Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Finanziato nell'ambito del Piano Nazionale di Ripresa e Resilienza PNRR. Missione 4, Componente 2, Investimento 1.3 Creazione di "Partenariati estesi alle università, ai centri di ricerca, alle aziende per il finanziamento di progetti di ricerca di base"

GRINS FOUNDATION

A COMPARATIVE ANALYSIS OF ESG QUESTIONNAIRES

Gennaio 2025 - Pubblicazione finanziata dal PNRR, Missione 4 (Infrastruttura e Ricerca), Componente 2 (Dalla Ricerca all'Impresa), Investimento 1.3 (Partenariati Estesi), Tematica 9 (Sostenibilità economica di sistemi e territori). Sito web: https://grins.it/

Document data	
Title	Work Package 4 Deliverable 3 **A comparative analysis of ESG questionnaires**
Owner	University of Turin
Contributor/s	Laura Corazza Eleonora Sommariva Francesco Marengo Elisa Giacosa
Document version	D3 Final
Last version date	24/11/2025

Gennaio 2025 - Pubblicazione finanziata dal PNRR, Missione 4 (Infrastruttura e Ricerca), Componente 2 (Dalla Ricerca all'Impresa), Investimento 1.3 (Partnership Estese), Tematica 9 (Sostenibilità economica di sistemi e territori). Sito web: https://grins.it/

A comparative analysis of ESG questionnaires

Laura CORAZZA Eleonora SOMMARIVA Francesco MARENGO Elisa GIACOSA

University of Turin

November 2025

This study systematically analyses the questionnaires used by rating agencies, financial institutions and companies to collect ESG data, highlighting significant heterogeneity in the structure, language and content of the tools currently in use. Through a three-level approach — descriptive analysis, measurement of textual similarity using cosine similarity and cluster analysis — the research shows how differences in data collection formats compromise the comparability and transparency of sustainability assessments. The empirical results, based on a diverse sample of questionnaires, reveal low levels of linguistic similarity, wide variations in the number and type of questions, and structural fragmentation reflected in six distinct clusters. This evidence is particularly relevant in light of the new Regulation (EU) 2024/3005 on ESG ratings, which aims to strengthen their integrity through more stringent requirements in terms of processes, transparency, and data quality. The study suggests that the effectiveness of the European regulatory framework could be further enhanced by extending harmonisation efforts to the information gathering phase, promoting minimum standards and greater methodological consistency. The analysis thus contributes to the debate on the quality of ESG metrics, providing useful evidence for improving the information base on which sustainability assessments in the European market are founded.

Keywords: ESG ratings; data collection; questionnaire heterogeneity; comparative analysis; non-financial reporting.

Gennaio 2025 - Pubblicazione finanziata dal PNRR, Missione 4 (Infrastruttura e Ricerca), Componente 2 (Dalla Ricerca all'Impresa), Investimento 1.3 (Partnership Estese), Tematica 9 (Sostenibilità economica di sistemi e territori). Sito web: https://grins.it/

1. Introduction

In recent years, environmental, social and governance (ESG) issues have taken on an increasingly central role in corporate strategies and investment decisions (Menicucci and Paolucci, 2024). Investors and stakeholders are paying increasing attention to the standards used to measure the social and environmental impact of companies (Bronzini et al., 2024; Soares, 2024), prompting many organisations to integrate sustainability practices into their policies and communications (Rao et al., 2025).

Reference regulation

On the regulatory front, the European Union has progressively tightened the framework governing non-financial disclosure. This evolution began with Directive 2014/95/EU and Legislative Decree 254/2016 (Carnini Pulino et al., 2022), and continued with the introduction of the Corporate Sustainability Reporting Directive (CSRD) in 2023 and Regulation (EU) 2024/3005, which entered into force on 1 January 2025. A key innovation of this latest regulation is the move toward standardising ESG ratings, with the explicit aim of enhancing their transparency, reliability and cross-company comparability.

Regulation (EU) 2024/3005 represents an attempt by the European Union to create a common regulatory framework to ensure transparency, integrity and quality in ESG rating activities. Its introduction responds to the need to direct financial flows towards sustainable investments, in line with the Sustainable Development Goals of the 2030 Agenda and the European Green Deal, which aims for climate neutrality by 2050. ESG ratings, in fact, play a crucial role in supporting investors, lenders, and insurance companies in their assessments of sustainability risks and opportunities.

The Regulation intervenes to address several critical issues that characterized the ESG rating market: the lack of transparency of the methodologies used, the presence of potential conflicts of interest, and, above all, the lack of common rules between Member States, which risked generating information asymmetries and distortions in investment decisions. To address these problems, the European legislator has introduced stringent process requirements: suppliers must adopt rigorous, systematic, independent and regularly updated methodologies; they must clearly and publicly communicate the models and assumptions underlying ratings; they must provide separate ratings for the three dimensions of Environment, Social and Governance or, in the case of aggregate ratings, make known the weightings attributed to the different pillars; furthermore, must ensure that the data used comes from reliable sources and is of adequate quality.

The Regulation covers all ESG rating providers operating in the EU, including those based in third countries when their ratings are intended for European entities, such as financial undertakings, companies subject to the reporting requirements of the Budgets Directive and the Transparency Directive, as well as Union public institutions and authorities. However, ratings produced for internal uses, private ones not intended for public dissemination and external verifications relating to European green bonds when not configured as ESG ratings are excluded.

In order to be able to operate in the European market, EU-based suppliers must obtain formal authorisation from ESMA, which also exercises ongoing supervision. Non-European suppliers can access the market through three channels: equivalence of their regulatory framework, endorsement by an already authorised EU supplier, or a recognition regime designed for small operators.

While not imposing a single formula or uniform rating model, the Regulation introduces a standardisation of processes, transparency and quality guarantees, reducing uncertainty and

inconsistencies that hitherto hampered the internal market. However, this is a framework that is still under development and implementation and is set to evolve in the coming years through technical guidelines, methodological updates and potential further regulatory interventions. In this sense, the Regulation provides a harmonized framework that raises the standards of reliability and comparability of ratings, while leaving room for methodological innovation by different operators.

Literature review

Companies have modified their governance systems by introducing sustainability-related incentives in their compensation plans and setting up dedicated committees (Buchetti et al., 2025). However, the expansion of the ESG ratings market has led to the proliferation of rating agencies operating with proprietary models, translating sustainability into synthetic scores that are often not comparable (Dorfleitner et al., 2015; Berg et al., 2022). Because rating agencies rely on heterogeneous methodologies and draw on predominantly qualitative information, their assessments often diverge substantially (Serafeim and Yoon, 2023), resulting in notable inconsistencies. This fragmentation is further exacerbated by the absence of standardised rating formats: while some agencies express ESG performance on a numerical scale from 0 to 100, others adopt alphabetic classifications (Anselmi et al., 2022). Such variability in scoring frameworks reduces the comparability of ratings assigned to the same firm, ultimately hindering an objective evaluation of its performance (Kim and Koo, 2023).

