

UCLA

UCLA Previously Published Works

Title

Building on Sand? Third-Party Sustainability Measures in the Business Literature

Permalink

<https://escholarship.org/uc/item/0640b047>

Journal

ORGANIZATION & ENVIRONMENT

ISSN

1086-0266

Authors

Timmer, Tyson
Delmas, Magali A
Corbett, Charles J
[et al.](#)

Publication Date

2026-05-27

DOI

10.1177/10860266261443020

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

BUILDING ON SAND? THIRD-PARTY SUSTAINABILITY MEASURES IN THE BUSINESS LITERATURE

Tyson Timmer, Magali A. Delmas, Charles Corbett, Olivier Boiral, Laurence Guillaumie¹

2026

Forthcoming *Organization & Environment*

Abstract

Third-party sustainability measures, such as ESG scores and rankings, are central to research linking corporate sustainability and financial performance. However, these measures lack transparency and vary significantly across providers, raising reliability concerns. This systematic review of 82 business journal articles (1995-2024) assesses how scholars engage with and critically assess these measures. We distinguish two sources of uncertainty that limit confidence in these measures: the quality of underlying data (accuracy, reliability, timeliness) and how data is combined (fungibility assumptions and weighting schemes). Our analysis reveals that discussions of measurement quality are rare, while methodological rigor is bimodal—researchers either scrutinize multiple dimensions or none at all. We observe systematic associations between attention to measurement elements, data-provider choices, and reported financial performance. We argue that choices about measure quality and aggregation are not neutral, but directly shape empirical findings and their interpretation. We outline practical recommendations to advance rigor and transparency in sustainability-performance research.

Keywords: Sustainability metrics, ESG, CSR, CSP, CEP, Systematic literature review, metrology, secondary data.

¹ Tyson Timmer - PhD Student, University of California, Los Angeles; tysonzimmer@gmail.com; Magali A. Delmas - Professor, University of California, Los Angeles, delmas@ucla.edu; Charles Corbett - Professor, University of California, Los Angeles; charles.corbett@anderson.ucla.edu; Olivier Boiral - Professor, Université Laval, Olivier.Boiral@mng.ulaval.ca; Laurence Guillaumie - Professor, Université Laval, laurence.guillaumie@fsi.ulaval.ca

Introduction

Third-party sustainability measures, such as Environmental, Social, and Governance (ESG) scores, rankings, and composite indicators, are central to a major theme in corporate sustainability research: examining the relationship between corporate sustainability and financial performance (Burbano et al., 2024). Yet despite their widespread use, these measures face growing scrutiny over their reliability, comparability, and validity (Berg et al., 2022; Boiral et al., 2020, 2021; Busch et al., 2022; Chatterji et al., 2016; Delmas et al., 2013; Kotsantonis & Serafeim, 2019; Widyawati, 2021). These measures range from fully aggregated scores (e.g., Bloomberg ESG) to disaggregated metrics (e.g., KLD/MSCI's binary strengths and concerns); we refer to them collectively as “sustainability measures” throughout this review.²

Much of the criticism targets providers' proprietary methods and limited disclosure regarding data sources, measurement approaches, and aggregation schemes. These concerns reflect longstanding problems in scientific measurement: all metrics have inherent technical and conceptual limitations. If left unacknowledged, these limitations foster false confidence (van der Bles et al., 2019; National Academies of Sciences, 2017, 2019). Research shows that how uncertainty is communicated, or ignored, shapes policy use: downplaying uncertainty can mislead decision-makers and distort expectations (Dhami & Mandel, 2022; Fischhoff & Davis, 2014; van der Bles et al., 2019; van der Bles et al., 2020). These insights are especially relevant

² The label “ESG” was originally coined by the United Nations Global Compact (2004) and has since become the dominant marketing term among financial data services providers (Serafeim, 2023). Other labels in common use include Corporate Social Performance (CSP), Corporate Environmental Performance (CEP), and Corporate Social Responsibility (CSR). We use “sustainability measures” as an umbrella term to encompass this variety of measures for brevity.

for third-party sustainability measures, where a lack of transparency leads researchers to unknowingly adopt the provider's assumptions.

Despite these concerns, third-party measures remain a cornerstone of empirical sustainability research largely because viable alternatives are limited (Berg et al., 2022; Dimson et al., 2020; Kotsantonis & Serafeim, 2019). While some call for abandoning these measures (Tayan, 2022), others advocate for greater justification and transparency in their application (Berg et al., 2022; Kotsantonis & Serafeim, 2019). We therefore ask: To what extent, and in what ways, do scholars engage with and critically assess third-party measures of corporate sustainability when examining the relationship between sustainability and financial performance? Specifically, we examine two domains of measurement uncertainty: the quality of underlying data (whether it is accurate, reliable, and timely) and the methods used to aggregate that data (decisions about fungibility—whether different sustainability dimensions can be meaningfully combined—and weighting).

Existing reviews have examined sustainability measurement from stakeholder-specific perspectives—regulatory requirements (Christensen et al., 2021), managerial decision-making (Grewal & Serafeim, 2020), and investor use (Frankel et al., 2025)—but none systematically assess how empirical business researchers engage with measurement uncertainty embedded in third-party data. This gap is consequential: while many reviews examine the sustainability–financial performance relationship with mixed results (Alshehhi et al., 2018; Busch & Friede, 2018; Endrikat et al., 2014; Huang, 2021; Kong et al., 2019; Van Beurden & Gössling, 2008), few examine how the inherent uncertainty of the metrics affects those evaluations. Even meta-analyses reporting positive associations (e.g., Friede et al., 2015) inherit limitations from primary

studies reliant on uncertain metrics, and more recent syntheses warn that structural weaknesses in third-party metrics can make them ill-suited for certain contexts (Coelho et al., 2023; Damtoft et al., 2025). Coelho et al. (2023) and Damtoft et al. (2025) do not, however, examine how empirical researchers engage with these weaknesses in practice.

To systematically examine how researchers address measurement uncertainty, we develop an analytical framework that integrates complementary foundations from measurement science (Byerly & Lazara, 1973; Simpson, 1981), social science research on secondary data use (Calantone & Vickery, 2010; Stewart & Kamins, 1993), and environmental and sustainability accounting research that has long treated information quality as a prerequisite for accountability and decision usefulness (Lamberton, 2005; Schaltegger & Burritt, 2000). The latter tradition highlights a foundational problem in accounting: high-quality information reduces the asymmetry between managers and outside stakeholders, yet producing and verifying non-financial data remains exceptionally difficult and costly (Dechow et al., 2010; Healy & Palepu, 2001). As Schaltegger (1997) warns, this creates a systemic risk that poor-quality sustainability information will crowd out the good. We then connect these foundations to recent corporate sustainability scholarship by drawing on Berg et al. (2022), who show that divergence across ESG ratings stems from identifiable differences in scope, measurement, and aggregation—the same dimensions our framework is designed to interrogate, and on King and Berchicci (2021), who conceptualize researchers’ analytic choices as “forking paths.” Taken together, these literatures underscore that selecting a third-party sustainability measure is itself a consequential, yet underexamined, forking path because it imports provider-side assumptions about both data quality and aggregation into researchers’ designs. When those foundations go unexamined, the

empirical structures built upon them rest on sand: findings that appear robust may instead reflect the unscrutinized choices of a data provider rather than the underlying reality of corporate sustainability.

Operationalizing these insights, our framework identifies five elements—accuracy, reliability, timeliness, fungibility, and weighting—that pinpoint where methodological assumptions enter research designs and potentially shape outcomes. Each element represents a distinct forking path: when researchers choose a third-party measure without interrogating its accuracy, reliability, timeliness, fungibility assumptions, or weighting scheme, they implicitly accept provider-side choices that may condition the direction and magnitude of their findings. We specified these elements prior to coding to reduce researcher’s degrees of freedom and to ensure consistent evaluation across studies. The first three concern data quality, while the latter two concern how dimensions are combined. By applying this framework, we demonstrate that choices regarding measure quality and aggregation are not neutral but condition the outcomes scholars report.

To explore this, we analyze 82 peer-reviewed articles published between 1995 and 2024. Our aim is not to comprehensively review the literature on the sustainability–financial performance relationship (see, e.g., Chen et al., 2023; Coelho et al., 2023; Gillan et al., 2021); rather, we examine the measurement infrastructure underpinning their findings. Our analysis reveals that while sustainability measures are ubiquitous, most studies pay limited attention to their underlying construction. Few articles critically assess key elements such as accuracy, reliability, timeliness, fungibility, or weighting, and even fewer examine how these elements might affect empirical outcomes. This lack of scrutiny is concerning, given the opacity of third-party methodologies. Notably, engagement with one element, such as accuracy, tends to coincide with

attention to others. Moreover, measurement choices are systematically associated with the direction and framing of reported results, underscoring that these decisions are not merely technical but substantively consequential.

The article proceeds as follows. We begin by presenting our framework, which highlights the key elements where methodological assumptions enter research designs and shape empirical results. Next, we detail our review methods, followed by results examining how these elements relate to methodological choices and reported outcomes. We conclude with recommendations to improve rigor and transparency in the use of third-party sustainability measures.

Quality and Aggregation Challenges in Sustainability Measures

Third-party sustainability measure providers typically rely on proprietary, opaque data collection and aggregation methodologies (Kotsantonis & Serafeim, 2019). As a result, researchers using these measures face two main domains of uncertainty. The first, measure quality uncertainty, concerns the credibility of the underlying data and encompasses issues of accuracy, reliability, and timeliness (Boiral et al., 2020, 2021). The second, aggregation uncertainty, stems from combining individual indicators into composite scores, raising issues of fungibility and weighting (Chatterji et al., 2016; Widyawati, 2020, 2021). Drawing on information quality literature (Byerly, 1973; Nelson et al., 2005; Rabinowich, 2005; Simpson, 1981; Wang & Strong, 1996), we emphasize that both the input quality and aggregation methods determine whether measures support valid conclusions. This aligns with accounting research showing that sustainability information differs fundamentally from traditional financial information. Specifically, sustainability information is multidimensional (encompassing environmental,

social, and governance dimensions that may not be naturally comparable), involves externalities (capturing impacts on stakeholders beyond firm boundaries), and serves heterogeneous users (investors, regulators, employees, communities) with divergent information needs (Bebbington & Larrinaga, 2014; Unerman et al., 2018). These characteristics mean that decisions about data quality and aggregation methods are not secondary technical matters but first-order determinants of whether sustainability measures can meaningfully inform decision-making (Christensen et al., 2021; Friedman & Ormazabal, 2024).

These uncertainties manifest in various ways. Inconsistent collection and reporting standards undermine accuracy and reliability (Berg et al., 2022; Busch et al., 2022), while reporting delays and horizon discrepancies affect timeliness (Delmas & Doctori-Blass, 2010). Fungibility, or treating different sustainability measures as interchangeable, can obscure firm-specific weaknesses (Capelle-Blancard & Petit, 2017; Munda & Nardo, 2005). Prior reviews note that such aggregation practices fuel persistent disagreement among third-party measures and blur the distinction between financially material and immaterial sustainability issues (Friedman & Ormazabal, 2024; Grewal & Serafeim, 2020). Additionally, third-party providers often assign weights based on subjective judgments, introducing bias and misalignment with stakeholder expectations (Capelle-Blancard & Petit, 2017; Gan et al., 2017; Greco et al., 2018). Aggregation and weighting schemes often reflect provider incentives and constraints rather than financial materiality, further complicating interpretation by users (Friedman & Ormazabal, 2024).

Together, these five elements capture the two principal entry points for uncertainty: the quality of the data inputs and the aggregation of those inputs into composite scores. Table 1 summarizes the definitions and scholarly grounding of these five elements.

[Insert Table 1 about here]

We investigate how business research acknowledges or assesses these five elements when using third-party sustainability measures. In doing so, we illuminate prevailing practices in the literature and identify opportunities to enhance the rigor and transparency of research relying on these measures.

Sustainability accounting scholars have long argued that non-financial information poses distinctive quality challenges that conventional accounting frameworks are ill-equipped to handle. Unlike financial data, sustainability information is multidimensional, difficult to verify, and serves a heterogeneous audience of investors, regulators, employees, and communities with divergent needs (Bebbington & Larrinaga, 2014; Unerman et al., 2018). It frequently relies on estimates, self-reported disclosures, and unverifiable proxies, and it captures externalities—impacts on parties beyond the firm—that fall outside traditional accounting boundaries (Gray, 2010). These characteristics mean that quality cannot be assessed through a single lens: a measure may be timely but inaccurate, or comprehensive but aggregated in ways that obscure material weaknesses.

Sustainability accounting research has responded by articulating information quality through a set of qualitative characteristics—most notably relevance (including materiality) and faithful representation, supported by comparability, verifiability, timeliness, and understandability (Bebbington & Larrinaga, 2014; Lamberton, 2005; Schaltegger & Burritt, 2010). Yet scholars have long questioned whether these frameworks can adequately capture sustainability performance, given the treatment of externalities, temporal horizons, and stakeholder plurality they require (Gray, 2010; Unerman et al., 2018). Third-party sustainability measures

operationalize only part of this ideal: providers must generate credible inputs, apply standardized data-collection and estimation procedures, and enable comparability across firms (Boiral et al., 2020).³ These challenges make the quality of third-party sustainability measures especially consequential for empirical research. In this review, we focus on the subset of quality characteristics most directly at stake when researchers select and use such measures: (1) the accuracy and reliability of inputs (e.g., audited data versus model-based estimates), (2) the timeliness of measurement windows, and (3) the aggregation logic—including fungibility and weighting—used to construct composite outputs. We do not evaluate whether a measure captures the “right” impacts for a given industry or reflects sufficient stakeholder materiality. For example, GHG emissions may be less material than data privacy controls for software firms; our concern is not whether GHG emissions are the “correct” sustainability metric in that context, but whether a third-party GHG score used in research rests on credible inputs or potentially inaccurate estimation models. Ultimately, we examine the extent to which scholars interrogate these technical foundations—specifically, whether they acknowledge and account for the inherent uncertainties of measurement and aggregation.

Elements of Measure Quality: Accuracy, Reliability, and Timeliness

Measurement science distinguishes three related but distinct properties: accuracy (closeness to the true value), reliability (repeatability across measurements), and validity (whether a measure

³ Standards from the Global Reporting Initiative (GRI), the World Economic Forum (WEF), and the Sustainability Accounting Standards Board (SASB, now part of IFRS) outline additional elements of sustainability measures. For example, GRI elements such as “balance,” “completeness,” and “sustainability context” (Global Reporting Initiative, 2016) focus on ensuring that sustainability reporting provides a fair and comprehensive view of a firm’s impacts. “Balance” ensures that firms report both positive and negative impacts, “completeness” assesses whether firms cover all relevant topics, and “sustainability context” examines how the firm’s performance aligns with broader sustainability goals. These elements help stakeholders evaluate whether reports provide an appropriate and accurate picture of a firm’s sustainability.

captures the construct it purports to represent) (Byerly & Lazara, 1973; Rabinowich, 2005; Simpson, 1981). For composite sustainability scores, validity is arguably the most fundamental concern: a score may be internally consistent—that is, reliable—yet still fail to represent corporate sustainability meaningfully if its scope is arbitrarily defined, its indicators poorly chosen, or its aggregation logic theoretically unjustified (Chatterji et al., 2016; Delmas et al., 2013). We focus on accuracy and reliability as the operationalizable elements of data quality that researchers can interrogate using available information, while recognizing that the construct validity of third-party scores is a prior and largely unresolved question. This connects directly to Berg et al.’s (2022) finding that divergence across ESG ratings stems from scope, measurement, and aggregation differences: scope decisions determine what the measure purports to represent (a validity question), measurement decisions determine how faithfully underlying data is captured (an accuracy and reliability question), and aggregation decisions determine how dimensions are combined (a fungibility and weighting question). Together, these sources of divergence map onto the five elements of our framework and explain why selecting a third-party measure is itself a consequential methodological choice. These are critical concerns for all research relying on secondary data, particularly data “collected by someone else for another primary purpose” (Johnston, 2017, p. 619). Stewart and Kamins (1993) warn that “[n]ot all information obtained from secondary sources is equally reliable or valid” (p. 17), urging researchers to remain skeptical of data quality and assess the potential impact on their analysis.

