberg, Refinitiv, and FTSE Russell score lower on average but still retain moderate opacity due to partial disclosure and continued use of proprietary methodologies (Christensen et al., 2022; IOSCO, 2021).