The literature identifies three main areas of divergence between agencies: scope, what is measured; measurement, relating to data collection and standardisation methods; and weighting of factors (Ferro et al., 2025). Measurement divergence is the main source of disagreement, as each agency adopts different protocols and incorporates its own interpretative bias in the definition of ESG variables, the so-called rater effect (Louche et al., 2023; Berg et al., 2022; Benuzzi et al., 2025).

The consequences of this fragmentation are clear: the average correlation between the ESG ratings of different agencies is significantly lower than that of credit ratings (Lopez et al., 2020), with values ranging from 0.30 to 0.54 depending on the study (Liang and Renneboog, 2020; Del Giudice et al., 2024; Berg et al., 2022). Billio et al. (2020) also note that the average disagreement between agencies exceeds one rating class (MAE = 1.32), pointing to a lack of common metrics in the definition of sustainability.

The heterogeneity of ratings stems from several factors: the lack of shared standards and a single definition of sustainability (Billio et al., 2020), poor transparency in assessment processes (Escrig-Olmedo et al., 2019) and differences in data weighting and aggregation criteria (Bronzini et al., 2024; González-Pozo et al., 2025). Furthermore, many systems tend to reward the amount of information disclosed rather than actual performance, with the risk of valuing communication more than substance (Cregan et al., 2025).

A prime example is the comparison between Refinitiv's methodology, based on percentile ranking, and the Performance Ratio (PR). While the former approach may distort actual performance differences due to relative normalisation, the PR is based solely on the minimum and maximum values of the sample, providing a more stable and representative measure of actual progress (Benuzzi et al., 2025).

In summary, the significant methodological heterogeneity that characterises ESG providers does not only concern what is measured or how factors are weighted, but above all how data is collected, interpreted and standardised (Louche et al., 2023; Berg et al., 2022). This jeopardises

investor confidence in ESG ratings (Mio et al., 2024), highlighting the need for greater methodological convergence and transparency in assessment processes (Soares, 2024; Lopez et al., 2020).

Research Approach

With this in mind, this study examines the linguistic, structural and content heterogeneity of ESG questionnaires used by rating agencies, public institutions, banks and companies, with the aim of highlighting the differences that characterise data collection tools in the initial phase of sustainability assessment processes. However, investigating these issues is very challenging, as the various actors involved in ESG processes almost never disclose the questionnaires or surveys used to collect ESG data (Assaf et al., 2024). The methodological approach adopted allows us to analyse the degree of consistency and divergence between questionnaires through three complementary levels of investigation.

Firstly, descriptive analysis provides an overview of the main structural characteristics of the questionnaires, considering aspects such as the number and type of questions, as well as providing an initial exploration of the language used. Subsequently, the linguistic component is explored in greater depth through cosine similarity analysis, which measures the lexical proximity between texts and assesses the degree of terminological affinity between the different instruments. Finally, through clustering analysis, the questionnaires are grouped into homogeneous sets based on common characteristics in order to highlight the main structural and semantic differences between the identified groups.

The results show significant heterogeneity among the questionnaires considered. The number of questions varies from a minimum of 7 to a maximum of 252, confirming the high variability in the complexity and depth of the information requested. From a linguistic point of view, the analysis also indicates a low similarity between the texts, with an average cosine similarity value of 0.141. Cluster analysis identifies six distinct groups of questionnaires, each characterised by specific combinations of key terms, thematic prevalence, average question length, response type and level of detail required.

Overall, the evidence gathered shows that the lack of uniformity in ESG questionnaires reflects the plurality of approaches and purposes adopted by the actors involved in the assessment processes. This fragmentation translates into considerable diversity in data collection methods, revealing the absence of shared standards and highlighting the need for greater methodological harmonisation. By highlighting the inconsistencies in the information acquisition phase, the research contributes to a broader debate on the quality, comparability and transparency of ESG rating systems. In doing so, it offers insights into how rating mechanisms can evolve towards practices that are more transparent, responsible and truly useful for corporate sustainability.

2. Methodology

This study takes an exploratory approach and aims to analyse the questionnaires used by various actors to collect the information needed to assess the environmental, social and governance (ESG) dimensions of organisations. These tools, despite their growing use in institutional and corporate contexts, have so far received limited attention in academia. One explanation for this also stems from the fact that these are tools internal to the organizations and are not publicly accessible (except at most for the few large rating agencies that publish the methodology).

The research is based on primary data, collected specifically for this work, with the aim of examining the degree of linguistic, structural and content heterogeneity of the ESG questionnaires used by rating agencies, public institutions, banks and companies. The inclusion of instruments from heterogeneous actors allows us to capture the variety of methodological approaches currently in use, which is evident both in external assessment processes for financial purposes and in internal practices for managing sustainability, supply chain and ESG risks.

Data collection

Data collection was conducted at national level in order to ensure greater linguistic, regulatory and institutional comparability between the questionnaires analysed. In a preliminary phase, an exploratory mapping of available sources was carried out by consulting official documentation, company websites and specialised platforms. This activity made it possible to identify an initial core of five freely accessible ESG questionnaires. In a second phase, with the aim of expanding the information base and including tools that are not publicly available, direct collection was initiated by contacting over 600 organisations in the CSR Piemonte project database, an initiative aimed at promoting corporate social responsibility among companies in the region.

The organisations involved, selected because they are considered socially responsible, were asked to share the ESG questionnaires used internally or received from third parties, such as rating agencies, banks or public institutions. The response rate, equal to 8.9%, made it possible to obtain a total sample of thirty-three questionnaires. Some tools were reported by more than one organisation, confirming their widespread use and representativeness on the national scene. The final sample is therefore consistent with the research objectives, as it includes questionnaires from a variety of actors with different purposes, reflecting the complexity and fragmentation of the current ESG assessment system.

A structured dataset was constructed from the questionnaires collected, in which each question was coded with the full text, the relevant subject area (environmental, social, governance or company profile), the length in terms of words and characters, the type of question (open, closed binary, closed multiple or Likert scale) and the unit of measurement associated with the answer. This systematisation phase made it possible to standardise the heterogeneous material collected and prepare it for subsequent statistical and textual analysis, ensuring the traceability and internal consistency of the data.

Descriptive analysis

The aim of this section is to provide a systematic representation of the main structural and linguistic characteristics of the questionnaires, laying the foundations for subsequent investigations into their semantic and content heterogeneity. This initial stage helped reveal the structural diversity among the questionnaires by systematically exploring and describing the characteristics of the sample (D'Este et al., 2025; Oyinlola, 2025).

Once the dataset containing all the information extracted from the ESG questionnaires had been constructed, an initial phase of analysis was carried out to summarise the characteristics of the sample and the main variables under study (Menicucci and Paolucci, 2024). In this preliminary phase, the unit of observation was represented by the individual questions in the questionnaires, on which descriptive and exploratory analyses were conducted. The frequency distributions of the questions were calculated according to the categorical variables identified and the summary measures relating to the average length of the questions, expressed in both characters and words.