These concerns are magnified in the sustainability context, where measurement often relies on estimates, self-reported data, and unverifiable proxies, making verification difficult (Grewal & Serafeim, 2020; Friedman & Ormazabal, 2024). Because high-quality environmental information

is costly to produce and difficult for external stakeholders to assess, there is a systemic risk that poor-quality information will overshadow the good (Schaltegger, 1997). Inaccurate or unreliable sustainability measures can lead to flawed conclusions about the relationship between sustainability performance and corporate financial performance (CFP). Berg et al. (2022) report that 56% of the variation among sustainability measures arises from differences in how third-party providers define and measure underlying metrics. This variation reflects “noisy measures of an underlying latent quality” (Berg et al., 2022, p. 1330) and can be amplified by reliance on estimates (Busch et al., 2022). Prior reviews interpret this as evidence that greater disclosure does not necessarily improve measurement quality or comparability, particularly when metrics lack standardized definitions or verifiable measurement protocols (Grewal & Serafeim, 2020; Christensen et al., 2021). Users thus face challenges in discerning variations in accuracy (Kotsantonis & Serafeim, 2019; Cort & Esty, 2020), while opacity in data collection and aggregation exacerbates uncertainty around reliability (Widyawati, 2020). These quality deficiencies limit researchers’ ability to distinguish signal from noise, even when data is publicly available (Christensen et al., 2021).

The third key element is timeliness, defined by GRI (2016, p. 16) as “the regularity of reporting [...] relative to the reporting period [and] the period to which [the measure] relates.” Timeliness requires clarity about the timeframe represented and ensuring consistency across all underlying metrics. For example, while Scope 1 and 2 GHG emissions may correspond to the stated reporting year, Scope 3 emissions may rely on data from prior years. Such temporal discrepancies can influence the predictive validity of an analysis (Clifford et al., 2016; Nelson et al., 2005; Stewart & Kamins, 1993). These mismatches reflect structural challenges in

sustainability reporting, where long horizons, supply-chain complexity, and estimation requirements make contemporaneous measurement difficult (Christensen et al., 2021; Friedman & Ormazabal, 2024). Consequently, analysts must contend with the risk that sustainability data aligned with financial outcomes actually describes earlier periods, complicating causal claims. Although researchers often align measures with the publication year, firms typically base their data on performance from earlier periods (Delmas et al., 2022). Establishing the precise timing of variables is critical for causal inference. For instance, if a 2025 sustainability measure reflects 2024 performance, lagging the measure to analyze effects on 2025 financial performance could create a two-year gap, potentially leading to inaccurate conclusions. These challenges reinforce broader concerns that sustainability disclosures may fail to support reliable inference when measurement noise and temporal misalignment are not explicitly addressed (Christensen et al., 2021; Frankel et al., 2025).

Elements of Aggregation: Fungibility and Weighting

Aggregation methods form the second major domain of uncertainty. Fungibility and weighting are not neutral technical choices; they embody assumptions about trade-offs and priority among sustainability dimensions. When these assumptions remain opaque or untested, they introduce aggregation uncertainty, which can strongly influence empirical results. Prior work in sustainability accounting warns that aggregation practices often lack clear decision-making relevance, risking the reduction of sustainability metrics to symbolic artifacts rather than durable analytical tools (Burritt & Schaltegger, 2010).

Fungibility refers to treating different metrics as interchangeable, allowing strong performance in one area to compensate for poor performance in another (Capelle-Blancard & Petit, 2017, p. 920). In sustainability metrics, fungibility enables overperformance in one domain to offset underperformance in another (Escrig-Olmedo et al., 2017; Graafland et al., 2004). For example, a company might mask poor environmental performance, such as excessive resource consumption, by excelling in social performance. This compensatory logic can obscure deficiencies or incomplete disclosures, complicating interpretation (Berg et al., 2022; Capelle-Blancard & Petit, 2017; Delmas & Doctore-Blass, 2010; Escrig-Olmedo et al., 2014). Aggregating extensive but potentially symbolic disclosures into composite measures may create metrics that appear comprehensive while masking substantive performance gaps (Michelon et al., 2015). Prior reviews caution that such aggregation blurs distinctions between financially material and immaterial sustainability issues (Grewal & Serafeim, 2020; Christensen et al., 2021). Consequently, highly aggregated measures often obscure the specific drivers of the overall sustainability rating. This demonstrates how fungibility injects uncertainty: by permitting compensation across domains, measures may hide weaknesses material to stakeholders while inflating confidence in composite scores. Early scholarship emphasized that combining indicators involves inherent trade-offs, warning that metrics lose interpretability when these assumptions remain implicit (Figge et al., 2002).

Weighting similarly shapes the construction and interpretation of sustainability measures, as it assigns “the relative importance of different dimensions in their contributions to [...] sustainability performance” (Gan et al., 2017, p. 492). Lacking a single objective function for sustainability performance, weighting schemes necessarily reflect normative judgments about

importance rather than purely technical considerations (Friedman & Ormazabal, 2024).

Weighting may reflect materiality considerations, stakeholder preferences, or an equal-weighting approach (Gan et al., 2017; Singh et al., 2009). Crucially, the choice of scheme determines the extent to which individual metrics compensate for one another. Research demonstrates that even minor weight adjustments can significantly alter measures (Chen & Delmas, 2011; Zhou et al., 2017). When weights are undisclosed or unjustified, aggregation uncertainty increases: small methodological changes can change effect sizes, significance, and even sign. Such opacity complicates assessments of whether measures capture relevant information or merely encode provider-specific preferences (Frankel et al., 2025).

However, third-party providers often lack transparency about their aggregation methods. For example, Bloomberg offers only a brief two-page overview of its methodology (Bloomberg, n.d.; Bloomberg, 2021). This opacity forces researchers to make assumptions about fungibility and weighting, introducing uncertainty and potential bias. Without clear information on how providers combine and weight metrics, assessing the validity of measures and the robustness of conclusions becomes a significant challenge (Berg et al., 2022; Boiral et al., 2020). Increased disclosure does not resolve this issue; instead, it may create an illusion of precision when underlying construction choices remain hidden (Christensen et al., 2021; Grewal & Serafeim, 2020).

Just as accuracy, reliability, and timeliness capture measure quality uncertainty, fungibility and weighting encapsulate aggregation uncertainty. In both domains, uncertainty is embedded in the sustainability metrics rather than arising only from downstream statistical analysis.

Methods

This review follows the systematic review approach defined by Grant and Booth's (2009) typology and adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Page et al., 2021; Panic et al., 2013) to ensure transparency and reproducibility.

Search Method

We searched the Web of Science and ProQuest databases for articles published in English between 1995 and 2024. These databases offer comprehensive coverage of high-quality journals in business, management, economics, and environmental studies and are commonly used in systematic literature reviews within these fields. We limited our search to journals indexed under the following categories: "Business," "Management," "Business Finance," "Operations Research Management Science," "Economics," "Environmental Studies," "Environmental Sciences," or "Green Sustainable Science Technology." Articles published in practitioner outlets, book chapters, conference proceedings, dissertations, and non-peer-reviewed sources were excluded. To define our keyword search, the research team first identified pivotal literature that examined the causal relationship between sustainability measures and CFP. These seminal articles served as an anchoring set against which candidate search terms were evaluated for relevance and coverage. This body of work informed the initial list of keywords used. Given the evolution and breadth of corporate sustainability research, spanning decades and employing a diverse, inconsistent vocabulary and disciplinary traditions, establishing a targeted, theoretically grounded keyword set was an essential early step. The field's vocabulary has undergone

significant shifts over time: early literature often referred to “corporate social responsibility” (CSR), “social performance,” or “triple bottom line,” while more recent research uses terms like “ESG performance,” “sustainability ratings,” or “non-financial disclosures.” Some terms have become more narrowly defined (e.g., ESG as an investment screening tool), while others have broadened or blurred in scope. Additionally, certain concepts like “stakeholder engagement,” “responsible business,” or “sustainability performance” may be used interchangeably in some contexts and distinctly in others.

This terminological inconsistency creates two risks: including irrelevant articles or missing relevant ones. Without anchoring to key papers, we would have needed hundreds of search terms to identify relevant work. To refine our approach, we conducted multiple rounds of iterative searches, beginning with broader keyword combinations and progressively narrowing the search string. At each iteration, we assessed whether the search results consistently captured the anchoring articles while minimizing the inclusion of clearly irrelevant studies. This process was conducted primarily in the Web of Science database due to its transparent indexing and reproducibility of search strings.

The final keyword string (see Appendix 1) included terms identified as foundational in the field: “Corporate Social Performance” (CSP), “Corporate Environmental Performance” (CEP), and “Corporate Social Responsibility” (CSR). These terms were among the most frequently used in corporate sustainability research between 1994 and 2021, according to a comprehensive bibliometric analysis by Burbano et al. (2023). We also included “ESG,” to reflect the term prominence in finance and investment research (Chytis et al., 2024; Friede et al., 2015).⁴ We

⁴ We searched for “ESG” specifically rather than “environmental, social, and governance” because the acronym is the standard terminology in the finance and investment literature we aimed to capture, reducing irrelevant results.

applied search strings to article titles, abstracts, and keywords. To further focus the search on empirical studies linking sustainability to financial outcomes, we added “Financial Performance” and “Shareholder Value” to the search string. We limited our search to those published between 1995 and 2024. Preliminary searches indicated that extending the range earlier did not yield additional relevant articles, and influential early work in this domain (e.g., Waddock & Graves, 1997) emerged shortly thereafter.

The Web of Science search returned 4,535 articles, and the ProQuest search returned 2,706 articles. After merging the two datasets, duplicate records were removed. To screen for journal quality, we included journals with a 2022 Journal Citation Indicator (JCI) score of 1.0 or higher. A JCI score above 1.0 indicates that a journal has more citations than the average of journals in the same academic categories (Szomszor, 2021). We also included journals with a 3, 4, or 4* rating, indicating “Highly Regarded,” “Top,” and “World Elite” journals, according to the Association of Business Schools (ABS) 2021 Academic Journal Quality Guide, as well as journals with an “A” or “A*” rating, indicating “Highly Regarded” and “Best or Leading” journals, in the Australian Business Deans Council (ABDC) 2023 Journal Ranking. Journals that did not meet at least one of these quality thresholds were excluded from further consideration. Both journal rating groups are considered two of the top lists for business literature (Harzing, 2023), and this screening approach has been employed in prior literature reviews (Damtoft et al., 2025). After removing duplicates and screening for journal quality, 714 articles remained and were advanced to content-based screening.

We selected articles using sustainability measures as an independent variable and one of four common measures of CFP: Return on Assets (ROA), Marginal Value Added (MVA), Tobin’s Q,

and Cumulative Abnormal Return (CAR). This selection follows Berchicci & King (2022), who utilized these metrics to investigate the effects of measurement uncertainty on highly influential sustainability studies. Narrowing our scope to these specific metrics facilitates a more direct and meaningful comparison among the reviewed articles. While this focus necessarily limits the total number of in-scope articles, there is no a priori reason to assume that studies using alternative performance indicators would be systematically more or less rigorous in their consideration of the quality of sustainability measures.

Using these screens, we identified 82 in-scope articles for complete data extraction (see Appendix 2 for the screening flowchart and Appendix 3 for the final sample). Two additional articles identified by the research team as key works were added to the screening set of articles (indicated in Appendix 3 with an “*”). The keyword search string did not capture these articles as their abstracts use more generic terms than those in our keyword search. Both were identified through prior work that was central to the development of our review, specifically Berchicci and King (2022) and Delmas et al. (2013). This approach yielded a representative set of influential studies published in highly regarded journals, enabling systematic analysis of how the field discusses third-party sustainability measures rather than attempting an exhaustive survey of all sustainability–CFP research.

Data Extraction

The data-extraction grid covered four main areas: firm characteristics, statistical analysis, outcomes, and the measure quality and aggregation elements (see Appendix 4). Firm characteristics included country, region, sector, and the analysis time range. Statistical analysis information included descriptive statistics of the sustainability measure, the frequency, and

names of measure providers, the ESG areas of the measures (e.g., those covering social, environmental, and/or governance aspects of corporate sustainability), and the statistical approach, along with its outcomes.

We coded each article for three quality elements (accuracy, reliability, timeliness) and two aggregation elements (fungibility, weighting), as summarized in Table 1. To capture substantive engagement beyond exact terminology, we coded elements as present whether explicitly named, described with synonyms, or implicitly operationalized. For example, we coded reliability as present if authors discussed data “precision,” “consistency,” or “verification.” We coded fungibility as implicitly present when authors described aggregation methods that treated positive and negative performance as interchangeable, such as subtracting “concerns” from “strengths” in KLD scores.

Data Synthesis

To synthesize the extracted data, we followed a two-step process. First, we categorized the study scope, third-party sources, and statistical models. We also recorded the statistical details on the relationship between the sustainability measures (independent variable) and financial measures (dependent variable), including direction, magnitude, and significance. Second, we conducted a thematic and descriptive analysis of the uncertainty elements (see Table 1 and Appendix 4).

Results

We begin with a high-level overview of the 82-article sample, followed by a detailed exploration of how the five key elements of our framework—accuracy, reliability, timeliness, fungibility, and

weighting—are addressed and how they relate to provider choice and reported statistical outcomes. While the search included articles published between 1995 and 2024, more than 50% of the articles in our sample were published in 2019 or later.

Reliance on third-party providers varied. Eighteen different sources were used, the most common being KLD/MSCI (purchased by MSCI in 2010), Refinitiv (formerly of Thomson Reuters, and now part of the London Stock Exchange Group, LSEG), and Bloomberg ESG (see Figure 1). KLD/MSCI dominated before 2015, while other providers gained traction more recently. Most articles (91%) relied on a single provider. Seventy-two percent employed sustainability measures that encompassed environmental, social, and governance dimensions, while the remaining articles focused on only one or two dimensions—most often (28%) excluding governance.

[Insert Figure 1 about here]

Next, we examine whether and how provider choice relates to reported statistical outcomes. Table 2 summarizes the direction and significance of the reported relationships between sustainability measures and CFP.

A majority of studies (59%) report only a positive and statistically significant relationship between sustainability measures and CFP. In contrast, 15% report only a negative relationship, and 9% find no significant effect. Notably, 18% of studies report both positive and negative effects, often due to disaggregated analyses across sustainability dimensions or the presence of non-linear relationships. This distribution underscores both the predominance of positive findings and the heterogeneity in reported effects. Mixed (both positive and negative) results occur mainly in KLD/MSCI-based and VigeoEIRIS-based research, indicating that aggregation choices (e.g., how “strengths” offset “concerns”) could shape observed effect patterns. In

contrast, Bloomberg and Refinitiv measures overwhelmingly yield unidirectional findings, with only a handful of papers documenting curvilinear relationships (e.g., Barnett & Salomon, 2012; Kumar et al., 2022).

