POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Linking Assurance to ESG Performance and Reporting Quality: Evidence from Italian and French Benefit Corporations

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

AUTHOR: MEROUR Juliette

ADVISOR: BONI Leonardo

CO-ADVISOR: KOPAL Hazal

ACADEMIC YEAR: 2025-2026

1. Introduction

ESG considerations have become increasingly important and are now widely integrated into investment decisions. In this context, non-financial reporting has become the main channel through which companies communicate their sustainability performance. However, investors have been raising concerns, not about the financial return of ESG investments, but about the veracity of the information disclosed. Non-financial reports still lack standardization, which leaves room for greenwashing and selective disclosure.

To address this credibility gap, external assurance has emerged. An independent party verifies the non-financial disclosures and expresses an opinion on their reliability [1].

This topic is particularly relevant in the European Union. Indeed, the Corporate Sustainability Reporting Directive (CSRD) requires limited assurance of sustainability statements, while the Omnibus I package has removed the planned move to reasonable assurance.

However, whether assurance actually improves what it certifies remains an open question, and evidence on its association with ESG performance and report quality is mixed. This raises a question of direct policy relevance: is mandating assurance the right way to secure its benefits?

Benefit Corporations make it possible to study this question because of their different institutional contexts. In France, every société à mission is required to have its mission and mission report verified by an independent third-party body (OTI), while in Italy assurance remains voluntary. This study exploits that institutional asymmetry to examine whether assurance, and the regime under which it is imposed, is associated with higher ESG performance and higher reporting quality.

2. Literature review

The literature review covers five main parts: ESG reporting, ESG ratings, ESG assurance, B Corp and EcoVadis, and Benefit Corporations.

Overall, disclosure alone does not reduce information asymmetry. To do so, high-quality disclosure is needed. Yet, there is no univocal definition of disclosure quality, which complicates its study.

ESG ratings and certifications have become a key tool for investors to assess the sustainability performance of companies, but the divergence among rating providers raises concerns about their usefulness. Moreover, it is not always clear whether higher ESG ratings are also associated with higher sustainability performance, as evidence is mixed.

There are different levels of assurance. Because assurance is based on the report, the current immaturity of non-financial reports slows down the development of assurance. Some studies highlight the important role of assurance in improving the internal organization and sustainability commitment of companies, but also show that a certain level of sustainability maturity is needed when implementing assurance in order to capture its benefits [2].

Some gaps in the literature were identified. First, few studies actually examine the relationship between assurance and ESG performance. Second, reporting quality assessment suffers from a lack of harmonization and high fragmentation. Third, most studies have focused on large listed firms, while few of them have examined SMEs and hybrid-purpose companies such as Benefit Corporations. This study tries to fill these gaps.

3. Research objectives

This study examines whether assurance is associated with higher ESG performance and report quality, beyond its role as a credibility signal. The research question is the following:

RQ: *Is the presence of external assurance associated with higher ESG performance and higher report quality?*

The theoretical framework built to answer this question is mainly based on agency theory.

Under agency theory, there is information asymmetry between managers and owners. Managers are incentivized to hide information from owners to protect their own interests [3]. With non-financial disclosure, this can translate into greenwashing or selective disclosure. Assurance helps reduce this behavior. Knowing that their reports will be assured, managers are incentivized to improve their sustainability performance ex ante. Since external ESG grades are assigned independently of the assurance engagement, assurance can only raise them indirectly, through real ESG improvements. A competing view is proposed by legitimacy theory: assurance would act as a symbolic signal decoupled from actual performance [1]. Given assurance's documented disciplining role, the agency-theoretic expectation is retained, which leads to the first hypothesis:

H1: *Assured companies have higher ESG performance than unassured companies.*

Assurance requires managers to take responsibility for what they are reporting and to benchmark it against recognized standards and frameworks. Based on agency theory, this should encourage more comprehensive, internally consistent, and standardized disclosure. Its effect may be limited, however, because companies can choose which sections to have assured, as legitimacy theory predicts. Nevertheless, assurance is expected to constrain managerial discretion and improve reporting, which leads to the second hypothesis:

H2: *ESG report quality is higher for assured companies than for unassured companies.*

4. Methodology

The sample is composed of French sociétés à mission and Italian Società Benefit. These two legal statuses are very similar. However, they differ in one aspect: under the PACTE law, French sociétés à mission are required to have their reports verified by an independent third party. Reports from 2024 were collected from companies' websites, and ESG grades were collected from B Corp and EcoVadis. These grades were normalized out of 100 and averaged when a company had both. Self-assessed scores, national ratings, and ratings from large agencies (available only for the largest firms) were excluded. Overall, there are 104 Italian companies (119 reports) and 84 French ones (92 reports).