This approach made it possible to observe how the formal complexity of the questionnaires varies according to their structure, revealing trends linked to the greater or lesser articulation of the ESG topics covered. At the same time, the relationships between the structural and thematic dimensions of the questionnaires were analysed, highlighting the differences in approach between more concise instruments and more extensive and detailed questionnaires.
In a subsequent phase, attention focused on the textual analysis of the entire corpus of questions, with the aim of exploring the linguistic composition of the questionnaires and identifying the main conceptual areas that characterise them. The text was subjected to linguistic cleaning procedures and natural language processing techniques, which made it possible to extract and quantify the most frequent words within the corpus. The exclusion of functional words and the subsequent analysis of frequencies made it possible to identify recurring key terms and construct a lexical profile of the questionnaires, highlighting significant differences in the terminology used in relation to the different ESG dimensions.
Finally, a structural profiling of the individual questionnaires was conducted, aimed at summarising the main quantitative and qualitative characteristics for each document. Summary indicators were calculated for each questionnaire, such as the total number of questions, the average length of the questions, the variety of response types and the internal distribution of thematic areas. This analysis made it possible to outline a specific profile for each questionnaire, highlighting the variety of methodological approaches and analytical depth adopted by the different actors. In addition, through targeted textual analysis, the most representative terms of each questionnaire were identified, providing a concise but informative overview of the linguistic peculiarities of the different assessment tools.

Cosine similarity

Textual similarity analysis was conducted to quantify the lexical proximity between the questionnaires and to identify any pairs of instruments characterised by similar wording (Oniani, 2020).
To conduct the analysis at an aggregate level, a dedicated dataset was constructed that included the full text of each questionnaire, the thematic distribution of the questions, the breakdown by question type, the average length of the questions in characters and words, the response detail index (measured through the number of response units), as well as information on the questionnaire's source and intended purpose.
After preliminary text cleaning and normalisation operations (removal of non-informative characters, management of missing values and spelling unification), a list of Italian stop words was defined and excluded from subsequent processing in order to focus the analysis on the key terms in the corpus. The full text of each questionnaire was then transformed into a numerical representation using the TF-IDF (Term Frequency–Inverse Document Frequency) technique (Guleria et al., 2025) technique, set to consider a limited vocabulary (to ensure parsimony and reduce noise) and oriented towards the use of unigrams, in order to capture the relative relevance of individual lemmas within the documents.
The cosine similarity between all documents was calculated on the TF-IDF matrix thus constructed: the result is a symmetric similarity matrix in which each cell measures the angular proximity between the vectors representing two questionnaires, and can therefore be interpreted as an index of normalised lexical overlap.

To summarise the results, descriptive statistics were calculated on the similarity coefficients between pairs: the mean, minimum and maximum (excluding the diagonal, i.e. the similarity of each document with itself) and the percentage distribution of pairs with similarities below or above the global mean. This operation allows us to evaluate not only cases of strong lexical convergence, but also the overall spread of terminological heterogeneity in the sample.
To complement the aggregate analysis, a procedure was implemented to compare TF-IDF vectors for selected pairs of questionnaires in order to inspect differences down to the level of individual words. In the example case of the questionnaires identified with IDs 23 and 32, the respective TF-IDF vectors were compared to verify whether they were practically identical; in the presence of disparities, the features (terms) with different weights were extracted and the TF-IDF values corresponding to each questionnaire were reported. This word-by-word comparison allows us to distinguish between overall similarities (largely overlapping vocabularies) and specific variations in the weighting of terms, providing clues about local adaptations of common text models or significant differences in the construction of questions.
Overall, the combination of a global analysis of the similarity matrix and detailed comparisons between individual pairs provides a robust assessment of the linguistic proximity of the questionnaires: the former offers an overview of the average lexical distance and its distributions, while the latter provides specific evidence on lexical divergences and convergences, which is useful for interpreting whether high similarities derive from a genuine overlap of content or from mere terminological coincidences.

Cluster analysis

The final stage of the methodology applied classification techniques to organise the questionnaires into groups that were internally as homogeneous and externally as distinct as possible, with the aim of uncovering patterns or underlying structures that illustrate the diversity present in their content (Bolliger et al., 2025).
The process was divided into three main stages: the preparation and transformation of variables, the construction of the integrated analysis matrix, and the application of different clustering techniques to validate the results.
During the first phase of variable preparation, the quantitative information relating to each questionnaire underwent a normalisation and standardisation procedure. In particular, the numerical variables included structural indicators (e.g., the total number of questions), measures of linguistic complexity (such as the average length of questions in words or characters), and internal proportions relating to ESG topics and question types. These variables were first cleaned and standardised by removing non-numerical symbols, homogenising the decimal format and imputing any missing values. Subsequently, the numerical variables underwent a standardisation process to ensure greater comparability and consistency within the dataset. This procedure, which transforms each variable into a distribution with a mean of zero and a standard deviation of one, eliminates distortions arising from heterogeneous scales or units of measurement. Standardisation is particularly important when using statistical and machine learning algorithms that are sensitive to scale differences, as it prevents variables with numerically higher values from disproportionately influencing the results.
At the same time, in order to quantitatively represent the textual content of the questionnaires, vectorisation was applied using the TF-IDF technique, widely used in text mining to assess the relevance of terms within a document in relation to the entire corpus analysed. This method

assigns each term a score based on its frequency in the individual document (TF) and its rarity in the overall corpus (IDF), allowing truly informative words to be distinguished from those that are very common and insignificant. The TF-IDF value, obtained as the product of the two components, thus allows the terms most representative of each questionnaire to be identified.
At the same time, the categorical variables, i.e. the purpose of the questionnaire (e.g. internal evaluation or external rating) and the origin of the actor who developed it, were transformed into numerical format using categorical coding procedures, in order to make them compatible with subsequent analyses.
The three components of the dataset — numerical, textual and categorical — were integrated into a single matrix that summarises the characteristics of each questionnaire. On this basis, the Elbow Method (Bonaventura Forleo and Bredice, 2025) was applied to identify the most appropriate number of clusters, observing how the sum of intra-cluster squares (WCSS) decreased as the number of groups increased. The inflection point of the curve indicated that a division into six clusters was the optimal configuration.
After establishing the composite structure, the clustering stage was carried out through two distinct unsupervised techniques. The first relied on the K-Means algorithm (Heiskanen and Ryynanen, 2024), which organised the questionnaires into six groups by optimising within-cluster similarity and maximising differentiation across clusters. The second approach employed a hierarchical procedure, constructing a dendrogram based on the same integrated matrix (Duong et al., 2025). Cluster formation followed Ward's linkage criterion, which reduces within-group variance and thus enhances the distinctiveness of the resulting clusters (Punj and Stewart, 1983). This method also suggested a six-cluster solution, enabling a direct and consistent comparison between the outcomes of the two clustering models.
The clusters generated by the K-Means and hierarchical models were then analysed at a structural and semantic level. In the first case, the averages of the main numerical variables were calculated for each group in order to identify recurring patterns, such as the presence of more concise or more extensive questionnaires, as well as differences in the wording of questions and the distribution of ESG topics. For the second analysis, the most representative words for each cluster were identified based on average TF-IDF scores.
Finally, the results produced by the two clustering models were subjected to a comparative evaluation using an internal validation procedure. This validation was conducted using three performance indicators — Silhouette Score, Calinski-Harabasz Index and Davies-Bouldin Score — which measure the degree of internal cohesion of the clusters and their mutual separation, thus providing an estimate of the overall reliability of the identified structure.
Overall, this procedure made it possible to identify coherent groups of ESG questionnaires, differentiated both in terms of structure and language, laying the foundations for an in-depth comparative analysis of the assessment practices adopted by the various actors.