Table 2 also shows considerable variation in how third-party sustainability scores are scaled and reported across the 82 studies. Most providers (e.g., Bloomberg, Newsweek, Sustainalytics, CSRHub) use a 0–100 scale, but reported values often fall within a narrower range (e.g., Bloomberg scores from 0 to 100, with observed minima around 2–5 and maxima below 85), raising the possibility of floor or ceiling effects. Other measures use 0–1 scales (e.g., Refinitiv, VigeoEIRIS), while KLD/MSCI’s disaggregated “strengths” and “concerns” scores span –20 to +20, and Fortune rankings use a 0–10 scale. This lack of standardization in both scaling and observed ranges complicates cross-study comparability.

***[Insert Table 2 about here] ***

Engagement with the framework elements also varied. Only a small fraction of the 82 reviewed articles addressed accuracy (24%), reliability (28%), or timeliness (6%), whereas weighting (59%) and fungibility (53%) received far greater attention. We coded each article to indicate whether it discussed the five elements (1/0), and we calculated ϕ correlation coefficients to examine the relationships among the elements, the statistical outcomes, and the sustainability measures used (see Table 3).

*** [Insert Table 3 about here]***

The analysis reveals two clusters among measurement-quality elements. Articles discussing accuracy are 10.5 times more likely to also discuss reliability (OR = 10.5, $\chi^2 = 9.8$, $p = 0.002$),

while fungibility discussions strongly align with weighting discussions ($\phi = 0.56$, $p < .01$). These clusters indicate a systematic divide: researchers either approach third-party data with comprehensive skepticism—addressing multiple quality dimensions simultaneously—or treat it as validated input requiring minimal scrutiny.

Emphasis on reliability is positively associated with reporting statistically significant positive effects ($\phi = 0.24$, $p < .05$) and inversely with reporting negative effects ($\phi = -0.24$, $p < .05$), while timeliness shows negative associations with fungibility discussions ($\phi = -0.27$, $p < .05$). Studies relying on KLD/MSCI data are far more likely to engage in weighting discussions ($\phi = 0.47$, $p < .01$). Bloomberg-based research reports fewer negative effects ($\phi = -0.23$, $p < .05$). Overall, these associations indicate that conceptual emphases (which elements authors discuss) and methodological choices (which provider they use) co-vary with the direction of reported effects. This reinforces our core argument: measure choices are not neutral, so their implications and limitations should be examined in studies that use them.

Measure Quality

Despite their importance, discussions of measurement quality were rare: only 24% addressed accuracy, 28% reliability, and 6% timeliness. When mentioned, these elements were typically brief justifications for dataset choice, not rigorous assessments of the measures' actual quality. For example, some articles cited reliability as a rationale for using a given dataset but did not explain why reliability was critical to their analysis (e.g., Blanco et al., 2013; Gao & Bansal, 2013; Oikonomou, Brooks & Pavelin, 2014; Petitjean, 2019). Illustrative statements include describing KLD data as “more objective” due to independent aggregation (Blanco et al., 2013, p.

70) or claiming Bloomberg ESG data is entirely “transparent back to a company document” (Petitjean, 2019, p. 504).

Only 10 articles discussed potential limitations related to reliability or accuracy—a striking gap given that 56% of variation among ESG measures stems from provider differences (Berg et al., 2022). Gao and Bansal (2013) noted that “KLD data [is] limited [...] due to the binary nature of the variables” (p. 252), while Duque-Grisales and Aguilera-Caracuel (2021) observed that the Refinitiv measure “is not free of subjective influences” (p. 330). However, most articles did not state their assumptions about data quality, raising questions about the validity of their statistical findings. This omission is especially consequential when results are borderline significant: among articles reporting statistically significant results, 17% reported p-values between 0.05 and 0.10 (i.e., meeting a $p < 0.10$ threshold but not the conventional $p < 0.05$ standard). A comparison of two studies illustrates how engagement with quality varies even among similar research designs. Lee et al. (2016) and Gangwani and Kashiramka (2024) both examined the relationship between sustainability measures and ROA, reporting significant associations at the 0.10 level only. Lee et al. (2016) explicitly discussed the construction and reliability of their Korean Corporate Governance Service (KCGS) data. In contrast, Gangwani and Kashiramka (2024) did not discuss the accuracy or reliability of Bloomberg’s data. It is therefore impossible to assess whether minor inaccuracy in the Bloomberg data would change their conclusions.

Temporal alignment is rarely scrutinized. Although over 45% of articles employed lagged sustainability measures in their analyses, only five explicitly addressed timeliness. These five varied in approach: Pichler et al. (2017) justified their temporal framing based on data availability constraints, while Sandberg et al. (2023) focused on the public release date of a

measure—when it becomes accessible to investors—rather than the performance period it reflects. Only Sandberg et al. (2023) explicitly justified their lag choice based on the period represented by the measures. Most studies seem to assume that sustainability measures are contemporaneous, with little scrutiny of how well the timing of sustainability and financial performance variables aligns. Preston and O’Bannon (1997) and Nakao et al. (2007) addressed this uncertainty by running models with both lagged and contemporaneous specifications.

Neglecting timeliness undermines longitudinal designs, which comprised over 95% of the sample. Researchers may assume that year-to-year changes in sustainability measures reflect actual shifts in firm performance. Yet variation may instead result from changes in providers’ methodologies (Albuquerque et al., 2019; Harrison et al., 2023) or from inherent instability (Awaysheh et al., 2020). Conversely, KLD/MSCI measures have exhibited limited variability over time, presenting a different timeliness challenge (Chatterji et al., 2009) that few studies explicitly address (see Table 2). While Waddock and Graves (1997, p. 308) note that “KLD staff members meet on a weekly basis [...] to assure that decisions [...] are being made consistently,” such endorsements rarely address potential consequences for studies using that data. Overall, discussions of accuracy, reliability, and timeliness were brief and seldom integrated into research design.

Measure Aggregation

Aggregation decisions—particularly fungibility and weighting—embody normative assumptions about trade-offs among sustainability dimensions, making substantive engagement with these elements essential for transparent research. The results are mixed. Seventy-two percent of studies relied on aggregated measures, raising questions about construct validity and aggregation logic,

and engagement was higher than for measure quality elements: 59% of articles discussed weighting and 54% addressed fungibility. These discussions typically appeared in the methods sections, focusing on weighting choices (commonly equal weighting or proprietary formulas) and decisions about whether to combine dimensions (e.g., using a composite ESG score versus separate E, S, and G scores). Even when aggregation was discussed, authors often treated it as a procedural choice rather than as a set of substantive assumptions about trade-offs among sustainability dimensions.

Most studies discussing weighting relied on unjustified equal weighting (see Figure 2). Some researchers, such as García-Sánchez & Martínez-Ferrero (2019), explicitly noted that equal weighting assumes all metrics are equally important, an assumption rarely empirically validated. Waddock and Graves (1997) and Zhao and Murrell (2016) established weights via “expert opinion” surveys. Additionally, 25% of the articles referencing weighting cite third-party providers that use proprietary schemes, such as Bloomberg ESG’s industry-group weighting, without providing methodological details (e.g., Radu & Smaili, 2021; Yu et al., 2018). Notably, articles using KLD/MSCI sustainability measures were more likely to engage in discussions about weighting (see Table 2), likely due to the disaggregated structure of KLD/MSCI data, because this forces researchers to make explicit decisions regarding fungibility (e.g., whether to combine “strengths” and “concerns”). In contrast, no statistically significant relationship was observed between weighting discussions and the use of Refinitiv or Bloomberg ESG measures.

*** [Insert Figure 2 about here] ***

Although aggregation inherently assumes fungibility, it was addressed less frequently than weighting. However, these two elements were often discussed together (see Table 2). Fungibility

discussions were most common in KLD/MSCI-based research (e.g., Waddock & Graves, 1997; Flammer, 2013; Kim, Kim & Qian, 2018; Zhao & Murrell, 2022). These articles explore whether a “strength” can offset a “concern” with several authors arguing that aggregating them into a single “net” score (strengths minus concerns) is inappropriate, rejecting the assumption that positive and negative social performance are fungible.

Comparing Busch et al. (2022) and Awaysheh et al. (2020) illustrates the consequence of fungibility assumptions. Both relied on KLD/MSCI data and used Tobin’s Q as their measure of financial performance as well as looking at a similar range of years (2005 to 2014 and 2003 to 2013, respectively). However, Busch et al. (2022) treated strengths and concerns as non-fungible and estimated their effects separately; they reported a negative relationship between KLD strengths and Tobin’s Q. In contrast, Awaysheh et al. (2020) treated them as fungible by collapsing them into a single score, reporting a positive relationship between this net KLD score and Tobin’s Q. This divergence demonstrates how the decision to allow positive and negative indicators to offset one another directly influences the sign and interpretation of reported effects. Similarly, twelve articles analyzed environmental, social, and governance measures independently. While these approaches suggest a more granular view of sustainability, few studies articulated the theoretical implications of these choices in detail.

Taken together, these findings reveal a striking asymmetry: while a majority of studies (59%) report positive relationships between sustainability measures and financial performance, critical engagement with measurement quality remains limited. Nearly three-quarters of articles fail to address accuracy or reliability, and only 6% engage with timeliness. Aggregation choices receive more attention, yet even weighting and fungibility discussions often lack theoretical justification.

This pattern suggests that reported findings may reflect not only underlying relationships but also unexamined assumptions embedded in measure construction and provider selection.

Discussion

This review examines how scholars engage with third-party sustainability measures when investigating the relationship between sustainability and financial performance. Unexamined assumptions about data quality and aggregation can influence research findings. Our analysis reveals a pattern: researchers who scrutinize one quality element (e.g., reliability) tend to scrutinize others, while many scrutinize none. This suggests researchers either approach sustainability data with critical skepticism or accept it wholesale. Rather than treating scores as “black boxes” to be unpacked, most researchers simply accept provider aggregation choices without discussion (Boiral et al., 2021; Gangi et al., 2022).

Most articles relied on an equal-weights approach (e.g., McWilliams & Siegel, 2000; Zhao & Murrell, 2022), adopting aggregation assumptions with little justification. This approach is not neutral; it implicitly assigns identical importance to all dimensions. While some scholars derived weights through expert panels (Waddock & Graves, 1997) or adopted provider-specific schemes (Petitjean, 2019; Yu & Luu, 2018), such transparency remains the exception. Regardless of the approach, researchers must document the trade-offs and potential biases implied by their weighting decisions. Hull and Rothenberg (2008, p. 784) provide a good example, noting: “[Our] approach has the advantage of providing a numerical score [and] it is more easily reproduced by future researchers than is the weighted index described by Waddock and Graves (1997), though

the weights they describe appear to correspond fairly well with ours.” However, such transparency is rare.

Fungibility is inherent in aggregation. This logic risks masking poor performance in one area with excellence in another (Delmas & Doctore-Blass, 2010). To evaluate these aggregation choices, researchers need to specify which indicators are combined and where compensatory logic is permitted. Our review found varied, often unarticulated, approaches: some treated all the data as fungible (e.g., Blanco et al., 2013; Godfrey et al., 2009), others treated sub-dimensions, like KLD/MSCI’s “strengths” and “concerns,” as non-fungible (e.g., Busch et al., 2022; Delmas et al., 2015; Flammer, 2013; Kim et al., 2018; Van der Laan et al., 2008), and some treated broader categories such as “environmental” and “social” as non-fungible (e.g., Janney & Gove, 2011; Oikonomou et al., 2014). These differences complicate cross-study comparison.

Jell-Ojobor and Raha (2022, p. 1970) provide a good example of explaining their fungibility approach: “We grouped the environmental scores into four main GSCM [Green Supply Chain Management] dimensions [from KLD]: (1) pollution and waste, (2) natural capital, (3) environmental opportunities, and (4) climate change. These scores reflect internal GSCM practices, such as sourcing and using water and energy for core business operations, and external GSCM practices, such as implementing programs with suppliers to reduce their carbon footprint.” This specificity helps move scholarship beyond implicit reliance on third-party scores toward more interpretable and theoretically grounded measurement choices, including explicit assumptions about where substitution across dimensions is (and is not) allowed.

These aggregation choices also intersect with contested notions of materiality in sustainability reporting. Traditional financial materiality, which focuses on information relevant to investor

decisions, can systematically exclude social and environmental impacts that are material to other stakeholders (Adams, 2015, 2017; Unerman et al., 2018). When third-party providers adopt investor-centric materiality frameworks, they may inadvertently reinforce a narrow view of corporate accountability that privileges financial returns over broader sustainability outcomes (Gray, 2010). This dynamic exacerbates the problem Schaltegger (1997) warned of: because high-quality environmental information is costly to produce and difficult for diverse stakeholders to verify, poor-quality or overly narrow metrics can easily dominate the landscape. This has direct implications for empirical research: studies using materiality-weighted measures (such as those from Bloomberg or Sustainalytics) embed these normative choices into their research designs, potentially limiting the generalizability of findings to contexts where stakeholder materiality differs from financial materiality.

Assumptions about accuracy and reliability also shape interpretation. Mainstream accounting literature emphasizes that information quality is not absolute, but is contingent on the specific decision-making context (Dechow et al., 2010). Consequently, directly observed measures (e.g., employee demographics) differ from modeled measures (e.g., Scope 2 GHG emissions estimates) and imputed measures (e.g., annualized charitable giving); each introduces distinct sources of error and ambiguity about what is being captured (e.g., a point-in-time snapshot versus a year-average, or a modeled baseline versus a firm-specific outcome). Unreliable measures with unknown margins of error can distort results. For example, Radu and Smaili (2021, p. 3354) note unreliability in Bloomberg's ESG measures, stating that "Given the nature of [Bloomberg's] sources of information, some gaps could exist between self-reported disclosure and actual performance." Similarly, Duque-Grisales and Aguilera-Caracuel (2021, p. 330) remark that "the

score assigned to each [Refinitiv] variable is not free of subjective influences, which may decrease the validity of our results.”

Timeliness is the most neglected element. While traditional financial accounting research treats timeliness as a core attribute of information quality that has a direct, measurable impact on a firm’s cost of equity (Francis et al., 2004), researchers utilizing sustainability scores typically assume sustainability scores reflect the reporting year, ignoring temporal mismatches.

Sustainability reports are typically published months after the end of a fiscal year, and third-party providers may take several more months to collect, verify, and aggregate that data into a formal score. Consequently, scores often lag the performance they supposedly measure. This lag can distort causal inference about sustainability-performance links (Delmas & Doctore-Blass, 2010). Without explicit attention to these temporal structures, researchers risk misaligning sustainability measures with financial metrics, obscuring the relevant window of market reaction or performance impact, and weakening the internal validity of causal claims.

In summary, the literature exhibits limited and uneven engagement with third-party sustainability measures: scrutiny is selective, attention clusters around certain elements, and many studies rely on implicit assumptions. This overlooks the measurement uncertainty introduced by provider-side design choices, from data modeling to analyst judgment. Because the choice of provider co-varies with the direction and framing of reported results, we conclude that these measurement attributes are not neutral.

Recommendations

We recommend several approaches to address the limitations of third-party sustainability data.

Use multiple independent sustainability measures. For instance, use Bloomberg's materiality-weighted score alongside equal-weighted KLD/MSCI's strengths/concerns. If both approaches yield similar results, confidence increases. If they diverge, this reveals how measurement choices shape conclusions. Such transparency clarifies the assumptions embedded in the chosen measure and acknowledges its limitations. Given the well-documented divergence among sustainability metrics (Berg et al., 2022), one effective strategy is to replicate analyses using multiple independent data sources. This approach, recommended by Dorfleitner et al. (2015) and Widyawati (2021), enhances robustness and supports broader generalizability. For instance, Delmas et al. (2013) combined environmental ratings from three major providers to extract key components of corporate environmental performance. In contrast, studies relying on a single measure may demonstrate a correlation between that specific metric and financial performance but risk overstating the generalizability of their findings to broader notions of sustainability.