Two different approaches were used, one for each hypothesis. For H1, the analysis was conducted at the company level (N = 188; 97 assured, 91 unassured). A company was considered assured when at least one of its reports was assured. For H2, the analysis was carried out at the report level (N = 211). Reports are grouped into impact/mission reports (N = 158), which Benefit Corporations must publish every year, and sustainability, CSR, integrated, and annual reports (N = 53), which are voluntary.

Both hypotheses were analyzed using quantile regression to account for differences across the distribution [4]. It allows the association between assurance and the outcome to differ between low performers (10th and 25th percentiles), the median, and high performers (75th and 90th percentiles). Some control variables were also included: sector, size, log(age), sustainability report publication (only for H1), type of report (only for H2), and the number of years since the company obtained Benefit Corporation status. A coefficient is considered significant when $p < 0.05$ (or its 95% confidence interval excludes zero). A 95% bootstrap confidence interval is used as a robustness check, and multicollinearity is checked with a Pearson correlation matrix and GVIF.

Endogeneity is addressed in two complementary ways: a Lewbel heteroskedasticity-based instrumental-variable approach for unobservable characteristics, complemented by Durbin and Wu–Hausman tests, and propensity score matching for observable ones [5].

4.1. Methodology H1

H1 was tested in two different ways. First, the relationship between assurance/assurance regime and external ESG grades was investigated. Because results on the link between sustainability performance and external ESG grades are mixed, a self-constructed ESG score was also built, aimed at capturing the sustainability performance of the company by applying Natural Language Processing to the reports. This score is based on three different components:

- **ESG sentiment score:** each sentence is classified into one of the SDGs with SDG BERT, scored for sentiment with an equally weighted combination of FinBERT and RoBERTa, and aggregated into E, S, and G pillars weighted by their share of sentences.
- **Sustainability redundancy (SR) score:** the share of words expressing vague, hedged, or superlative language, which adds length but not information [6]. It is applied as a penalty on the overall score.
- **Business model innovation (BMI) score:** Sentence-BERT similarity between report sentences and literature-based seed sentences on sustainable value proposition, creation and delivery, and capture, amplified by the share of transformative (rather than incremental) sentences.

The final score is then given by:

$$ESG\ score = ESG\ sentiment\ score \times (1 - SR\ score) \times BMI\ score \quad (4.\ 1)$$

4.2. Methodology H2

H2 was tested in three different ways. The relationship between assurance and report quality was tested first on impact/mission reports, then on sustainability reports, and finally on all the reports together. This distinction was made because impact/mission reports and sustainability reports differ in their goal and orientation, and sustainability reports are often published by companies with higher sustainability maturity.

To assess the quality of the reports, an assessment framework was built based on the literature. It covers six pillars: **completeness** (coverage of the material topics), **balance** (proportion of environmental, social, and governance topics), **KPIs** (quantitative and forward-looking indicators), **readability**, **conciseness** (number of pages), and **comparability** (whether the report follows reporting guidelines). Each pillar was assigned a specific weight reflecting its relative importance. The final score is the following:

$$Quality = \frac{(C + R) \times 5 + (K + B + Comp) \times 10 + CP \times 20}{60} \quad (4.\ 2)$$

where C is conciseness, R readability, B balance, K KPIs, CP completeness and Comp comparability, each scored out of 10.

5. Results

5.1. Results H1

Assurance is associated with higher external ESG grades for low-to-mid performers at the **25th** (p = 0.019) and **50th** (p = 0.003) percentiles. Assured companies score between **4.0** (at the 25th percentile) and **5.7** (at the 50th percentile) points (out of 100) higher than non-assured companies. At lower and higher quantiles, there is no significant association. When replacing assurance with the mandatory regime, the results are highly similar. These results pass the bootstrap robustness test.