3. Results

The methodological framework adopted in this study assesses the consistency of ESG questionnaires across three complementary dimensions. First, the descriptive analysis offers insight into their structural properties and provides a preliminary overview of their linguistic features. This linguistic examination is then deepened through cosine similarity, a metric that captures how closely the questionnaires align in terms of vocabulary and terminology. The final

component, cluster analysis enables a content-level assessment by identifying clusters of questionnaires that share common traits, thereby uncovering broader patterns of similarity within the dataset.

Descriptive analysis

Analysis of the questionnaires collected revealed a diverse and complex picture, reflecting the multiplicity of entities involved in the development and use of ESG assessment tools. The sources of the questionnaires show a balanced distribution among different types of actors: seven tools come from banks, seven from rating agencies, seven from public administration bodies, six from private companies that use them internally for the management or self-assessment of their practices, and a further six from research bodies or social organisations. This variety of origins provides a significant basis for observing how different purposes influence the formulation of questions and the overall structure of the questionnaires.

Overall, the corpus analysed comprises 33 unique questionnaires with a total of 1,839 questions, which represent the basic analytical unit of the first phase of the study. The resulting dataset, structured into nine main variables, includes information on the questionnaire ID, the text of the question, the ESG theme to which it belongs, the type and form of response expected, the unit of measurement of the response, and the length expressed in characters and words.

From a thematic perspective, there is a clear prevalence of environmental issues, which account for approximately 35% of the total, with 651 questions overall. This is followed by governance, with 520 questions, and social issues, with 497. The sections dedicated to cross-cutting themes, such as organisational profile information, are marginal, with around 80 questions. This distribution reflects the centrality attributed to environmental issues in ESG debate and practice, but also the difficulty of balancing the three pillars of sustainability evenly in assessment processes.

Analysis of the types of questions reveals a clear trend towards structured and easily codifiable forms. Closed questions are in fact the most frequent category, with 1,145 occurrences, followed by open questions (417) and those based on Likert scales (276). This prevalence is further confirmed by the fact that over two-thirds of the questionnaires have a proportion of closed questions exceeding 70% of the total. This suggests a widespread preference for formats that facilitate coding, comparability of responses and quantitative processing of the information collected. Open-ended questions, although less frequent, continue to play an important role, especially in questionnaires developed by research institutions or social organisations, where there is a prevailing interest in the quality of practices and the narration of processes. Finally, Likert scale questions appear to be concentrated in banking questionnaires and internal company assessments, where they are used to measure perceptions and levels of maturity in specific ESG areas.

The type of responses varies considerably between questionnaires, but shows a clear predominance of the text format, which represents the absolute majority of cases (1,233 out of 1,839). This is followed, with significantly lower frequencies, by responses structured on a Likert scale (258) and those expressed as a percentage (115). The limited recurrence of more specific quantitative units of measurement, such as tonnes of CO_2, euros or kWh, suggests that only some of the questionnaires take a truly metric approach to ESG assessment. On average, each questionnaire includes about three different types of measurement, but the distribution is highly asymmetrical: most questionnaires (19 out of 33) are characterised by a low level of detail, with only two units of measurement, generally textual units combined with numerical units. The level of detail increases slightly, with about two-thirds of questionnaires including three units of

measurement in the responses. Only five questionnaires show a high level of detail in the responses, including between eight and ten different measurement units. In these more complex cases, the combined presence of quantitative indicators and descriptive responses allows for a more accurate understanding of the multidimensionality of ESG practices. However, the majority of questionnaires appear to be structurally simplified, confirming a still elementary approach to data collection.

A more in-depth examination of individual questionnaires revealed marked variability in the structure, length and depth of the questions. The number of questions ranges from a minimum of 7 in the most concise questionnaires to a maximum of 252 in questionnaire 23, which is the most extensive and detailed. These quantitative differences reflect profoundly different approaches to data collection: some instruments aim at a broad but standardised exploration, while others prefer a few targeted and qualitative questions. The average length of the questions also varies considerably, ranging from about nine words in the shortest questionnaires to over thirty-four in the most complex ones. In particular, questionnaires 5 and 10, with average lengths of 207 and 239 characters respectively, include complex, often multi-part questions that require in-depth reflection on the part of the respondent. In contrast, more concise instruments, such as questionnaire 8, are characterised by concise wording, focusing on quick and codifiable answers.

Additional variation among the questionnaires emerges from the way topics are distributed within their items. Six instruments place a predominant emphasis on environmental matters, with more than 45% of their questions devoted to this area. Two questionnaires show an equivalent concentration but oriented toward social issues, while three focus primarily on governance, dedicating a comparable share of their questions to organisational, ethical, and compliance-related themes.

The vocabulary also varies depending on the source and purpose of the tool. The most extensive questionnaires, such as 23 and 32, are dominated by technical vocabulary focused on environmental metrics and governance, with terms such as “emissions”, “GHG”, “scope”, “targets” and “total”, typical of quantitative reporting practices. Other questionnaires, such as 11, show a more markedly social orientation, with recurring words such as “women”, “gender”, “CCNL” and “flexibility”, while instruments such as 17 focus on balance sheet and time elements, with high frequencies of terms such as “balance sheet” and “years”. Questionnaire 24 stands out for its use of technical language related to information management and compliance, while questionnaire 10 emphasises the relational dimension, with frequent references to “stakeholders”, “community” and “territory”.

Overall, the results confirm a strong structural and linguistic heterogeneity of ESG questionnaires, reflecting the plurality of approaches, purposes and skills of the actors involved. This variability has direct implications for both the quality and comparability of the data collected and the level of effort required of respondents. While some tools show a more mature and systematic approach, capable of integrating quantitative measurements and qualitative assessments, others are limited to descriptive and partial surveys, a sign of a field still in evolution and a methodological standardisation that is far from consolidated.

Cosine similarity

The cosine similarity analysis made it possible to assess the degree of lexical similarity between the texts of the questions contained in the various questionnaires, in order to understand how

similar or divergent the linguistic and terminological formulations were. For each pair of questionnaires, the entire textual content was compared, appropriately vectorised using the TF-IDF technique (Guleria et al., 2025), with the removal of stopwords in Italian and the limitation of the number of features to 500, in order to capture the linguistic essence of the documents in a concise but effective manner.