Analyze the sensitivity of sustainability measures to inaccuracy and unreliability. One approach is to introduce random noise into sustainability measures within a realistic range and rerun analyses. If conclusions remain stable, measurement quality may be less critical to the findings. If results change substantially, this indicates that findings depend heavily on provider accuracy. By incorporating a defined range of measurement error as a robustness check, scholars can evaluate how robust their inferences are to deteriorating data quality. While no articles in our review employed such approaches, Berchicci and King (2022) provide a compelling example of how this methodological rigor could be applied.

Align lagged and unlagged variables based on the period the sustainability measures may represent. The timing of sustainability data is critical for research design, yet it is often

uncertain. Providers frequently revise their methodologies, potentially altering scores in ways unrelated to a firm's actual performance (Cho et al., 2015; Gillan et al., 2021). Moreover, ratings may reflect prior-year performance rather than the year of publication. For instance, a 2025 score may be based on 2024 data, making it appropriate to pair with a 2025 financial outcome in a study using ESG data to predict CFP. In the absence of clear information about the period the rating represents, researchers should test both lagged and unlagged specifications to assess the sensitivity of their results. Preston and O'Bannon (1997) and Nakao et al. (2007) employed this strategy and found consistent directional outcomes. Without explicit alignment between the timing of sustainability measures and financial variables, claims about causality or performance effects remain tentative.

Ensure an appropriate level of aggregation. Aggregation of sustainability measures should reflect the study's analytical aims. Broader questions may warrant composite indicators, whereas disaggregated measures are better suited to isolate the effects of specific sustainability dimensions. For example, Petitjean (2019) found no relationship between Bloomberg's environmental score and ROA, whereas Minutolo et al. (2019) identified a positive relationship using Bloomberg's aggregated ESG score. Disaggregated approaches help reduce assumptions associated with fungibility and weighting by treating environmental, social, and governance dimensions as distinct constructs, allowing for a clearer understanding of each measure's contribution to financial performance.

Assess sensitivity to weighting schemes. Few studies evaluate how alternative weighting schemes affect results. Surroca et al. (2020), using Refinitiv data, offer a good example through testing multiple weighting schemes to assess robustness. When working with disaggregated data, such

as KLD/MSCI’s “Strengths” and “Concerns,” one can simulate a range of aggregate scores by applying different weightings to the individual components. This approach reveals how sensitive findings are to methodological choices. Hall and Rothenberg (2008) followed this strategy by comparing results across two prior studies that used different aggregation methods with KLD/MSCI data (i.e., McWilliams & Siegel, 2000, and Waddock & Graves, 1997).

The measurement limitations documented in this review do not affect researchers alone.

Investors, managers, employees, and regulators all rely on third-party sustainability measures and are equally exposed to the uncertainties embedded in them (Amel-Zadeh & Serafeim, 2018; Clementino & Perkins, 2021; European Parliament & Council of the European Union, 2024; Welch & Yoon, 2023).

Investors should be careful not to over-interpret any given study and build an investment strategy on its findings. They should consult various ratings to avoid overreliance on a single source, just as they seek diverse perspectives in traditional financial research. In addition, changes in sustainability measures, as some of the articles in this review show, can affect investment returns. Investors need to understand what can drive these changes, encompassing both the underlying metrics and their relative weighting. This requires an understanding of when sustainability measures are updated and the period they reflect. Considering these temporal structures can help investors more accurately assess the impact of sustainability measures on investment analyses.

Managers who rely on sustainability ratings should recognize that provider methodologies differ substantially—in what they measure, how they weight dimensions, and when scores are updated. Optimizing for a single provider's score risks overlooking dimensions that matter to other raters.

A more robust approach is to monitor performance across multiple providers, recognizing that ratings can shift due to changes in provider methodology that are unrelated to the firm's actual sustainability performance.

Policy makers face challenges in detecting greenwashing when sustainability measures allow strengths in one area to offset weaknesses in another (Montgomery et al., 2023). Requiring more disaggregated disclosures and limiting compensatory scoring can improve transparency and reduce misleading claims. Although policymakers may not rely directly on third-party scores, research using this data can help identify misalignments between fund-level sustainability claims and actual portfolio composition. More broadly, emerging mandatory reporting requirements—including the International Sustainability Standards Board (ISSB) framework, California's SB 261 and SB 253, and the European Union's Corporate Sustainability Reporting Directive (CSRD)—represent a structural opportunity to reduce dependence on opaque third-party scores. By mandating standardized, verifiable disclosures, these frameworks could substantially improve the accuracy, reliability, and comparability of underlying sustainability data, directly addressing the quality concerns this review documents.

In sum, these recommendations can strengthen the validity of sustainability–CFP research and support more informed analysis by stakeholders relying on widely used sustainability measures.

Conclusions, Limitations, and Future Research

Our systematic review of 82 articles yields two main findings. First, engagement with the five framework elements is limited and uneven: only 24% of reviewed studies addressed accuracy, 28% reliability, and a mere 6% timeliness. Second, methodological rigor is distinctly bimodal:

articles discussing accuracy are 10.5 times more likely to also discuss reliability ($OR = 10.5, p = 0.002$), indicating that researchers either approach third-party data with comprehensive skepticism or treat it as validated input requiring no scrutiny. These patterns matter because unexamined measurement choices are not merely omissions—they are active confounds. When researchers adopt a provider's score without interrogating its construction, they import that provider's scope, estimation, and aggregation decisions directly into their research design, where those decisions shape the variance and direction of the independent variable in ways that standard model controls cannot address. This scrutiny, however, remains rare.

We examined the quality elements related to “how” sustainability measures are created and how uncertainty is addressed. By focusing on use rather than construction alone, we extend prior conceptual critiques (Damtoft et al., 2025) and bring attention to researchers' practices. Whereas Damtoft et al. (2025) develop a normative framework for context-specific measurement design, the present review documents how practicing researchers actually engage—or fail to engage—with the uncertainty embedded in the third-party scores they adopt.

These findings have several implications. First, the field lacks shared standards for methodological rigor—researchers either scrutinize several dimensions of measurement or not at all, with little middle ground. Second, the provider differences, which are often not critically examined in the studies using that data, challenge the assumption that sustainability ratings measure a common underlying construct.

Provider opacity complicates researcher efforts to assess measurement elements (Boiral et al., 2021; Delmas et al., 2013; Escrig-Olmedo et al., 2014; Kotsantonis & Serafeim, 2019).

However, when researchers cannot determine the quality and aggregation methods of the

underlying data, they must examine how inaccuracy, unreliability, timeliness, and different aggregation approaches may affect their analyses. Our analysis shows that when such examination occurs, it clusters around reliability, fungibility, and weighting, while timeliness remains neglected. Neglecting these uncertainties risks overstating the validity of findings.

This review is not without limitations. Our Web of Science and ProQuest searches focused on business journals, limiting our scope to how business research has used sustainability measures. Other fields and interdisciplinary research also use sustainability measures, but we excluded publications from those fields from our review. As a result, the systematic associations we identify reflect patterns within business research and may differ in fields where sustainability measures are theorized or operationalized differently.

Restricting the analysis to high-quality peer-reviewed journals introduces potential selection bias. First, articles with serious measurement problems may have been rejected during peer review, meaning our findings may underestimate the extent of measurement issues in the broader research population. Second, publication bias may favor positive findings regardless of measurement quality, which could explain some of the associations we observe between measurement attention and reported effects. Future research could address this by analyzing working papers or rejected manuscripts, or by surveying researchers directly about their measurement decision-making—approaches that would reveal whether the selective engagement we document is even more pronounced outside the published record.

We deliberately do not evaluate the substantive dimensions of what sustainability measures are intended to capture. These dimensions, such as completeness, commensurability, materiality, and balance, have been emphasized in broader information quality research (e.g., Nelson et al.,

2005), as well as standards for sustainability reporting (Schaltegger & Burritt, 2000), and are essential for understanding the substantive validity of sustainability indicators. However, incorporating these dimensions would require engaging with firm-level disclosure choices, stakeholder prioritization, and normative judgments about sustainability performance, issues that lie outside of the focus of this review.

Future literature reviews could extend our approach by examining how researchers justify the selection of sustainability measures with respect to substantive dimensions such as materiality, balance, and completeness, and whether attention to these dimensions is likewise associated with systematic differences in research findings. Such research could move the corporate sustainability field closer to developing measures that better reflect firms' "true" sustainability. Nonetheless, our study shows that understanding and explicitly managing challenges related to quality and aggregation remain essential for credible and transparent research.

While comparing effect sizes across providers would be valuable, heterogeneity in model specifications prevented such analysis. For instance, Jeriji et al. (2023) and Yu et al. (2018) both predict Tobin's Q using global samples, but with different providers (Refinitiv and Bloomberg respectively) and different specifications. Yu et al. (2018) include ESG and ESG² as independent variables, while Jeriji et al. (2023) focus on GRI disclosure with CSP as a control variable, including a GRI × CSP interaction. Attributing coefficient differences to provider choice rather than model specification would therefore be unjustified. Similarly, Busch et al. (2022) and Minutolo et al. (2019) both examine US large cap firms using KLD/MSCI and Bloomberg data respectively, but with incompatible designs: Busch et al. (2022) focus on GHG emissions with KLD/MSCI strengths and weaknesses as separate controls, while Minutolo et al. (2019) use

Bloomberg scores as the main independent variable and include ROA as a predictor of Tobin's Q. These specification differences make meaningful effect size comparisons infeasible. Even Friede et al.'s (2015) comprehensive meta-analysis relied on vote-counting and simple correlations rather than standardized effect size comparisons across studies.

A further boundary condition concerns financial performance measurement—our review focused only on articles using four indicators: Tobin's Q, ROA, MVA, and CAR. While these are widely accepted metrics, many other financial performance measures exist. For example, profitability ratios such as profit margins or return on equity, and market value indicators such as price-to-earnings ratios or earnings per share, are also well-established methods for assessing CFP. Moreover, financial performance is not the only relevant outcome for research. Studies examining the impact of sustainability measures on environmental and social outcomes have reported concerning findings (Kathan et al., 2025; Raghunandan & Rajgopal, 2022). Future studies could examine whether the patterned relationships we observe between measurement emphases and outcome direction persist across alternative financial and non-financial performance indicators. Overall, further research is needed to explore how firm performance, whether financial or non-financial, is measured and understood.

Lastly, we specifically examined sustainability measures from third-party providers. Researchers can and do construct their measures from data they collect instead of using third-party data (Delmas et al. 2025). Although this approach introduces additional challenges, the elements of measurement quality and aggregation remain relevant, without the opacity created by third-party providers. With advancements in large language models and machine learning, the rapid and accurate collection of firsthand sustainability data is becoming less labor-intensive, reducing

researchers' and stakeholders' dependence on proprietary third-party sources. These developments may help bridge the gap between voluntary disclosures and mandatory reporting requirements, while also enabling researchers to test whether greater transparency reduces the systematic associations between measurement choices and reported outcomes identified in this review.

This review does not make a blanket argument against the use of sustainability measures in research. Instead, it responds to a desire to simplify the complex task of measuring corporate sustainability, attempting to provide a clear and comprehensive overview of how sustainability measures are used. Our five-elements framework echoes long-standing principles in information quality assessment (Nelson et al., 2005), while tailoring them to the unique challenges of sustainability measurement. Its goal is not to highlight the well-established weaknesses of these metrics but to introduce a framework of elements that define their construction quality and to offer actionable guidance for addressing known limitations. By showing that measurement choices co-vary with the direction and framing of empirical results, our review underscores why interrogating these assumptions is a substantive concern for theory development. The failure to do so contributes to the continued legitimacy of opaque and inconsistent metrics (King & Berchicci, 2021; Berg et al., 2022). As the title of this article suggests, without explicit scrutiny of these methodological “forking paths,” scholarship risks being built on sand, where findings are less a reflection of corporate reality and more an artifact of unexamined measurement infrastructure.

Disclosures

Use of AI: This work was supported by AI for assistance with copy editing of the text and survey questions. The author maintains full control over the final content.

Conflicts: The author declares no potential conflicts of interest with respect to the research or authorship of this article.

Data Availability Statement: The data for this study is available at Harvard Dataverse.

<https://doi.org/10.7910/DVN/VXTAOA>.

References

- Adams, C. A. (2015). The international integrated reporting council: A call to action. *Critical Perspectives on Accounting*, 27, 23–28.
- Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10), 4451–4469.
- Alshehhi, A., Nobanee, H., & Khare, N. (2018). The impact of sustainability practices on corporate financial performance: Literature trends and future research potential. *Sustainability*, 10(2), 494.
- Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74(3), 87–103.
- Awaysseh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, 41(6), 965–987. <https://doi.org/10.1002/smj.3122>
- Bebbington, J., & Larrinaga, C. (2014). Accounting and sustainable development: An exploration. *Accounting, Organizations and Society*, 39(6), 395–413.
- Berchicci, L., & King, A. (2022). Building knowledge by mapping model uncertainty in six studies of social and financial performance. *Strategic Management Journal*, 43(7), 1319–1346.
- Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.
- Blanco, B., Guillamón-Saorín, E., & Guiral, A. (2013). Do non-socially responsible companies achieve legitimacy through socially responsible actions? The mediating effect of innovation. *Journal of Business Ethics*, 117(1), 67–83.
- Bloomberg. (2021). *Bloomberg ESG disclosure scores: Methodology overview*. Bloomberg Professional Services.

- Bloomberg. (n.d.). *Environmental & social (ES) scores: Assess sustainability performance with transparent, data-driven scores*.
https://data.bloomberglp.com/professional/sites/10/ESG_Environmental-Social-Scores.pdf
- Boiral, O., Brotherton, M. C., & Talbot, D. (2020). Building trust in the fabric of sustainability ratings: An impression management perspective. *Journal of Cleaner Production*, 260, Article 120942.
- Boiral, O., Talbot, D., Brotherton, M. C., & Heras-Saizarbitoria, I. (2021). Sustainability rating and moral fictionalism: Opening the black box of nonfinancial agencies. *Accounting, Auditing & Accountability Journal*, 34(8), 1740–1768.
- Burbano, V. C., Delmas, M. A., & Cobo, M. J. (2024). The past and future of corporate sustainability research. *Organization & Environment*, 37(2), 133–158.
- Burritt, R. L., & Schaltegger, S. (2010). Sustainability accounting and reporting: Fad or trend? *Accounting, Auditing & Accountability Journal*, 23(7), 829–846.
- Busch, T., & Friede, G. (2018). The robustness of the corporate social and financial performance relation: A second-order meta-analysis. *Corporate Social Responsibility and Environmental Management*, 25(4), 583–608.
- Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate carbon and financial performance revisited. *Organization & Environment*, 35(1), 154–171.
- Busco, C., Granà, F., & Achilli, G. (2021). Understanding integrated thinking: Evidence from the field, the development of a framework and avenues for future research. *Meditari Accountancy Research*, 29(4), 673–690.
- Byerly, H. C., & Lazara, V. A. (1973). Realist foundations of measurement. *Philosophy of Science*, 40(1), 10–27.
- Calantone, R. J., & Vickery, S. K. (2010). Introduction to the special topic forum: Using archival and secondary data sources in supply chain management research. *Journal of Supply Chain Management*, 46(4), 3.
- Capelle-Blancard, G., & Petit, A. (2017). The weighting of CSR dimensions: One size does not fit all. *Business & Society*, 56(6), 919–943.
- 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.
- Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, 18(1), 125–169.
- Chen, C. M., & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and Operations Management*, 20(6), 789–804.
- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of Environmental Management*, 345, Article 118829.
- Cho, C. H., Michelon, G., Patten, D. M., & Roberts, R. W. (2015). CSR disclosure: The more things change...? *Accounting, Auditing & Accountability Journal*, 28(1), 14–35.
- Christensen, D. M., Serafeim, G., & Sikochi, A. (2021). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, 97, 147–175.

- Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248.
- Chytis, E., Eriotis, N., & Mitroulia, M. (2024). ESG in business research: A bibliometric analysis. *Journal of Risk and Financial Management*, 17(10), 460.
- Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. *Journal of Business Ethics*, 171(2), 379–397.
- Clifford, N., Cope, M., Gillespie, T. W., & French, S. (Eds.). (2016). *Key methods in geography*. SAGE.
- Coelho, R., Jayantilal, S., & Ferreira, J. J. (2023). The impact of social responsibility on corporate financial performance: A systematic literature review. *Corporate Social Responsibility and Environmental Management*, 30(4), 1535–1560.
- Cort, T., & Esty, D. (2020). ESG standards: Looming challenges and pathways forward. *Organization & Environment*, 33(4), 491–510.
- Damtoft, N. F., van Liempd, D., & Lueg, R. (2025). Sustainability performance measurement: A framework for context-specific applications. *Journal of Global Responsibility*, 16(1), 162–201.
- Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3), 344–401.
- Delmas, M. A., Buskard, D., Li, J., & Timmer, T. (2025). *The state of corporate sustainability disclosure 2025*. <https://doi.org/10.2139/ssrn.5269059>
- Delmas, M. A., Clark, K., Timmer, T., & McClellan, M. (2022). *The state of corporate sustainability disclosure*. SSRN. <https://ssrn.com/abstract=4194032>
- Delmas, M. A., & Doctori-Blass, V. (2010). Measuring corporate environmental performance: The trade-offs of sustainability ratings. *Business Strategy and the Environment*, 19(4), 245–260.
- Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture? *Academy of Management Perspectives*, 27(3), 255–267.
- Dhami, M. K., & Mandel, D. R. (2022). Communicating uncertainty using words and numbers. *Trends in Cognitive Sciences*, 26(6), 514–526.
- Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG ratings. *The Journal of Portfolio Management*, 47(1), 75–87.
- Dobbie, M. J., & Dail, D. (2013). Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecological Indicators*, 29, 270–277.
- Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility: An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, 16(7), 450–466.
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123.
- Endrikat, J., Guenther, E., & Hoppe, H. (2014). Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal*, 32(5), 735–751.

- Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. Á., & Rivera-Lirio, J. M. (2014). Lights and shadows on sustainability rating scoring. *Review of Managerial Science*, 8(4), 559–574.
- Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. Á., & Rivera-Lirio, J. M. (2017). Measuring corporate environmental performance: A methodology for sustainable development. *Business Strategy and the Environment*, 26(2), 142–162.
- European Parliament, & Council of the European Union. (2024). Regulation (EU) 2024/3005 of the European Parliament and of the Council of 27 November 2024 on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities, and amending Regulations (EU) 2019/2088 and (EU) 2023/2859. *Official Journal of the European Union*, L 2024/3005, 1–48.
- Figge, F., Hahn, T., Schaltegger, S., & Wagner, M. (2002). The sustainability balanced scorecard: Linking sustainability management to business strategy. *Business Strategy and the Environment*, 11(5), 269–284.
- Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences*, 111(Suppl. 4), 13664–13671.
- Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758–781.
- Francis, J., LaFond, R., Olsson, P. M., & Schipper, K. (2004). Costs of equity and earnings attributes. *The Accounting Review*, 79(4), 967–1010.
- Frankel, R., Kothari, S. P., & Raghunandan, A. (2025). The economics of ESG disclosure regulation. *Review of Accounting Studies*. Advance online publication.
- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233.
- Friedman, H. L., & Ormazabal, G. (2024). The role of information in building a more sustainable economy: A supply and demand perspective. *Journal of Accounting Research*, 62(1), 1–54.
- Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., & Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81, 491–502.
- Gangi, F., Varrone, N., Daniele, L. M., & Coscia, M. (2022). Mainstreaming socially responsible investment: Do environmental, social and governance ratings of investment funds converge? *Journal of Cleaner Production*, 353, Article 131684.
- Gao, J., & Bansal, P. (2013). Instrumental and integrative logics in business sustainability. *Journal of Business Ethics*, 112(2), 241–255.
- García-Sánchez, I., & Martínez-Ferrero, J. (2019). Chief executive officer ability, corporate social responsibility, and financial performance: The moderating role of the environment. *Business Strategy and the Environment*, 28(4), 542–555.
- Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66, Article 101889.
- Global Reporting Initiative. (2016). *GRI 101: Foundation 2016*.
<https://www.globalreporting.org/standards/media/1036/gri-101-foundation-2016.pdf>

- Graafland, J. J., Eijffinger, S. C., & Smid, J. H. (2004). Benchmarking of corporate social responsibility: Methodological problems and robustness. *Journal of Business Ethics*, 53(1), 137–152.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.
- Gray, R. (2010). Is accounting for sustainability actually accounting for sustainability... and how would we know? An exploration of narratives of organisations and the planet. *Accounting, Organizations and Society*, 35(1), 47–62.
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1), 61–94.
- Grewal, J., & Serafeim, G. (2020). Research on corporate sustainability: Review and directions for future research. *Foundations and Trends in Accounting*, 14(2), 73–173.
- Harrison, J. S., Yu, X., & Zhang, Z. (2023). Consistency among common measures of corporate social and sustainability performance. *Journal of Cleaner Production*, 391, Article 136232. <https://doi.org/10.1016/j.jclepro.2023.136232>
- Harzing, A. (2023, September 5). *Journal Quality List*. <https://harzing.com/download/jq170a-title.pdf>
- Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1-3), 405–440.
- Huang, D. Z. X. (2021). Environmental, social and governance (ESG) activity and firm performance: A review and consolidation. *Accounting & Finance*, 61, 335–360.
- Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries*, 3(3), 619–626.
- Kathan, M. C., Utz, S., Dorfleitner, G., Eckberg, J., & Chmel, L. (2025). What you see is not what you get: ESG scores and greenwashing risk. *Finance Research Letters*, 74, Article 106710.
- Kim, K.-H., Kim, M., & Qian, C. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management*, 44(3), 1097–1118.
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284.
- King, A., & Berchicci, L. (2021). Mapping the garden of forking paths: The case of social & financial performance. *Strategic Management Journal*, 43(7), 1319–1346.
- Kong, Y., Antwi-Adjei, A., & Bawuah, J. (2019). A systematic review of the business case for corporate social responsibility and firm performance. *Corporate Social Responsibility and Environmental Management*, 27(2), 444–454.
- Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, 31(2), 50–58.
- Lamberton, G. (2005). Sustainability accounting—a brief history and conceptual framework. *Accounting Forum*, 29(1), 7–26.

- Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*, 33, 59–78.
- Montgomery, W., Lyon, T., Barg, J., & Lynch, M. (2023). *Greenwashing 3.0: Why addressing greenwashing remains as important as ever (and what can be done about it)*. Ivey Business School. https://issuu.com/erbminstitute/docs/greenwashing_3.0_-_spreads?fr=xIAEoAT3_NTU1
- Munda, G., & Nardo, M. (2005). *Constructing consistent composite indicators: The issue of weights* (EUR 21834 EN). European Commission Joint Research Centre.
- Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics*, 41(12), 1513–1523.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Communicating science effectively: A research agenda*. The National Academies Press.
- National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press.
- Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, 21(4), 199–235.
- Oikonomou, I., Brooks, C., & Pavelin, S. (2014). The financial effects of uniform and mixed corporate social performance. *Journal of Management Studies*, 51(6), 898–925.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, Article 105906.
- Panic, N., Leoncini, E., De Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLOS ONE*, 8(12), Article e83138.
- Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: Voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–634.
- Petitjean, M. (2019). Eco-friendly policies and financial performance: Was the financial crisis a game changer for large US companies? *Energy Economics*, 80, 502–511.
- Pizzi, S., Principale, S., & De Nuccio, E. (2023). Material sustainability information and reporting standards. Exploring the differences between GRI and SASB. *Meditari Accountancy Research*, 31(6), 1654–1674.
- Rabinovich, S. (2005). *Measurement errors and uncertainties: Theory and practice* (3rd ed.). AIP Press.
- Raghunandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, 27(3), 822–863.
- Schaltegger, S. (1997). Information costs, quality of information and stakeholder involvement: The necessity of international standards of ecological accounting. *Eco-Management and Auditing*, 4(3), 87–97.
- Schaltegger, S., & Burritt, R. (2000). *Contemporary environmental accounting: Issues, concepts and practice*. Greenleaf Publishing.
- Schaltegger, S., & Burritt, R. L. (2010). Sustainability accounting for companies: Catchphrase or decision support for business leaders? *Journal of World Business*, 45(4), 375–384.

- Serafeim, G. (2023). ESG: From process to product. *Harvard Business School Accounting & Management Unit Working Paper*, (23-069).
- Simpson, J. A. (1981). Foundations of metrology. *Journal of Research of the National Bureau of Standards*, 86(3), 281.
- Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, 9(2), 189–212.
- Stewart, D. W., & Kamins, M. A. (1993). *Secondary research: Information sources and methods* (Vol. 4). SAGE Publications.
- Szomszor, M. (2021, May 20). *Introducing the Journal Citation Indicator: A new, field-normalized measurement of journal citation impact*. Clarivate.
<https://clarivate.com/blog/introducing-the-journal-citation-indicator-a-new-field-normalized-measurement-of-journal-citation-impact/>
- Tayan, B. (2022). *ESG ratings: A compass without direction*. Harvard Law School Forum on Corporate Governance. <https://corpgov.law.harvard.edu/2022/08/24/esg-ratings-a-compass-without-direction/>
- The Global Compact. (2004). *Who cares wins*. United Nations.
https://www.unepfi.org/fileadmin/events/2004/stocks/who_cares_wins_global_compact_2004.pdf
- Unerman, J., Bebbington, J., & O'Dwyer, B. (2018). Corporate reporting and accounting for externalities. *Accounting and Business Research*, 48(5), 497–522.
- Van Beurden, P., & Gössling, T. (2008). The worth of values—A literature review on the relation between corporate social and financial performance. *Journal of Business Ethics*, 82, 407–424.
- van der Bles, A. M., van der Linden, S., Freeman, A. L., & Spiegelhalter, D. J. (2020). The effects of communicating uncertainty on public trust in facts and numbers. *Proceedings of the National Academy of Sciences*, 117(14), 7672–7683.
- van der Bles, A. M., van der Linden, S., Freeman, A. L., Mitchell, J., Galvao, A. B., Zaval, L., & Spiegelhalter, D. J. (2019). Communicating uncertainty about facts, numbers and science. *Royal Society Open Science*, 6(5), Article 181870.
- Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1-3), 97–180.
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
- Welch, K., & Yoon, A. (2023). Do high-ability managers choose ESG projects that create shareholder value? Evidence from employee opinions. *Review of Accounting Studies*, 28(4), 2448–2475.
- Widyawati, L. (2020). A systematic literature review of socially responsible investment and environmental social governance metrics. *Business Strategy and the Environment*, 29(2), 619–637.
- Widyawati, L. (2021). Measurement concerns and agreement of environmental social governance ratings. *Accounting & Finance*, 61, 1589–1623.
- Yu, E. P., Guo, C. Q., & Luu, B. V. (2018). Environmental, social and governance transparency and firm value. *Business Strategy and the Environment*, 27(7), 987–1004.

Zhao, X., & Murrell, A. (2022). Does a virtuous circle really exist? Revisiting the causal linkage between CSP and CFP. *Journal of Business Ethics*, 177(1), 173–192.

Tables and Figures

Table 1. Elements of Measure and Aggregation Quality

Element Category	Element	Element Definition	References
Measure Quality	Accuracy	“The accuracy of a measurement reflects how close the result is to the true value of the measured quantity” (Rabinowich, 2005, p. 2)	Cort & Esty, 2020; Nelson et al., 2005; Rabinowich, 2005; Simpson, 1981; Sørensen et al., 1996; Stewart & Kamins, 1993
	Reliability	“Reliability is the extent to which measurements are repeatable” (Drost, 2011, p. 106)	Drost, 2011; Kimberlin & Winterstein, 2008; Kotsantonis & Serafeim, 2019; Nelson et al., 2005; Simpson, 1981; Widyawati, 2020
	Timeliness	“Timeliness refers to the regularity of reporting [...] relative to the reporting period [and] the time period to which [the measure] relates” (GRI, 2016, p. 16)	Clifford et al., 2016; Global Reporting Initiative, 2016; Nelson et al., 2005; Simpson, 1981; Stewart & Kamins, 1993
Measure Aggregation	Fungibility	“[Fungibility is when] a good score may compensate for a bad score” (Capelle-Blancard & Petit, 2017, p. 920)	Capelle-Blancard & Petit, 2017; Delmas & Doctori-Blass, 2010; Chen & Delmas, 2011; Escrig-Olmedo et al., 2017; Escrig-Olmedo et al., 2014; Graafland et al., 2004; Munda & Nardo, 2005; Paruolo et al. 2013
	Weighting	“Weights of [measures] reflect the relative importance of different dimensions in their contributions to the sustainability performance of a system” (Gan et al., 2017, p. 492)	Capelle-Blancard & Petit, 2017; Chan & Delmas, 2011; Dobbie & Dail, 2011; Gan et al., 2017; Graafland et al., 2004; Greco et al., 2019; Munda & Nardo, 2009; Singh et al., 2009; Zhou et al., 2017

Table 2. Summary of Scales and Reported Ranges of Third-Party Scores Used ^{5,6}

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Gangwani & Kashiramka, 2024	Bloomberg	0.1 to 100	4.89	73.71	Positive	Bloomberg on Tobin's Q; Bloomberg on ROA
Hoang et al., 2020	Bloomberg	0.01 to 100	1.3793	82.1795	Positive	Bloomberg on ROA
Kumar et al., 2022	Bloomberg	0 to 100	NR	NR	Curvilinear	Bloomberg on ROA
Minutolo et al., 2019	Bloomberg	0 to 100	0	76	Positive	Bloomberg on Tobin's Q and ROA
Petitjean, 2019	Bloomberg	0.1 to 100	2.069	82.171	None	None
Radu & Smaili, 2021	Bloomberg	0.01 to 100	0.775	84.211	Positive	Bloomberg E score & S Score on ROA
Yu et al., 2018	Bloomberg	0 to 1	0.2	0.8678	Positive	Bloomberg on Tobin's Q
Taddeo et al., 2024	Bloomberg & LSEG	NR	57.34	135.712	Positive	Bloomberg + LSEG on ROA
Sroufe & Gopalakrishna-Remani, 2019	Bloomberg & Newsweek	0 to 100	NR	NR	Positive	Newsweek on ROA
Hossain et al., 2024	Choice's ESG & Rankings ESG	NR	0	23	Positive	Choices on CAR; Rankings on CAR
Sandberg et al., 2023	CSRHub	0 to 100	31	77	Positive	CSRHub on ROA
Preston & O'Bannon, 1997	Fortune	NR	NR	NR	Positive	Fortune on ROA
Woodroof et al., 2019	Fortune	0-10	NR	NR	Positive	Fortune on CAR
Cao et al., 2023	Hexun	NR	NR	NR	Positive	Hexun on Tobin's Q; Hexun on ROA

⁵ “NR” indicates that a value or range was not reported by the article. “IV” refers to “independent variable.”

⁶ Note: This table is not intended to summarize the literature on the sustainability measure–financial performance relationship, but is intended to detail the various sources and ways research has described sustainability measures.