However, when looking at the association between the self-constructed ESG score and assurance, the results point in the other direction. Assurance is almost never significant, and most of the results do not pass the bootstrap robustness check, which indicates that they are still fragile. The only robust finding (for both assurance and assurance regime) concerns the sustainability redundancy score: at the 90th percentile, assurance is associated with a higher redundancy score **(+0.002, p < 0.001**, bootstrap CI [0.000, 0.004]). The total score (75th and 90th percentiles), the sentiment score (90th percentile), and the BMI score (10th and 25th percentiles) also show negative coefficients, but these do not pass the bootstrap check.

5.2. Results H2

The association between assurance and reporting quality is positive but not robust, as none of the tests pass the bootstrap robustness check. For impact/mission reports, assurance is associated with **+0.596** points (out of 10) at the 90th percentile (p = 0.**037**). For sustainability reports, no significant relationship was found, and for all the reports together, assurance is associated with **+0.453** points at the 10th percentile (**p = 0.031**).

In this analysis, the strongest robust predictor of quality is report type: in the pooled sample, classification as a sustainability report is associated with **+0.918** and **+1.216** points at the 25th and 50th percentiles.

5.3. Endogeneity

The endogeneity tests are largely inconclusive, mainly because of weak instruments and low matching rates. At best, they weakly support a failure to reject exogeneity. Reverse causality is less plausible in this setting, since French assurance is imposed by law and the quality framework was built for this study. A joint-determination story, in which reporting maturity drives both assurance and outcomes, cannot be ruled out. All results should therefore be read as associations rather than causal effects.

6. Discussion

Across all tests, a firm's organizational and reporting maturity appears to matter more than the presence of assurance itself.

The results only partially support H1. When looking at external ESG grades as the dependent variable, the gains are concentrated among low-to-mid ESG performers. This result is in line with agency theory. These firms have less mature sustainability practices, and assurance may help them improve their performance by acting on their internal organization. Firms at the top already have a strong internal sustainability organization, and assurance does not bring much to them. For firms at the bottom, however, the absence of an effect could be linked to a lack of internal control and insufficient sustainability maturity: these companies do not have the capacity to make use of assurance. Assurance therefore appears to add value only when the sustainability maturity of the company is high enough, but not so high that the information asymmetry has already been resolved.

Looking at the self-constructed ESG score, the story is different, though not completely contradictory. Assurance is associated with more vague, hedged, and superlative language, which the literature treats as a marker of poor ESG performance and greenwashing. Legitimacy theory offers an explanation: assurance may be adopted for its signaling value. Agency theory complements this: under greater scrutiny, managers are incentivized to use more cautious and vague language to pass the assurance process. Combined with the results found with external ESG grades, this suggests that assurance may improve how a firm is rated rather than its underlying ESG performance. Managerial effort may shift from performance to the report, concentrating on symbolic rather than substantive performance. Part of this divergence may also stem from the score itself, which relies on NLP models

not validated for this population and on translated reports.

Looking at the relationship between assurance and report quality, no robust relationship was found. This may be due to the fact that assurance is, above all, designed to verify that the information is accurate, not to check whether the report is well balanced, complete, or clearly written. It confirms content but does not build the reporting competence that determines quality. The strong effect of report type reinforces this reading: quality is driven primarily by reporting maturity, not by whether a report is verified.

Finally, it is interesting to look at these results in light of the debate between voluntary and mandatory assurance regimes. Overall, this study suggests that compulsion changes how many firms are assured, but not who benefits from assurance, and it does not correct the tendency toward vague, hedged, and superlative language observed for H1. A uniform assurance mandate carries the risk of missing the firms that most need to improve, while imposing high costs on firms that do not need it and extending the incentive toward symbolic compliance. It is important to note that these results are not an argument against assurance. They are an argument against treating assurance as a uniform tool that would benefit every company in the same way.

7. Conclusion

This study finds that external assurance is associated with higher external ESG grades for low-to-mid performers and with more redundant and symbolic language. No robust relationship with report quality was found.