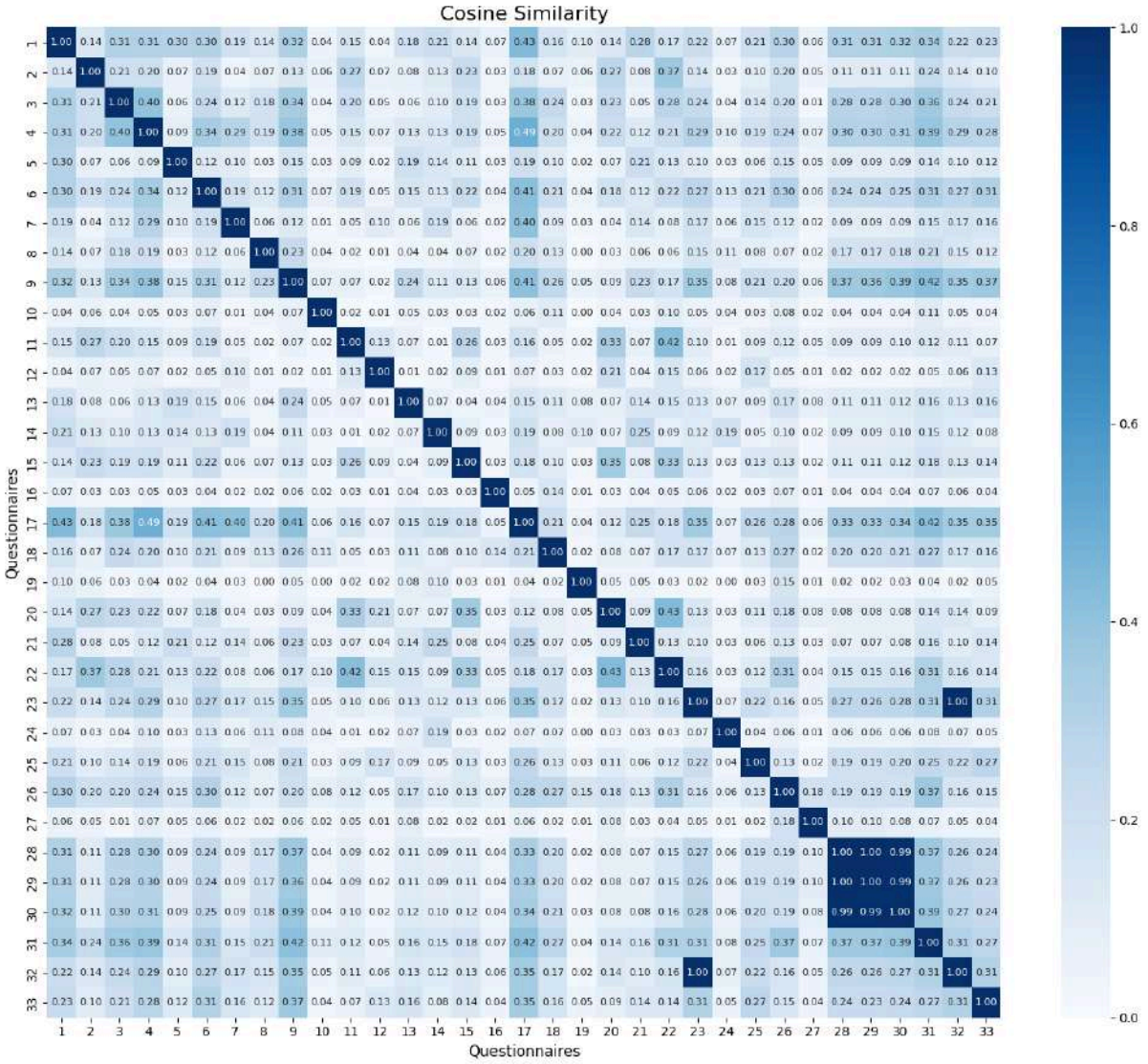

Figure 1 - Source: Authors own work

The results of the analysis show an average cosine similarity value of 0.140, with variability ranging from a minimum of 0.000 to a maximum, excluding 1, of 0.491, as illustrated in Figure 1. This extremely low average value highlights the poor similarity between the questionnaires, which share a limited number of terms and have widely divergent linguistic profiles. Even the highest value recorded, 0.491, while indicating partial lexical convergence, still suggests substantial differences in the structure and distribution of terms. Only in one case was a cosine similarity close to 1.00 found, a value that indicates an almost perfect coincidence in the direction of the TF-IDF vectors and therefore a proportional equivalence in the presence and relative frequency of terms, a sign of an almost complete overlap in the formulation of the questions: questionnaires 23 and 32 deal with

substantially similar topics, with almost complete adherence in the choice of key terms. However, the differences in TF-IDF weights, however small, indicate that these are not perfect duplicates: the versions may derive from a common but adapted model.
Overall, most pairs of questionnaires (638, or 60.4% of the total) have below-average similarity values, confirming a high degree of linguistic heterogeneity within the analysed corpus. This result suggests that, while sharing the general objective of investigating aspects related to ESG issues, the questionnaires differ profoundly in their choice of terminology, question construction and textual layout, reflecting heterogeneous conceptual and editorial approaches in the design of the survey instruments.

Cluster analysis

Cluster analysis conducted on the aggregated data identified six distinct groups of questionnaires, using both the k-means algorithm and the agglomerative hierarchical method. The main objective was to explore the structural, linguistic and semantic similarities between the questionnaires in order to outline homogeneous profiles in terms of both construction and content. Both approaches produced a segmentation into six clusters, albeit with differences in composition and degree of internal cohesion, reflecting the intrinsic nature of the two methods and the different logic with which they identify the distance and membership of cases to groups. A comparison of the two approaches indicates that the hierarchical solution yields more coherent and higher-quality groupings. The internal validation measures point to an improvement across all metrics: the Silhouette Score rises from 0.163 under k-means to 0.197 with the hierarchical method, signalling tighter clusters and greater separation on average. The Calinski–Harabasz index also increases, moving from 5.757 to 6.324, which reflects a clearer delineation of cluster boundaries. At the same time, the Davies–Bouldin index, where lower values indicate better cluster compactness, falls from 1.460 to 1.375, further suggesting that the hierarchical approach produces a more stable and internally consistent segmentation.

Table II shows how the different clusters obtained with the k-means algorithm differ in terms of the most representative words, average number of questions, ESG theme and type of most frequent questions, average number of words per question and unit of measurement of responses.

Summary description of the clusters obtained with KMeans

Cluster	Cluster size	Most representative words	Average number of questions	Most frequent ESG topic	Type of most frequent questions	Average number of words per question	Average number of response measurement units
0	5	certifications, company, customer	29,60	Environmental	Closed - binary	14,16	1,40
1	7	company, percentage, scope	106,14	Environmental	Closed - multiple	26,81	8,14
2	10	company, selection, activity	49,60	Governance	Closed - multiple	18,04	3,00

3	3	ESG, bank, solution	24,33	ESG	Likert	20,19	2,33
4	4	business, rules, procedures	29,00	Social	Closed - multiple	11,58	2,75
5	4	company, technology, management	65,75	Governance	Likert	17,93	1,5

Table II - Source: Authors own work

Cluster 1 is the most detailed, distinguished by the highest average number of questions (106.14), the greater length of the formulations (26.81 words on average) and the level of detail required in the answers, which reaches the highest value among all clusters (8.14). In contrast to this configuration, cluster 0, while sharing the environmental focus with cluster 1, has a decidedly more basic structure: the average number of questions is low (29.60), the average length is short (14.16 words, the second lowest after cluster 4) and the degree of detail required in the answers is the lowest recorded in the entire segmentation (1.40).