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Xia et al., 2023	Hexun	NR	-6.39	85.8	Positive	Hexun on ROA
Xu et al., 2019	Hexun	NR	NR	NR	Negative	Hexun on Tobin's Q
He et al., 2023	Hexun & CNRDS	NR	-4.558	2.809	Negative	Hexun & CNRDS on ROA
Afrin et al., 2022	KLD/MSCI	NR	NR	NR	Both	KLD on CAR
Apaydin et al., 2021	KLD/MSCI	NR	NR	NR	Both	KLD on ROA
Awaysheh et al., 2020	KLD/MSCI	-20 to 20	-11	19	Positive	KLD on Tobin's Q
Barnett & Salomon, 2012	KLD/MSCI	0 to 30	0	27	Curvilinear	KLD on ROA
Blanco et al., 2013	KLD/MSCI	NR	-10	20	Positive	KLD on Tobin's Q
Brower & Dacin, 2020	KLD/MSCI	NR	2	22	Positive	KLD on Tobin's Q
Brower et al., 2017	KLD/MSCI	NR	-3	5	Positive	KLD on Tobin's Q
Busch et al., 2022	KLD/MSCI	NR	NR	NR	Negative	KLD Strengths on Tobin's Q
Chang et al., 2013	KLD/MSCI	NR	-3	-0.36	None	None
Delmas et al., 2015	KLD/MSCI	0 to 5	0	5	Negative	KLD Strengths on Tobin's Q
Deng et al., 2013	KLD/MSCI	NR	NR	NR	Positive	KLD on CAR
Flammer, 2013	KLD/MSCI	0 to 7	NR	NR	Both	KLD E Strengths on CAR; KLD E Concerns on CAR
Gao & Bansal, 2013	KLD/MSCI	0-5; 0 - 21	NR	NR	Positive	KLD Social on Tobin's Q
Godfrey et al., 2009	KLD/MSCI	NR	NR	NR	Positive	KLD on CAR
Hannah et al., 2021	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Hasan et al., 2018	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Hillman & Keim, 2001	KLD/MSCI	NR	NR	NR	Both	KLD on MVA
Hsu et al., 2019	KLD/MSCI	NR	NR	NR	Negative	KLD Concerns on CAR
Hull & Rothenberg, 2008	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Hyun et al., 2023	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Inoue & Lee, 2011	KLD/MSCI	NR	-3	5	Both	KLD on ROA & Tobin's Q
Janney & Gove, 2011	KLD/MSCI	0 to 7	NR	NR	Both	KLD Strengths on CAR; KLD Governance Strengths on CAR
Jell-Ojobor & Raha, 2022	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Kashmiri et al., 2017	KLD/MSCI	-7 to 7	NR	NR	Positive	KLD on CAR
Kim et al., 2018	KLD/MSCI	0 to 1	NR	NR	Positive	KLD Strengths on Tobin's Q; KLD Concerns on Tobin's Q
Lee et al., 2018	KLD/MSCI	NR	-3	6	None	None
Liu et al., 2020	KLD/MSCI	NR	-1.196	2.804	Positive	KLD on CAR
Pichler et al., 2018	KLD/MSCI	0 to 1	0	1	None	None
McWilliams & Siegel, 2000	KLD/MSCI	0 to 1	NR	NR	Positive	KLD on ROA
Sadovnikova & Pujari, 2017	KLD/MSCI	NR	NR	NR	Negative	KLD E Strengths on CAR
Shahzad & Sharfman, 2017	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Tang et al., 2012	KLD/MSCI	NR	NR	NR	Positive	KLD on ROA
Theodoulidis et al., 2017	KLD/MSCI	NR	-3	2	Positive	KLD on ROA and Tobin's Q

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Van der Laan et al., 2008	KLD/MSCI	NR	NR	NR	Negative	KLD Concerns on ROA
Waddock & Graves, 1997	KLD/MSCI	-2 to 2	NR	NR	Positive	KLD on ROA
Wang & Choi, 2013	KLD/MSCI	NR	NR	NR	Positive	KLD on Tobin's Q
Zhao & Murrell, 2016	KLD/MSCI	0 to 8	-1.347	2.462	None	None
Zhao & Murrell, 2022	KLD/MSCI & Sustainalytics	0 to 5; -5 to 5; 0 to 100	NR	NR	Both	KLD on Tobin's Q; Sustainalytics on ROA and Tobin's Q
Delmas et al., 2013	KLD/MSCI & SAM	NR	1.03	3.24	Positive	Composite of Scores on Tobin's Q
Lee et al., 2016	Korean Corporate Governance Service	NR	0	250	Positive	KCGS on ROA
Lee & Kwon, 2019	Newsweek	NR	22.3	85.58	Positive	Newsweek on MVA
Yadav et al., 2017	Newsweek	NR	NR	NR	Both	Newsweek on ROA; Newsweek on Tobin's Q
Nakao et al., 2007	Nikkei Environmental Management Survey	NR	NR	NR	Positive	Nikkei on ROA; Nikkei on Tobin's Q
Schreck, 2011	Oekom Research	1 to 4	1	3.86	Both	Oekom on Tobin's Q
Ben Lahouel et al., 2022	Refinitiv	0 to 100	8.43	97.22	Negative	Refinitiv on Tobin's Q and ROA
Brinette et al., 2023	Refinitiv	-100 to 0	-100	-0.143	Negative	Refinitiv on Tobin's Q
Candio, 2024	Refinitiv	0 to 100	NR	NR	Both	Refinitiv on ROA
Chen et al., 2023	Refinitiv	NR	6.23	84	Positive	Refinitiv on ROA
Duque-Grisales & Aguilera-Caracuel, 2021	Refinitiv	0 to 100	NR	NR	Negative	Refinitiv on ROA
Garcia & Orsato, 2020	Refinitiv	0 to 100	2.56	97.46	Both	Refinitiv on ROA

Article	Sustainability Measure(s) Used as IV	Scale of IV	Lowest Reported IV Value	Highest Reported IV Value	Direction of Statistically Significant Effect	Statistically Significant Relationship Found
Ibishova et al., 2024	Refinitiv	NR	2.33	94.77	Negative	Refinitiv on ROA
Iurkov et al., 2024	Refinitiv	NR	NR	NR	Positive	Refinitiv on CAR
Jeriji et al., 2023	Refinitiv	NR	NR	NR	Positive	Refinitiv on Tobin's Q
Lys et al., 2015	Refinitiv	0 to 1	NR	NR	Positive	Refinitiv on ROA
McGuinness et al., 2020	Refinitiv	NR	3.05	96.64	None	None
Moneva et al., 2020	Refinitiv	0 to 100	1.34	98.84	Negative	Refinitiv on ROA
Nekhili et al., 2021	Refinitiv	0 to 1	0.1925	0.9709	Positive	Refinitiv on Tobin's Q
Shin et al., 2023	Refinitiv	NR	NR	NR	Positive	Refinitiv on ROA
Shin et al., 2024	Refinitiv	NR	0	95.552	None	None
Surroca et al., 2020	Refinitiv	0 to 100	5	88	Positive	Refinitiv on Tobin's Q
Wang et al., 2011	Southern Weekend	0 to 100	27.735	79.572	Positive	Southern Weekend on CAR
Surroca et al., 2010	Sustainalytics	0 to 100	NR	NR	Positive	Sustainalytics on Tobin's Q
García-Sánchez & Martínez-Ferrero, 2019	VigeoEIRIS	NR	NR	NR	Both	EIRIS on Tobin's Q; EIRIS on ROA
Meier et al., 2021	VigeoEIRIS	0 to 1	0	1	Both	EIRIS on ROA
Wu & Shen, 2013	VigeoEIRIS	1 to 4	NR	NR	Positive	EIRIS on ROA

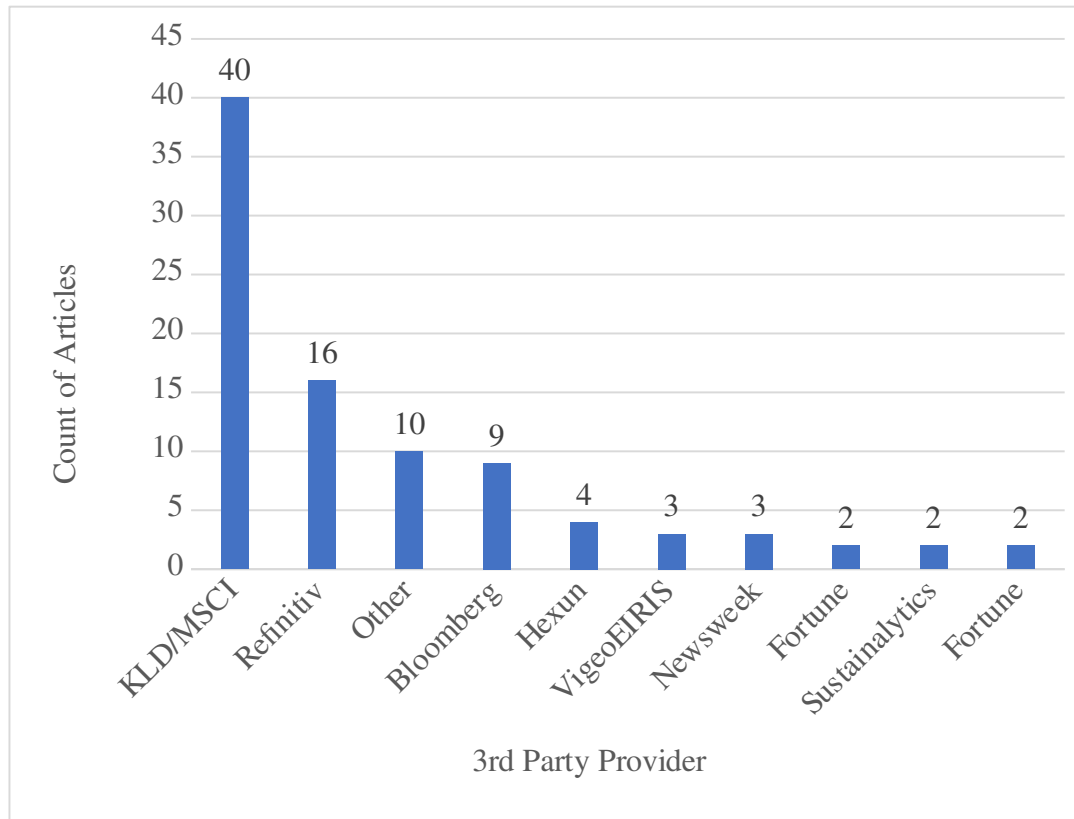
Table 3. ϕ Coefficient of Elements, Statistical Outcomes, and Measure Providers (N=82)⁷⁸

#	Variable	N	%	1	2	3	4	5	6	7	8	9	10
1	Accuracy Discussed	20	24.4%	-									
2	Reliability Discussed	23	28.0%	0.40**	-								
3	Timeliness Discussed	5	6.1%	-0.03	0.07	-							
4	Fungibility Discussed	44	53.7%	0.02	0.04	-0.27*	-						
5	Weighting Discussed	48	58.5%	-0.04	0.03	-0.20	0.56**	-					
6	Positive Stat. Sig. Effect	61	74.4%	0.14	0.24*	-0.08	-0.10	0.02	-				
7	Negative Stat. Sig. Effect	25	30.5%	-0.07	-0.24*	-0.17	0.08	0.02	-0.34**	-			
8	KLD/MSCI Used	38	46.3%	0.07	-0.01	-0.15	0.08	0.47**	0.01	0.04	-		
9	Refinitiv Used	14	17.0%	0.04	-0.07	0.02	-0.16	-0.21	-0.18	0.12	-0.44**	-	
10	Bloomberg Used	9	11.0%	-0.02	0.13	-0.09	-0.14	-0.10	0.03	-0.23*	-0.34**	-0.16	-

⁷ * denotes a p-value of 0.05; ** denotes a p-value of 0.01.

⁸ Other third-party providers were excluded from this table, as their sample sizes were insufficient to support statistically meaningful conclusions in this analysis.

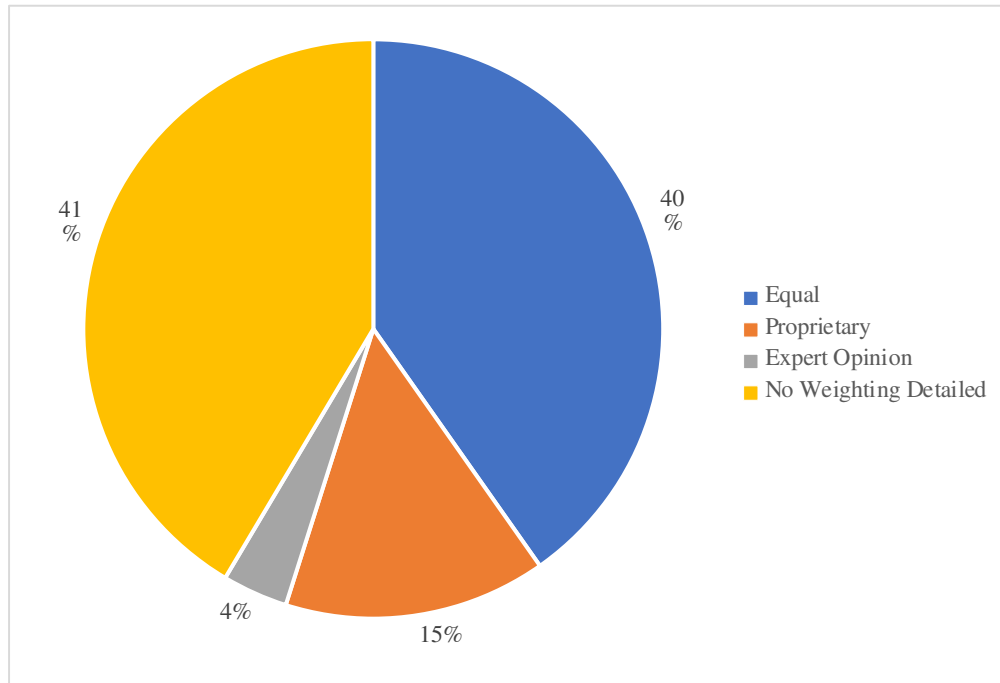
Figure 1. Sustainability Measure Providers Used (N = 89)^{9,10}



⁹ Total sample number is greater than the number of articles reviewed due to some articles using more than one third-party measure provider.