The most important conclusion is that the value of assurance does not lie in the instrument itself but in the organizational maturity of the company that uses it. Assurance can confirm what a firm discloses, but it cannot create the internal systems needed to generate credible information. In the context of the CSRD and the ongoing European debate on assurance requirements, this supports a differentiated, maturity-based approach rather than a uniform mandate.

7.1. Contributions

First, by using external ESG grades as well as a self-constructed NLP-based ESG score, this study shows that the effects of assurance are not evenly distributed among companies. Organizational and reporting maturity appear to be determinants of the usefulness of assurance.

Second, by exploiting the contrast between France and Italy, it suggests that a mandatory regime does not extend the benefits of assurance, only the number of firms that use it.

Third, it builds two instruments that could be reused in other studies: an ESG score and a reporting quality framework.

Finally, it extends assurance research to small and medium-sized Benefit Corporations, a population largely absent from prior studies.

7.2. Implications

For regulators, the evidence suggests that mandatory assurance should be phased in depending on companies' reporting and organizational maturity. Not every company can benefit from assurance. Regulators should first focus on helping companies build strong ESG competences as well as strong non-financial reporting competences. In practice, maturity could be proxied by observable indicators, such as the number of years of sustainability reporting, the existence of internal controls over sustainability data, or a dedicated sustainability function.

For companies, assurance is more useful for low-to-mid performers. Low performers should first focus on building ESG competences and improving their internal organization. High performers gain little from assurance in terms of ESG performance and may be better served by maintaining the internal capacity that secures reporting credibility. All firms should avoid relying on redundant, symbolic language.

For assurance providers, engagements should extend beyond data reliability to the relevance and quality of reported information. Given the association between assurance and vaguer language, providers could also flag vague, hedged, or unsupported claims during the engagement.

7.3. Limitations

The sample is small (188 companies, 211 reports) and composed mainly of SMEs already committed to sustainability. Moreover, because the sample is made up mainly of SMEs, some controls such as revenue, ROA, or leverage were unavailable. The assurance variable does not distinguish OTI verification from limited or reasonable assurance. The external grade averages two providers with different methodologies. The ESG score is not validated against an external benchmark and relies on translated reports, and the quality framework was applied by a single coder. The central role of organizational and reporting maturity is inferred from the pattern of results rather than measured directly. Finally, the inconclusive endogeneity tests mean that no result can be interpreted causally.

7.4. Future research

Future research could examine more closely the link between assurance and boilerplate language, including how much management time is devoted to the assurance process relative to substantive ESG action. It could also measure organizational and reporting maturity directly, for example through surveys or indicators of internal control over sustainability data, to test whether it moderates the association between assurance and ESG performance. The analysis could also be replicated with a single rating provider, additional financial controls, a larger sample, and alternative ESG scores and quality frameworks. Finally, distinguishing between limited and reasonable assurance would clarify whether the level of assurance changes its association with performance and quality.

8. Bibliography

[1] R. Simnett, A. Vanstraelen and W. F. Chua, "Assurance on sustainability reports: An international comparison," *The Accounting Review,* vol. 84, no. 3, pp. 937-967, 2009.

[2] J. Krasodomska, R. Simnett and D. L. Street, "Extended external reporting assurance: Current practices and challenges," *Journal of International Financial Management & Accounting,* vol. 32, no. 1, pp. 104-142, 2021.

[3] M. C. Jensen and W. H. Meckling, "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics,* vol. 3, no. 4, pp. 305-360, 1976.

[4] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica,* vol. 46, no. 1, pp. 33-50, 1978.

[5] A. Lewbel, "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models," *Journal of Business & Economic Statistics,* vol. 30, no. 1, pp. 67-80, 2012.

[6] M. Lang and L. Stice-Lawrence, "Textual analysis and international financial reporting: Large sample evidence," *Journal of Accounting and Economics,* vol. 60, no. 2-3, pp. 110-135, 2015.

9. List of Equations

(4. 1) 3
(4. 2) 3

10. Acknowledgments

I would like to thank Leonardo Boni, Hazal Kopal and Nicholas Algie for their support throughout this work. Their availability, advice and the time they spent discussing my ideas and reviewing my progress were very helpful in developing this thesis. I am also grateful for their patience and encouragement along the way.

Finally, I would like to thank my family and friends for their constant support and encouragement throughout my studies.