Table III shows how the clusters derived from the hierarchical method have different configurations compared to those obtained with the k-means method and have distinct configurations in terms of most representative words, average number of questions, prevailing ESG theme, most used types of questions, average length of the formulations and level of detail required in the answers through the relative units of measurement.

Summary description of the clusters obtained with Hierarchical							
Cluster	Cluster size	Most representative words	Average number of questions	Most frequent ESG topic	Type of most frequent questions	Average number of words per question	Average number of response measurement units
1	3	ESG, banking, CSR	24,33	Environmental	Closed - binary	20,19	2,33
2	5	Company, technology, impact	56,20	Environmental	Closed - multiple	16,42	1,60
3	3	Company, shares, scopes	191,33	Governance	Closed - multiple	43,89	9,67
4	3	Company, rules, procedures	32,67	ESG	Likert	11,98	3,00
5	7	Company, selection, sustainability	44,71	Social	Closed - multiple	15,52	1,86
6	12	Company, employees, emissions	41,67	Governance	Likert	16,55	4,33

Table III - Source: Authors own work

The examination of the clustering outcomes indicates that Cluster 3 is the most information-rich group: it features the highest average number of measurement units attached to responses (9.67) and includes some of the most extensive questionnaires, both in terms of the volume of questions and the length of their formulations. In contrast, Cluster 4 exhibits a substantially lower degree of detail; the questionnaires within this group address ESG topics in a broad and generalised manner, rely predominantly on Likert-scale items, and contain the shortest questions on average across all clusters.
Taken together, these findings illustrate the distinct profiles that characterise each cluster, underscoring the pronounced variability among the questionnaires with respect to structural configuration, linguistic choices and substantive content.

4. Discussion

The descriptive examination of the ESG questionnaires reveals substantial structural diversity, reflecting the wide range of strategies employed to assess sustainability. Notably, the instruments analysed differ sharply both in the total number of questions, spanning from as few as 7 to as many as 252, and in the average length of their items, which varies between 8.7 and 34.6 words. Beyond their formal aspects, these discrepancies point to pronounced methodological variation among the organisations that design ESG assessment tools, each of which adopts its own approach to gathering and organising information (Louche et al., 2023).
Additional indications of this heterogeneity emerge when considering the level of detail required in the responses. The degree of specificity expected from respondents varies widely across instruments: on average, a questionnaire includes around 3.4 distinct response measurement units, with values ranging from 1 to as many as 10. Only a small subset (five questionnaires) exhibits a notably high degree of granularity, incorporating between eight and ten measurement units and thus signalling a more rigorous and comprehensive orientation toward ESG data collection. In contrast, the majority of the tools rely on simpler structures, suggesting that sustainability assessment practices remain heavily anchored in quantitative and standardisable metrics.
Moreover, the predominance of closed-ended questions reinforces a broader movement toward standardisation. While this format enhances efficiency and facilitates comparison across respondents and organisations, it also constrains the ability to capture richer qualitative insights that are more likely to emerge through open-ended questions, which inherently demand greater interpretive and analytical effort (Baid and Vaidyanathan, 2022).
While these methodological choices facilitate the aggregation of information into summary indicators, they raise questions about the actual ability of the questionnaires to represent the complexity of ESG practices and the informational quality of the resulting ratings.
The thematic differences observed further contribute to this fragmentation. Six questionnaires focus mainly on environmental aspects, two favour the social dimension and three focus on governance, with a distribution of questions that exceeds 45% on the dominant theme in each case. This diversification raises a crucial question about the actual object of measurement, as the preponderance of some areas over others reflects the different purposes of the evaluators (Louche et al., 2023) .

Linguistic analysis further confirms the heterogeneity between the instruments, showing how terminological choices reflect different evaluation priorities. The frequency of terms varies according to the thematic orientation of the questionnaires: those focused on the environment use technical and quantitative language, with recurring terms such as emissions, resources, consumption, tCO_2 and waste (Sharma and Bandyopadhyay, 2023); questionnaires with a predominantly social orientation use more relational and qualitative vocabulary, centred on words such as employees, training, work, gender and diversity; while those focused on governance use terms such as model, organisational, code and ethical, indicating a more regulatory and procedural approach.

Cosine similarity analysis enabled a numerical assessment of the linguistic variation across questionnaires, revealing an overall mean similarity of 0.140. This low value indicates minimal lexical overlap and markedly different linguistic patterns among the instruments. More than 60% of questionnaire pairs (638 out of 1,056) fall below this already modest average, underscoring a substantial level of conceptual and terminological divergence. Such variation mirrors the multiplicity of ways in which sustainability is interpreted (Amini et al., 2028) and points to the need for greater terminological alignment to foster clearer communication and more coherent data collection, conditions that are crucial for enhancing the comparability and robustness of ESG evaluations, which are currently hindered by divergent measurement and weighting approaches (Berg et al., 2022).

The cluster analysis offered an additional layer of insight by grouping the questionnaires into six distinct sets, each characterised not only by its most salient terms and its dominant ESG dimension, but also by structural attributes such as the average number of questions (ranging from 191.33 to 24.33), the average length of the questions (from 43.89 to 11.98 words), and the mean number of measurement units used in responses (from 9.67 to 1.60).

These contrasts suggest that each cluster embodies a different methodological orientation to ESG assessment, distinguished by both the thematic focus and the degree of analytical sophistication required. The resulting fragmentation shows that, despite the shared goal of evaluating corporate sustainability, the tools rely on widely divergent operational logics, producing a landscape marked by pronounced structural, linguistic, and functional heterogeneity (Chen, 2023).

The lack of methodological consistency compromises the comparability between questionnaires, generating substantial differences in data collection and processing protocols. As noted by Ferro et al. (2025), these discrepancies in the measurement phase are one of the main causes of misalignment between ESG ratings, directly influencing the construction of final scores.

Taken together, the findings indicate that the absence of consistency across ESG questionnaires mirrors a wider fragmentation in the ways sustainability is measured. Reducing these divergences could help address long-standing issues of inconsistency and variability in ESG ratings (Mio et al., 2024). Without common standards for data collection and evaluation, comparability across assessments remains limited (Benuzzi et al., 2025), and companies may be incentivised to highlight only those evaluations that cast them in a favourable light. At the same time, the coexistence of numerous tools, structures, and linguistic conventions contributes to ambiguity and complexity for reporting entities.

Enhancing alignment in the structure, terminology, and content of questionnaires would support the development of more rigorous approaches to assessing ESG performance (Soares, 2024) and foster the production of standardised, high-quality data. This is crucial for the generation of ratings

that are more comparable, transparent, and meaningful for both investors and corporate decision-makers (Lopez et al., 2020).