¹⁰ “Other” includes the following measure providers used by only 1 article: Trucost, Oekom, Ranking’s, SAM, Southern Weekend, Korean Corporate Governance Service, Choice’s, CSRHub, CNRDS, LSEG

Figure 2. Weighting Schemes Identified (N=82)

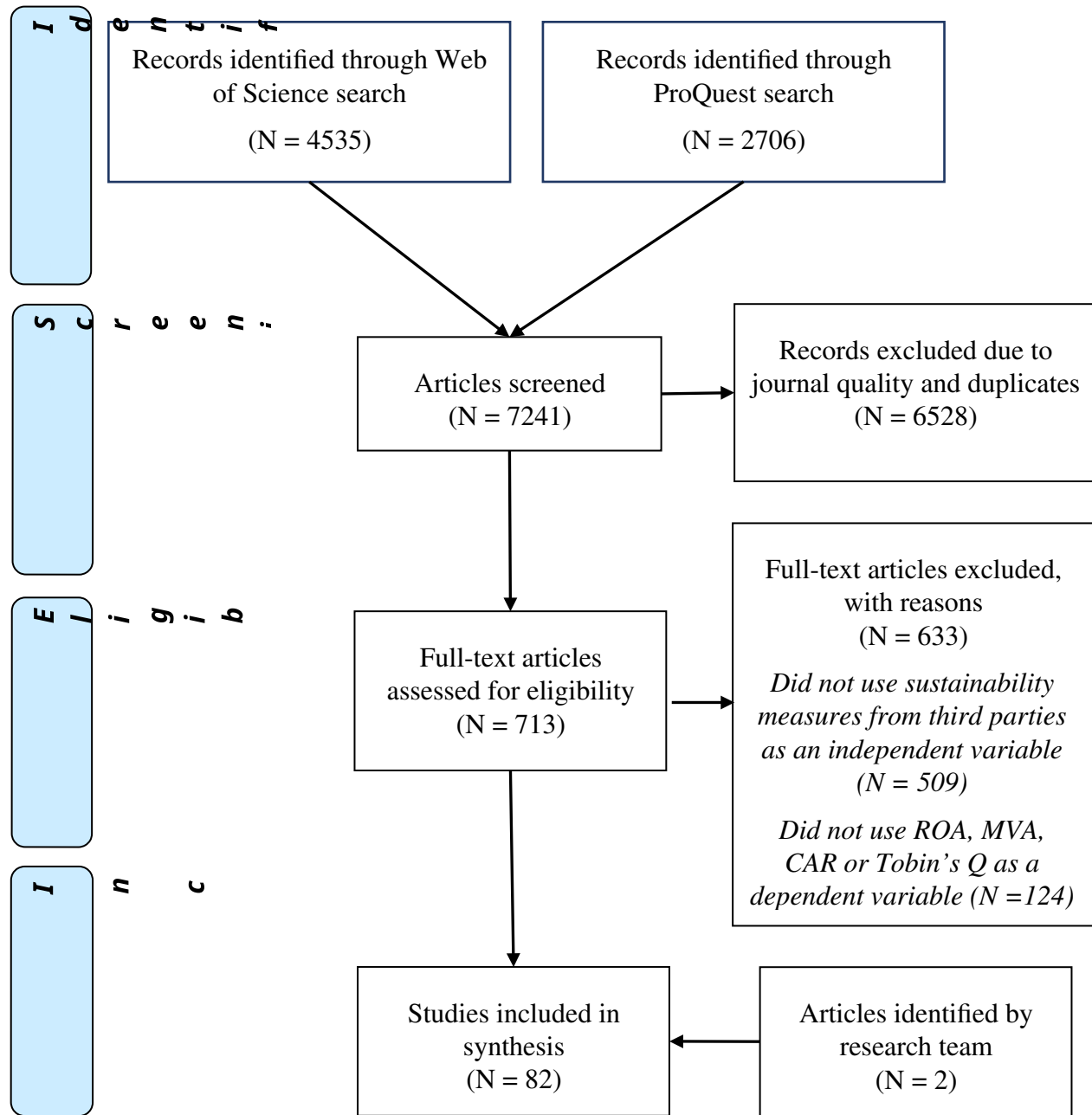


Appendices

Appendix 1. Keyword Search String

AB = ((Corporate Social Performance OR Corporate Environmental Performance OR Corporate Social Responsibility OR ESG OR Social Issue) AND (Financial Performance OR Shareholder Value)) AND WC = (Business OR Management OR Business Finance OR Operations Research Management Science OR Economics OR Environmental Studies OR Environmental Sciences OR Green Sustainable Science Technology) AND PY = (1995-2024)

Appendix 2. PRISMA Flow Chart



Appendix 3. List of Articles Included

- Afrin, R., Peng, N., & Bowen, F. (2022). The wealth effect of corporate water actions: How past corporate responsibility and irresponsibility influence stock market reactions. *Journal of Business Ethics*, 180(1), 105–124. <https://doi.org/10.1007/s10551-021-04881-6>
- Apaydin, M., Jiang, G. F., Demirbag, M., & Jamali, D. (2021). The importance of corporate social responsibility strategic fit and times of economic hardship. *British Journal of Management*, 32(2), 399–415. <https://doi.org/10.1111/1467-8551.12402>
- Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. *Strategic Management Journal*, 41(6), 965–987. <https://doi.org/10.1002/smj.3122>
- Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33(11), 1304–1320. <https://doi.org/10.1002/smj.1980>
- Ben Lahouel, B., Ben Zaied, Y., Managi, S., & Taleb, L. (2022). Re-thinking about U: The relevance of regime-switching model in the relationship between environmental corporate social responsibility and financial performance. *Journal of Business Research*, 140, 498–519. <https://doi.org/10.1016/j.jbusres.2021.11.019>
- Blanco, B., Guillamón-Saorín, E., & Guiral, A. (2013). Do non-socially responsible companies achieve legitimacy through socially responsible actions? The mediating effect of innovation. *Journal of Business Ethics*, 117(1), 67–83. <https://doi.org/10.1007/s10551-012-1503-3>
- Brinette, S., Sonmez, F. D., & Tournus, P. S. (2023). ESG controversies and firm value: Moderating role of board gender diversity and board independence. *IEEE Transactions on Engineering Management*. Advance online publication. <https://doi.org/10.1109/TEM.2023.3236667>
- Brower, J., & Dacin, P. A. (2020). An institutional theory approach to the evolution of the corporate social performance - corporate financial performance relationship. *Journal of Management Studies*, 57(4), 805–836. <https://doi.org/10.1111/joms.12550>
- Brower, J., Kashmiri, S., & Mahajan, V. (2017). Signaling virtue: Does firm corporate social performance trajectory moderate the social performance-financial performance relationship? *Journal of Business Research*, 81, 86–95. <https://doi.org/10.1016/j.jbusres.2017.08.013>
- Busch, T., Bassen, A., Lewandowski, S., & Sump, F. (2022). Corporate carbon and financial performance revisited. *Organization & Environment*, 35(1), 154–171.
- Candio, P. (2024). The effect of ESG and CSR attitude on financial performance in Europe: A quantitative re-examination. *Journal of Environmental Management*, 354, Article 120390. <https://doi.org/10.1016/j.jenvman.2024.120390>
- Cao, S. P., Yao, H. J., & Zhang, M. (2023). CSR gap and firm performance: An organizational justice perspective. *Journal of Business Research*, 158, Article 113692. <https://doi.org/10.1016/j.jbusres.2023.113692>
- Chang, Y. K., Oh, W. Y., & Messersmith, J. G. (2013). Translating corporate social performance into financial performance: Exploring the moderating role of high-performance work practices. *The*

- International Journal of Human Resource Management*, 24(19), 3738–3756.
<https://doi.org/10.1080/09585192.2013.778312>
- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of Environmental Management*, 345, Article 118829. <https://doi.org/10.1016/j.jenvman.2023.118829>
- Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture? *Academy of Management Perspectives*, 27(3), 255–267. <https://doi.org/10.5465/amp.2012.0123>
- Delmas, M. A., Nairn-Birch, N., & Lim, J. (2015). Dynamics of environmental and financial performance: The case of greenhouse gas emissions. *Organization & Environment*, 28(4), 374–393.
- Deng, X., Kang, J. K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, 110(1), 87–109. <https://doi.org/10.1016/j.jfineco.2013.04.014>
- Duque-Grisales, E., & Aguilera-Caracuel, J. (2021). Environmental, social and governance (ESG) scores and financial performance of multilatinas: Moderating effects of geographic international diversification and financial slack. *Journal of Business Ethics*, 168(2), 315–334. <https://doi.org/10.1007/s10551-019-04177-w>
- Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56(3), 758–781. <https://doi.org/10.5465/amj.2011.0744>
- Gangwani, M., & Kashiramka, S. (2024). Does ESG performance impact value and risk-taking by commercial banks? Evidence from emerging market economies. *Business Strategy and the Environment*, 33(7), 7562–7589. <https://doi.org/10.1002/bse.3882>
- Gao, J. J., & Bansal, P. (2013). Instrumental and integrative logics in business sustainability. *Journal of Business Ethics*, 112(2), 241–255. <https://doi.org/10.1007/s10551-012-1245-2>
- Garcia, A. S., & Orsato, R. J. (2020). Testing the institutional difference hypothesis: A study about environmental, social, governance, and financial performance. *Business Strategy and the Environment*, 29(8), 3261–3272. <https://doi.org/10.1002/bse.2570>
- García-Sánchez, I. M., & Martínez-Ferrero, J. (2019). Chief executive officer ability, corporate social responsibility, and financial performance: The moderating role of the environment. *Business Strategy and the Environment*, 28(4), 542–555. <https://doi.org/10.1002/bse.2263>
- Godfrey, P. C., Merrill, C. B., & Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal*, 30(4), 425–445. <https://doi.org/10.1002/smj.750>
- Hannah, S. T., Sayari, N., Harris, F. H. D., & Cain, C. L. (2021). The direct and moderating effects of endogenous corporate social responsibility on firm valuation: Theoretical and empirical evidence from the global financial crisis. *Journal of Management Studies*, 58(2), 421–456. <https://doi.org/10.1111/joms.12586>
- Hasan, I., Kobeissi, N., Liu, L. L., & Wang, H. Z. (2018). Corporate social responsibility and firm financial performance: The mediating role of productivity. *Journal of Business Ethics*, 149(3), 671–688. <https://doi.org/10.1007/s10551-016-3066-1>

- He, C., Jia, F., Wang, L., Chen, L., & Fernandes, K. (2023). The impact of corporate social responsibility decoupling on financial performance: The role of customer structure and operational slack. *International Journal of Operations & Production Management*, 43(12), 1859–1890. <https://doi.org/10.1108/IJOPM-08-2022-0521>
- Hillman, A. J., & Keim, G. D. (2001). Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal*, 22(2), 125–139.
- Hoang, T. H. V., Przychodzen, W., Przychodzen, J., & Segbotangni, E. A. (2020). Does it pay to be green? A disaggregated analysis of US firms with green patents. *Business Strategy and the Environment*, 29(3), 1331–1361. <https://doi.org/10.1002/bse.2437>
- Hossain, M. M., Wang, L. F., & Yu, J. (2024). The reputational costs of corporate environmental underperformance: Evidence from China. *Business Strategy and the Environment*, 33(2), 930–948. <https://doi.org/10.1002/bse.3528>
- Hsu, A., Koh, K., Liu, S., & Tong, Y. H. (2019). Corporate social responsibility and corporate disclosures: An investigation of investors' and analysts' perceptions. *Journal of Business Ethics*, 158(2), 507–534. <https://doi.org/10.1007/s10551-017-3767-0>
- Hull, C. E., & Rothenberg, S. (2008). Firm performance: The interactions of corporate social performance with innovation and industry differentiation. *Strategic Management Journal*, 29(7), 781–789. <https://doi.org/10.1002/smj.675>
- Hyun, S., Kim, J. M., & Liu, Y. (2023). Equal gains and pains? Analyzing corporate financial performance for industrial corporate social performance leaders and laggards. *Journal of Business Research*, 155, Article 113414. <https://doi.org/10.1016/j.jbusres.2022.113414>
- Ibishova, B., Misund, B., & Tveterås, R. (2024). Driving green: Financial benefits of carbon emission reduction in companies. *International Review of Financial Analysis*, 96, Article 103757. <https://doi.org/10.1016/j.irfa.2024.103757>
- Inoue, Y., & Lee, S. (2011). Effects of different dimensions of corporate social responsibility on corporate financial performance in tourism-related industries. *Tourism Management*, 32(4), 790–804. <https://doi.org/10.1016/j.tourman.2010.06.019>
- Iurkov, V., Koval, M., Misra, S., Pedada, K., & Sinha, A. (2024). Impact of ESG distinctiveness in alliances on shareholder value. *Journal of Business Research*, 171, Article 114395. <https://doi.org/10.1016/j.jbusres.2023.114395>
- Janney, J. J., & Gove, S. (2011). Reputation and corporate social responsibility aberrations, trends, and hypocrisy: Reactions to firm choices in the stock option backdating scandal. *Journal of Management Studies*, 48(7), 1562–1585. <https://doi.org/10.1111/j.1467-6486.2010.00984.x>
- Jell-Ojobor, M., & Raha, A. (2022). Being good at being good—The mediating role of an environmental management system in value-creating green supply chain management practices. *Business Strategy and the Environment*, 31(5), 1964–1984. <https://doi.org/10.1002/bse.2993>
- Jeriji, M., Louhichi, W., & Ftiti, Z. (2023). Migrating to Global Reporting Initiative guidelines: Does international harmonization of CSR information pay? *British Journal of Management*, 34(2), 555–575. <https://doi.org/10.1111/1467-8551.12603>
- Kashmiri, S., Nicol, C. D., & Hsu, L. (2017). Birds of a feather: Intra-industry spillover of the Target customer data breach and the shielding role of IT, marketing, and CSR. *Journal of the Academy of Marketing Science*, 45(2), 208–228. <https://doi.org/10.1007/s11747-016-0486-5>

- Kim, K. H., Kim, M., & Qian, C. L. (2018). Effects of corporate social responsibility on corporate financial performance: A competitive-action perspective. *Journal of Management*, 44(3), 1097–1118. <https://doi.org/10.1177/0149206315602530>
- Kumar, A., Gupta, J., & Das, N. (2022). Revisiting the influence of corporate sustainability practices on corporate financial performance: Evidence from the global energy sector. *Business Strategy and the Environment*, 31(7), 3231–3253. <https://doi.org/10.1002/bse.3073>
- Lee, J., & Kwon, H. B. (2019). The synergistic effect of environmental sustainability and corporate reputation on market value added (MVA) in manufacturing firms. *International Journal of Production Research*, 57(22), 7123–7141. <https://doi.org/10.1080/00207543.2019.1578430>
- Lee, K. H., Cin, B. C., & Lee, E. Y. (2016). Environmental responsibility and firm performance: The application of an environmental, social and governance model. *Business Strategy and the Environment*, 25(1), 40–53. <https://doi.org/10.1002/bse.1855>
- Lee, S., Kim, B., & Ham, S. (2018). Strategic CSR for airlines: Does materiality matter? *International Journal of Contemporary Hospitality Management*, 30(12), 3592–3608. <https://doi.org/10.1108/IJCHM-10-2017-0697>
- Liu, A. Z., Liu, A. X., Wang, R., & Xu, S. X. (2020). Too much of a good thing? The boomerang effect of firms' investments on corporate social responsibility during product recalls. *Journal of Management Studies*, 57(8), 1437–1472. <https://doi.org/10.1111/joms.12525>
- Lys, T., Naughton, J. P., & Wang, C. (2015). Signaling through corporate accountability reporting. *Journal of Accounting and Economics*, 60(1), 56–72. <https://doi.org/10.1016/j.jacceco.2015.03.001>
- McGuinness, P. B., Vieito, J. P., & Wang, M. (2020). Proactive government intervention, board gender balance, and stakeholder engagement in China and Europe. *Asia Pacific Journal of Management*, 37(3), 719–762. <https://doi.org/10.1007/s10490-018-9611-y>
- McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5), 603–609. [https://doi.org/10.1002/\(SICI\)1097-0266\(200005\)21:5<603::AID-SMJ101>3.3.CO;2-V](https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.3.CO;2-V)
- Meier, O., Naccache, P., & Schier, G. (2021). Exploring the curvature of the relationship between HRM-CSR and corporate financial performance. *Journal of Business Ethics*, 170(4), 857–873. <https://doi.org/10.1007/s10551-019-04332-3>
- Minutolo, M. C., Kristjanpoller, W. D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. *Business Strategy and the Environment*, 28(6), 1083–1095. <https://doi.org/10.1002/bse.2303>
- Moneva, J. M., Bonilla-Priego, M. J., & Ortas, E. (2020). Corporate social responsibility and organisational performance in the tourism sector. *Journal of Sustainable Tourism*, 28(6), 853–872. <https://doi.org/10.1080/09669582.2019.1707838>
- Nakao, Y., Amano, A., Matsumura, K., Genba, K., & Nakano, M. (2007). Relationship between environmental performance and financial performance: An empirical analysis of Japanese corporations. *Business Strategy and the Environment*, 16(2), 106–118. <https://doi.org/10.1002/bse.476>
- Nekhili, M., Boukadhaba, A., Nagati, H., & Chtioui, T. (2021). ESG performance and market value: The moderating role of employee board representation. *The International Journal of Human Resource Management*, 32(14), 3061–3087. <https://doi.org/10.1080/09585192.2019.1629989>

- Petitjean, M. (2019). Eco-friendly policies and financial performance: Was the financial crisis a game changer for large US companies? *Energy Economics*, 80, 502–511. <https://doi.org/10.1016/j.eneco.2019.01.028>
- Pichler, S., Blazovich, J. L., Cook, K. A., Huston, J. M., & Strawser, W. R. (2018). Do LGBT-supportive corporate policies enhance firm performance? *Human Resource Management*, 57(1), 263–278. <https://doi.org/10.1002/hrm.21831>
- Preston, L. E., & O'Bannon, D. P. (1997). The corporate social-financial performance relationship: A typology and analysis. *Business & Society*, 36(4), 419–429. <https://doi.org/10.1177/000765039703600406>
- Radu, C., & Smaili, N. (2021). Corporate performance patterns of Canadian listed firms: Balancing financial and corporate social responsibility outcomes. *Business Strategy and the Environment*, 30(7), 3344–3359. <https://doi.org/10.1002/bse.2806>
- Sadovnikova, A., & Pujari, A. (2017). The effect of green partnerships on firm value. *Journal of the Academy of Marketing Science*, 45(2), 251–267. <https://doi.org/10.1007/s11747-016-0490-9>
- Sandberg, H., Alnoor, A., & Tiberius, V. (2023). Environmental, social, and governance ratings and financial performance: Evidence from the European food industry. *Business Strategy and the Environment*, 32(4), 2471–2489. <https://doi.org/10.1002/bse.3259>
- Schreck, P. (2011). Reviewing the business case for corporate social responsibility: New evidence and analysis. *Journal of Business Ethics*, 103(2), 167–188. <https://doi.org/10.1007/s10551-011-0867-0>
- Shahzad, A. M., & Sharfman, M. P. (2017). Corporate social performance and financial performance: Sample-selection issues. *Business & Society*, 56(6), 889–918. <https://doi.org/10.1177/0007650315590399>
- Shin, H., Song, H. J., & Kang, K. H. (2024). The moderating effect of interest rates on the relationship between ESG and firm performance in the US restaurant industry. *Journal of Sustainable Tourism*, 1–18. <https://doi.org/10.1080/09669582.2024.2382427>
- Shin, J., Moon, J. J., & Kang, J. (2023). Where does ESG pay? The role of national culture in moderating the relationship between ESG performance and financial performance. *International Business Review*, 32(3), Article 102071. <https://doi.org/10.1016/j.ibusrev.2022.102071>
- Sroufe, R., & Gopalakrishna-Remani, V. (2019). Management, social sustainability, reputation, and financial performance relationships: An empirical examination of US firms. *Organization & Environment*, 32(3), 331–362. <https://doi.org/10.1177/1086026618756611>
- Surroca, J., Tribó, J. A., & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal*, 31(5), 463–490. <https://doi.org/10.1002/smj.820>
- Surroca, J. A., Aguilera, R. V., Desender, K., & Tribó, J. A. (2020). Is managerial entrenchment always bad and corporate social responsibility always good? A cross-national examination of their combined influence on shareholder value. *Strategic Management Journal*, 41(5), 891–920. <https://doi.org/10.1002/smj.3132>
- Taddeo, S., Agnese, P., & Busato, F. (2024). Rethinking the effect of ESG practices on profitability through cross-dimensional substitutability. *Journal of Environmental Management*, 352, Article 120115. <https://doi.org/10.1016/j.jenvman.2024.120115>

- Tang, Z., Hull, C. E., & Rothenberg, S. (2012). How corporate social responsibility engagement strategy moderates the CSR-financial performance relationship. *Journal of Management Studies*, 49(7), 1274–1303. <https://doi.org/10.1111/j.1467-6486.2012.01068.x>
- Theodoulidis, B., Diaz, D., Crotto, F., & Rancati, E. (2017). Exploring corporate social responsibility and financial performance through stakeholder theory in the tourism industries. *Tourism Management*, 62, 173–188. <https://doi.org/10.1016/j.tourman.2017.03.018>
- Van der Laan, G., Van Ees, H., & Van Witteloostuijn, A. (2008). Corporate social and financial performance: An extended stakeholder theory, and empirical test with accounting measures. *Journal of Business Ethics*, 79(3), 299–310. <https://doi.org/10.1007/s10551-007-9398-0>
- Waddock, S. A., & Graves, S. B. (1997). The corporate social performance - financial performance link. *Strategic Management Journal*, 18(4), 303–319. [https://doi.org/10.1002/\(SICI\)1097-0266\(199704\)18:4<303::AID-SMJ869>3.3.CO;2-7](https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.3.CO;2-7)
- Wang, H. L., & Choi, J. (2013). A new look at the corporate social-financial performance relationship: The moderating roles of temporal and interdomain consistency in corporate social performance. *Journal of Management*, 39(2), 416–441. <https://doi.org/10.1177/0149206310375850>
- Wang, M. B., Qiu, C., & Kong, D. M. (2011). Corporate social responsibility, investor behaviors, and stock market returns: Evidence from a natural experiment in China. *Journal of Business Ethics*, 101(1), 127–141. <https://doi.org/10.1007/s10551-010-0713-9>
- Woodroof, P. J., Deitz, G. D., Howie, K. M., & Evans, R. D. (2019). The effect of cause-related marketing on firm value: A look at Fortune's most admired all-stars. *Journal of the Academy of Marketing Science*, 47(5), 899–918. <https://doi.org/10.1007/s11747-019-00660-y>
- Wu, M. W., & Shen, C. H. (2013). Corporate social responsibility in the banking industry: Motives and financial performance. *Journal of Banking & Finance*, 37(9), 3529–3547. <https://doi.org/10.1016/j.jbankfin.2013.04.023>
- Xia, L., Li, Z., Wei, J., & Gao, S. (2024). Doing well by doing good: Unpacking the black box of corporate social responsibility. *Asia Pacific Journal of Management*, 41(3), 1601–1631. <https://doi.org/10.1007/s10490-023-09878-5>
- Xu, J., Wei, J. C., & Lu, L. D. (2019). Strategic stakeholder management, environmental corporate social responsibility engagement, and financial performance of stigmatized firms derived from Chinese special environmental policy. *Business Strategy and the Environment*, 28(6), 1027–1044. <https://doi.org/10.1002/bse.2299>
- Yadav, P. L., Han, S. H., & Kim, H. (2017). Sustaining competitive advantage through corporate environmental performance. *Business Strategy and the Environment*, 26(3), 345–357. <https://doi.org/10.1002/bse.1921>
- Yu, E. P. Y., Guo, C. Q., & Luu, B. V. (2018). Environmental, social and governance transparency and firm value. *Business Strategy and the Environment*, 27(7), 987–1004. <https://doi.org/10.1002/bse.2047>
- Zhao, X. P., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, 37(11), 2378–2388. <https://doi.org/10.1002/smj.2579>

Zhao, X. P., & Murrell, A. (2022). Does a virtuous circle really exist? Revisiting the causal linkage between CSP and CFP. *Journal of Business Ethics*, 177(1), 173–192.
<https://doi.org/10.1007/s10551-021-04769-5>

Appendix 4. Codebook

ESG Systematic Literature Review Codebook - Version 20240422

The following describes the data points to be collected from each article. The format is as follows.

[Variable Name], column header on the coding matrix. **[Coding Type]** as a 0/1 binary to indicate “no” coded as a “0” or “yes” coded as a “1” or as a text response to be copied verbatim from the text of the article, or a variation thereof. **[Description]** as a narrative description of how the variable should be identified from the text and, if applicable, instructions on what to do if data for the variable is not present in the article.

Criteria for Selecting Firms; Copied Text – Looking throughout the article, copy the criteria used to select the firms in scope for the article’s analysis. This can include geographical focuses (e.g., US Markets), index or fund focuses (e.g., S&P500 listed firms, MSCI 400 Firms), sector focuses (e.g., GCIS Information Technology firms), and firms included in the ESG data source’s scoring (e.g., only firms with KLD scores for years 2010-2020).

Start of Years Covered; Copied Text – Looking through the methods section of the article, copy the year in which the analysis is started. E.g., if the analysis looks at a firm’s performance between 2000 and 2020, then “2000” should be copied and pasted.

End of Years Covered; Copied Text – Looking through the methods section of the article, copy the year in which the analysis ends. E.g., if the analysis looks at a firm’s performance between 2000 and 2020, then “2020” should be copied and pasted.

ESG Data Provider 1; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the ESG data provider as printed in the article.

ESG Data Provider 2; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the second ESG data provider as printed in the article. If a second ESG data source is not used, enter “NA.”

ESG Data Provider 3; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the third ESG data provider as printed in the article. If a third ESG data source is not used, enter “NA.”

ESG Data Provider 4; Copied Text – Looking in the methods or data sources section of the article, copy out the name of the fourth ESG data provider as printed in the article. If a fourth ESG data source is not used, enter “NA.”

ESG Data Used Includes Environmental Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include aspects of environmental performance, impacts, or strategies? If so, enter “1.”

ESG Data Used Includes Social Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include aspects of social performance, impacts, or strategies? If so, enter “1.”

ESG Data Used Includes Governance Aspects; 0/1 – Looking at the scope of the ESG data used in the article, does it include governance performance, impacts, or strategies? If so, enter “1.”

ROA Used as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use return on assets as a dependent variable in the statistical analysis?

Tobin's Q as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use Tobin's Q as a dependent variable in the statistical analysis?

CAR as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use cumulative abnormal return as a dependent variable in the statistical analysis?

MVA as DV; 0/1 – Look in the methods or data sources section of the article. Does the article use market value added as a dependent variable in the statistical analysis?

Type of Statistical Analyses; Copied Text – Look in the methods or data sources section of the article, copy out the types of statistical techniques used in the analyses of the article. Include these as a list separated by a semicolon.

Scale of ESG Score; Copied Text – As defined by the article, list the theoretical minimum and maximum values for the ESG score used (e.g., 0 – 100; -10 to +10). If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0 – 100; -10 to +10). Note that theoretical scores are different from the minimum and maximum scores observed.

Reported Range of ESG Score; Copied Text – As reported by the article, list the minimum and maximum ESG scores observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 55 – 98; -1 to +9).

Reported Mean of ESG Score; Copied Text – As reported by the article, list the mean ESG score observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the values separated by a semicolon (e.g., 65; 2).

Reported Median of ESG Score; Copied Text – As reported by the article, list the median ESG score observed by the article. If none is provided, enter “Not Provided.” If more than one ESG score provider is used, list the values separated by a semicolon (e.g., 65; 2).

IV DV Lag Used; 0/1 – Does the article describe using a time lag between the observation of the ESG score (as the IV) and the measures of financial performance (as the DV)?

Direction of Effect; Positive/Negative – As reported by the article, is the direction of the effect between the ESG score (as the IV) and the measures of financial performance (as the DV) negative or positive?

Magnitude of Effect; Copied Text – As reported by the article, list the magnitude of the effect(s) between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 6.5; 0.2).

Std Error of Effect; Copied Text – As reported by the article, list the standard error of the coefficient(s) between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0.065; 0.002).

Significance Level of Effect; Copied Text – As reported by the article, list the p-value(s) reported between the ESG score (as the IV) and the measures of financial performance (as the DV). If more than one ESG score provider is used, list the ranges separated by a semicolon (e.g., 0.0065; 0.002).

Author Conflict of Interest ID's; 0/1 – As reported by the article, is a conflict of interest identified between the research into corporate sustainability or ESG data providers and the

authors of the article? This could include providing consulting services to a firm or firms covered, or serving as an investor in or advisor to ESG data providers.

Accuracy as Justification; 0/1 – Looking throughout the entire article, does the article discuss the accuracy of the ESG score used as a justification in its selection? A discussion of accuracy will include some aspects of the quantitative nature of the ESG score or the use of standardized methods of measurement.

Accuracy as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the accuracy of the ESG score used as a limitation of using said score? A discussion of accuracy will include some aspects of the quantitative nature of the ESG score or the use of standardized methods of measurement.

Reliability as Justification; 0/1 – Looking throughout the entire article, does the article discuss the reliability of the ESG score used as a justification in its selection? A discussion of reliability will include aspects of assurance for the measurements used to create the ESG score, as well as for other research used to verify the ESG score's methods.

Reliability as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the reliability of the ESG score used as a limitation of using said score? A discussion of reliability will include aspects of assurance for the measurements used to create the ESG score, as well as for other research used to verify the ESG score's methods.

Timeliness as Justification; 0/1 – Looking throughout the entire article, does the article discuss the timeliness of the ESG score used as a justification in its selection? A discussion of timeliness will include aspects such as how often the score/data is updated, the “look-back” period the data used to create the ESG score represents, and whether the “year” associated with the ESG score reflects the year of the data or the year the score was issued.

Timeliness as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the timeliness of the ESG score used as a limitation of using said score? A discussion of timeliness will include aspects such as how often the score/data is updated, the “look-back” period the data used to create the ESG score represents, and whether the “year” associated with the ESG score reflects the year of the data or the year the score was issued.

Fungibility as Justification; 0/1 – Looking throughout the entire article, does the article discuss the fungibility decisions made in creating the ESG score used as a justification in its selection? A discussion of fungibility will include some aspect of what subcomponent measures are being aggregated (i.e., treated as fungible) or decidedly being kept unaggregated (i.e., treated as non-fungible).

Fungibility as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the fungibility decisions made in the ESG score used as a limitation of using said score? A discussion of fungibility will include some aspect of what subcomponent measures are being aggregated (i.e., treated as fungible) or decidedly being kept unaggregated (i.e., treated as non-fungible).

Normalization as Justification; 0/1 – Looking throughout the entire article, does the article discuss the normalization methods used in creating the ESG score used as a justification in its selection? A discussion of normalization methods will include some aspects of compensating for differences between companies. This could be for attributes like company size or industry/sector.

Normalization as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the normalization methods used in creating the ESG score as a limitation of using said

score? A discussion of normalization methods will include some aspects of compensating for differences between companies. This could be for attributes like company size or industry/sector.

Weighting as Justification; 0/1 – Looking throughout the entire article, does the article discuss the weighting method used in creating the ESG score as a justification in its selection? A discussion of weighting methods will include statements of how the subcomponents are balanced against each other. This can include a statement of everything being “equally” weighted.

Weighting as Limitation; 0/1 – Looking throughout the entire article, does the article discuss the weighting methods used in creating the ESG score as a limitation of using said score? A discussion of weighting methods will include statements of how the subcomponents are balanced against each other. This can include a statement of everything being “equally” weighted.

****General Notes****

For the discussion of justifications and limitations, do not collect general statements that refer to all aggregate measures of ESG. Only capture those specifically called out in the article being coded regarding the measures used. However, if the article states something generally about all ESG scores and indicates that it applies to the specific ESG source used in the article, that would constitute a specific statement and should be captured.

Justifications will typically be discussed in the introduction, background/review literature, and methods/data sections of an article. At times, it may be discussed in the article’s conclusions. Limitations will typically be discussed near the end of the article in the limitations or conclusions sections. At times, a discussion of the limitations can be in the introduction, background literature, or methods sections of an article.

Limitations can include acknowledging “unknowns” about the scores. This can include stating that proprietary methods are used to create the aggregate score from its constituent components.