5. Policy Recommendation

The empirical evidence presented in this study reveals a marked fragmentation in the structure, language, and content of ESG questionnaires, which represent the very foundation of sustainability assessments. This heterogeneity is manifested in the number of questions, the depth and nature of the information requested, the degree of linguistic complexity and the diversity of units of measurement, and has direct implications for the comparability, transparency and reliability of ESG ratings. These findings are particularly relevant in relation to Regulation (EU) 2024/3005, which establishes a harmonised framework to ensure integrity, transparency and accountability in ESG rating activities across the European Union.

While the Regulation does not impose a single methodology or uniform scoring models, it places strong emphasis on standardizing processes, including the adoption of systematic and robust methodologies, transparent disclosure of rating models, clear communication of assumptions and weighting schemes, and the use of reliable information sources. The results of this study suggest that these requirements cannot be fully met unless greater attention is paid to the initial phase of the evaluation process, the data collection. The Regulation, in relation to this, highlights how providers must take all necessary measures to ensure that the information used for the issuance of ESG ratings is of sufficient quality and comes from reliable sources, however it does not express precise measures for the collection of information.

The marked heterogeneity documented in this research therefore indicates an important area in which policymakers and regulators can intervene. First, in order to improve the reliability of final ratings, minimum quality standards could be established for ESG data collection tools. Such standards could include clearer definitions of ESG themes, harmonized terminology and basic requirements for the structure of questionnaires, but without imposing an overly prescriptive or rigid format in such a way as to leave room for the discretion of the various agents to maintain an adequate level of competition.

Second, it may be optimal to introduce guidelines to standardize the language used in data collection tools and how they are measured. Doing so would reduce diversity in wording and response formats which not only complicates comparability but can also introduce distortions related to interpretation, respondent burden or different disclosures. Clearer language conventions and more consistent units of measurement would strengthen the reliability of the data on which ESG ratings are based, improving their alignment with the integrity and comparability objectives of the Regulation.

Finally, policymakers could consider promoting interoperability frameworks aimed at organizations that could provide ESG data only once, diminishing the administrative honor of companies that would otherwise be required to complete several highly heterogeneous questionnaires. This would achieve greater alignment in data collection processes by contributing to a more coherent and integrated ESG reporting ecosystem, ultimately supporting the Regulation's goal of steering capital flows towards sustainable activities.

In summary, while Regulation (EU) 2024/3005 represents a decisive step towards strengthening the transparency and reliability of ESG ratings, the findings of this study highlight the need to extend the harmonization approach to the data collection phase, as refining the structure, clarity,

and methodological consistency of ESG questionnaires would not only improve the quality of the underlying information but also strengthen the credibility of the entire ESG rating system, enabling ratings to more effectively support sustainable and responsible investment decisions across the European Union.

6. Conclusion

ESG evaluations have become a central instrument for assessing corporate sustainability performance, shaping investment choices, regulatory adherence, reputation management, and strategic decision-making (Berg et al., 2022). Their rapid diffusion underscores how essential they have become for encouraging organisations to align their practices with environmental, social, and governance principles. Yet this growing importance also heightens the need for assessment processes that are reliable, coherent, and transparent. Paradoxically, while ESG ratings increasingly influence economic and managerial decisions, the earliest stage of the evaluation process—the collection of underlying data—has remained largely overlooked, despite its fundamental role in shaping sustainability metrics.

This study addresses this gap by examining the questionnaires employed by rating agencies, public authorities, financial institutions, and companies to gather ESG-related information. Using a three-tier analytical framework—descriptive analysis, lexical analysis through cosine similarity, and cluster analysis—the research uncovers substantial variation in how these tools are designed and formulated. The findings reveal wide disparities not only in the volume and typology of questions but also in linguistic style, terminological precision, and the depth of detail expected in the responses. While these divergences may seem merely operational, they point to more fundamental differences in how the notion of sustainability is interpreted and translated into evaluative practices across different actors.

This fragmentation has significant implications. In the absence of shared and transparent data collection protocols, the comparability and interpretability of ESG assessments are compromised. This reduces the reliability of these tools as decision-making aids and paves the way for opportunistic behaviour, whereby companies can select or favour assessments that represent them in a more favourable light. Furthermore, the proliferation of heterogeneous and sometimes opaque tools creates further confusion for companies called upon to respond to multiple information requests, often based on definitions and measurements that are not aligned with each other.

From a regulatory perspective, the results of this analysis highlight the urgent need for greater harmonization in ESG data collection and reporting. The European Union is already advancing in this direction with tools such as CSRD and, most importantly, with the new Regulation (EU) 2024/3005 on ESG ratings, which introduces rigorous standards of transparency, data quality and integrity of evaluation processes.

However, although the regulation significantly strengthens the rating processing phase, it only marginally affects the preliminary data collection phase, which this study shows to be highly heterogeneous. Without more consistent criteria for the design of information-gathering instruments the standardization required of rating providers is likely to rest on a weak and non-comparable information base.

Consequently, the effectiveness of the new European framework will increasingly depend on the ability to extend the logic of harmonisation to the data acquisition phase as well, so as to ensure

that ESG ratings are based on homogeneous, reliable and genuinely comparable information in the single market.

The main contribution of this study lies in its systematic exploration of a hitherto neglected dimension of ESG assessments: the questionnaires that form their information base. By highlighting the inconsistencies and structural fragmentation present at this fundamental level, the research offers an innovative perspective for understanding the broader challenges associated with measuring sustainability. Future developments could extend the analysis to a larger sample, including different geographical areas and sectors, or directly involve rating agencies to investigate the interpretative processes that guide the construction of their models. Only through greater transparency and methodological convergence can ESG assessments truly fulfil their function as tools capable of promoting concrete and measurable sustainability outcomes.

From a regulatory standpoint, the findings of this study underscore the pressing need to harmonise ESG data collection and disclosure practices. Although initiatives such as the European Union's Corporate Sustainability Reporting Directive (CSRD) mark important progress toward greater standardisation, they must be accompanied by clearer, more widely shared methodologies for gathering ESG information if the quality and comparability of sustainability assessments are to be meaningfully improved.

Gennaio 2025 - Pubblicazione finanziata dal PNRR, Missione 4 (Infrastruttura e Ricerca), Componente 2 (Dalla Ricerca all'Impresa), Investimento 1.3 (Partnership Estese), Tematica 9 (Sostenibilità economica di sistemi e territori). Sito web: https://grins.it/

References

Amini, M., Bienstock, C. C., & Narcum, J. A. (2018). Status of corporate sustainability: a content analysis of Fortune 500 companies. *Business Strategy and the Environment*, *27*(8), 1450–1461. https://doi.org/10.1002/bse.2195

Anselmi, G., & Petrella, G. (2023). ESG Ratings: Disagreement across Providers and Effects on Stock Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4328468

Assaf, C., Monne, J., Harriet, L., & Meunier, L. (2024). ESG investing: Does one score fit all investors' preferences? *Journal of Cleaner Production*, *443*, 141094. https://doi.org/10.1016/j.jclepro.2024.141094

Baid, V., & Jayaraman, V. (2022). Amplifying and promoting the "S" in ESG investing: the case for social responsibility in supply chain financing. *Managerial Finance*, *48*(8), 1279–1297. https://doi.org/10.1108/MF-12-2021-0588

Benuzzi, M., Bax, K., Paterlini, S., & Taufer, E. (2025). Chasing ESG performance: How methodologies shape outcomes. *International Review of Financial Analysis*, *104*. https://doi.org/10.1016/j.irfa.2025.104239

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, *26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2020). Inside the ESG Ratings: (Dis)Agreement and Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3659271

Bolliger, R., Fischer, B., Amaral, M. G. do, de Faria, A. F., & Serafim, M. P. (2025). Are technology parks the same? A typology proposal for an emerging economy. *Innovation & Management Review*, 1–18. https://doi.org/10.1108/INMR-12-2023-0245

Bonaventura Forleo, M., & Bredice, M. (2025). Italian Gen Z and sustainable coastal tourism: a segmentation analysis of knowledge, attitudes and pro-environmental behaviours. *Journal of Tourism Futures*. https://doi.org/10.1108/JTF-07-2024-0146

Bronzini, M., Nicolini, C., Lepri, B., Passerini, A., & Staiano, J. (2024). Glitter or gold? Deriving structured insights from sustainability reports via large language models. *EPJ Data Science*, *13*(1), 41. https://doi.org/10.1140/epjds/s13688-024-00481-2

Buchetti, B., Arduino, F. R., & Perdichizzi, S. (2025). A literature review on corporate governance and ESG research: Emerging trends and future directions. *International Review of Financial Analysis*, *97*, 103759. https://doi.org/10.1016/j.irfa.2024.103759

Carnini Pulino, S., Ciaburri, M., Magnanelli, B. S., & Nasta, L. (2022). Does ESG Disclosure Influence Firm Performance? *Sustainability*, *14*(13), 7595. https://doi.org/10.3390/su14137595

Chen, R. (2023). Assessment of Potential Risks of Current ESG Investment. *Advances in Economics, Management and Political Sciences*, *39*(1), 188–193. https://doi.org/10.54254/2754-1169/39/20231967

Cregan, C., Kelly, J. A., & Clinch, J. P. (2025). The Use of ESG Ratings in Corporate Compensation Plans: Promises and Pitfalls. *Compensation and Benefits Review*. https://doi.org/10.1177/08863687251329843

Del Giudice, A., Gallucci, C., & Santulli, R. (2024). *FIN-GOV I rating ESG: un confronto internazionale*. www.unicatt.it

D'Este, C., Galavotti, I., & Cantoni, F. (2025). The cascade effect of women on boards: how firm-level gender diversity management develops intellectual capital. *Corporate Governance: The International Journal of Business in Society*, *25*(8), 132–155. https://doi.org/10.1108/CG-11-2024-0577

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. https://doi.org/10.1057/jam.2015.31

Duong, C., Sung, B., Lee, S., & Easton, J. (2025). Assessing the Australian abalone market: a best-worst approach on 46 attributes. *British Food Journal*, *127*(13), 535–557. https://doi.org/10.1108/BFJ-05-2024-0485

Escrig-Olmedo, E., Fernández-Izquierdo, M. ángeles, Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability (Switzerland)*, *11*(3). https://doi.org/10.3390/su11030915

Ferro, A., Marazzina, D., & Stocco, D. (2025). Uncovering ESG Ratings: The (Im)Balance of Aspirational and Performance Features. *Corporate Social Responsibility and Environmental Management*. https://doi.org/10.1002/csr.70007

González-Pozo, R., Arenas-Parra, M., Quiroga-García, R., & Bilbao-Terol, A. (2025). A proposal for refining the ESG methodology used by rating agencies. *International Transactions in Operational Research*, *32*(4), 2003–2033. https://doi.org/10.1111/itor.13550

Guleria, P., Frnda, J., & Srinivasu, P. N. (2025). NLP based text classification using TF-IDF enabled fine-tuned long short-term memory: An empirical analysis. *Array*, *27*. https://doi.org/10.1016/j.array.2025.100467

Heiskanen, A., & Ryynänen, T. (2024). Optimists, moderates and sceptics – identifying consumer groups and their willingness to consume cultured proteins in Finland. *British Food Journal*, *126*(13), 658–671. https://doi.org/10.1108/BFJ-03-2024-0268

Kim, R., & Koo, B. (2023). The impact of ESG rating disagreement on corporate value. *Journal of Derivatives and Quantitative Studies*, *31*(3), 219–241. https://doi.org/10.1108/JDQS-01-2023-0001

Liang, H., & Renneboog, L. (2020). Corporate Social Responsibility and Sustainable Finance: A Review of the Literature. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3698631

Lopez, C., Contreras, O., & Bendix, J. (2020). *Disagreement among ESG rating agencies: shall we be worried?* https://mpra.ub.uni-muenchen.de/103027/

Louche, C., Delautre, G., & Balvedi Pimentel, G. (2023). Assessing companies' decent work practices: An analysis of ESG rating methodologies. *International Labour Review*, *162*(1), 69–97. https://doi.org/10.1111/ilr.12370

Mele, E., Dubosson, M., & Schegg, R. (2024). Are all luxury guests the same? A benefit segmentation of 5-star hotel customers. *Journal of Hospitality and Tourism Insights*, *8*(11), 39–54. https://doi.org/10.1108/JHTI-04-2024-0336

Menicucci, E., & Paolucci, G. (2024). Board gender equality and ESG performance. Evidence from European banking sector. *Corporate Governance: The International Journal of Business in Society*, *24*(8), 147–174. https://doi.org/10.1108/CG-04-2023-0146

Mio, C., Fasan, M., Costantini, A., Scarpa, F., & Fitzpatrick, A. C. (2024). *Unveiling the Consequences of ESG Rating Disagreement: An Empirical Analysis of the Impact on the Cost of Equity Capital*. https://doi.org/10.2139/ssrn.5054102

Oniani, D. (2020). *Cosine Similarity and Its Applications in the Domains of Artificial Intelligence*.

Oyinlola, B. (2025). Do CEO and board characteristics matter in the ESG performance of their firms? *Corporate Governance: The International Journal of Business in Society*, *25*(8), 21–39. https://doi.org/10.1108/CG-01-2024-0052

Punj, G., & Stewart, D. W. (1983). Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research*, *20*(2), 134. https://doi.org/10.2307/3151680

Rao, S., Juma, N., & Srinivasan, K. (2025). Textual Analysis of Sustainability Reports: Topics, Firm Value, and the Moderating Role of Assurance. *Journal of Risk and Financial Management*, *18*(8), 463. https://doi.org/10.3390/jrfm18080463

Serafeim, G., & Yoon, A. (2023). Stock Price Reactions to ESG News: The Role of ESG Ratings and Disagreement. *Springer, vol. 28(3), pages 1500-1530, September.* https://doi.org/10.1007/s11142-022-09675-3

Sharma, P., & Bandyopadhyay, S. (2023). A quantitative framework for sustainability assessment. *Clean Technologies and Environmental Policy*, *25*(9), 2971–2985. https://doi.org/10.1007/s10098-023-02541-z

Soares, C. P. (2024). *Leveraging Natural Language and Item Response Theory Models for ESG Scoring*. http://arxiv.org/abs/2407.20377