

AUDITING THEORY / AUDITING AND ASSURANCE SERVICES

WILLIAM KWESI DONKOH

(UD95680AC104902)

Beyond the Paymaster: A Critical Interrogation of Auditing Theory in the Age of Symbolic Assurance and ESG Complexity

1. Executive Summary & Introduction

The audit function rests on a deceptively simple promise: an independent third party examines evidence and renders a conclusion that reduces information asymmetry between those who manage an entity and those who rely on its representations. Financial auditing has spent decades codifying this promise into standards (ISAs, PCAOB ASs), ethical codes (IESBA), and regulatory oversight (PCAOB, FRC). Yet the expansion of auditing beyond financial statements—into quality management (ISO 9001), environmental performance (ISO 14001), and sustainability reporting (GRI, AA1000)—has exposed a fracture in this theoretical edifice. The three papers comprising this report's core evidence collectively demonstrate that non-financial auditing operates in a space where the traditional principal-agent model collapses under the weight of a single, corrosive fact: the auditee is the paymaster.

Prajogo, Castka, and Searcy (2020) state this bluntly. Firms select their assurance providers. Firms contract with them. Firms pay their fees. This arrangement, which would raise immediate red flags in a financial audit context (imagine a publicly traded company directly hiring and paying its own external financial auditor without any audit committee oversight), remains the default operating model for ISO 9001 certification, sustainability report assurance, and eco-label verification. The consequence, demonstrated empirically across 597 certified firms, is a systematic divergence between

two audit characteristics: strictness (substantive verification of actual practices) and market credibility (the brand reputation of the assurance provider). Firms pursuing external legitimacy select the latter. Those pursuing genuine operational improvement select the former. The two characteristics do not reliably co-occur.

Perego and Kolk (2012) extend this problem to the sustainability domain, tracking Fortune Global 250 firms across a decade. Their longitudinal data reveal a troubling pattern: adoption of sustainability assurance increased significantly (from 21.4% in 1999 to 55.8% in 2008), but the quality of assurance statements improved only marginally (mean content analysis score from 9.72 to 12.93 on a 27-point scale). More concerning, the market for assurance providers fragmented. Accounting firms, specialists, certification bodies, and a residual category called "Others" (including academic institutions) each offered different quality profiles. Accounting firms scored highest (13.71 mean), but still delivered statements capped at 22 points out of 27. The "Others" category, increasingly prevalent in Japan, scored a dismal 5.89. These numbers are not merely descriptive; they represent a failure of the assurance promise. Stakeholders reading a statement from an "Other" provider have no reasonable basis to trust its conclusions.

Boiral, Heras-Saizarbitoria, Brotherton, and Bernard (2018) take us inside the heads of the assurance providers themselves. Through 38 semi-structured interviews, they extract uncomfortable admissions. Forty-five percent of interviewees explicitly acknowledged commercial pressures undermining independence. The same percentage admitted that verification processes were often superficial. Practitioners described a world where limited assurance (the lowest level) dominates, where negative phrasing like "nothing has come to our attention" replaces positive affirmations of accuracy, and where the line between auditing and consulting blurs to the point of invisibility. One respondent recounted a Mexican firm maintaining two separate identities—consultant and auditor—with auditors physically relocating to a Starbucks when an unscheduled client visit threatened to expose the arrangement. These are not edge cases. They are systemic symptoms of a theoretical model disconnected from commercial reality.

The Core Argument

This report advances a specific thesis: while traditional auditing theory assumes a linear, evidence-based path from client engagement to independent assurance to reduced information asymmetry, the proliferation of non-financial auditing reveals a decoupled reality where symbolic legitimacy and market credibility systematically supersede substantive verification. The paymaster relationship does not merely create a *potential* conflict of interest. It creates an *inevitable* structural bias that no amount of professional skepticism—however earnestly applied by individual auditors—can fully

overcome. The only reliable safeguard against this bias would be the removal of provider selection discretion from the auditee. That safeguard is absent across virtually all non-financial assurance contexts.

This argument draws on three theoretical traditions. **Agency theory** (Section 2.1) provides the baseline model of why independent assurance matters in the first place. **Institutional theory and legitimacy theory** (Section 2.3) explain why firms pursue assurance even when they have no intention of substantive implementation—they seek the legitimacy of certification as an end in itself. **The information and insurance hypotheses** (Section 2.2) clarify what assurance promises to deliver (credible information) versus what it often delivers in practice (a liability shield for the provider). The tension between these theoretical lenses is productive; each captures part of the phenomenon, none captures the whole.

What This Report Does Not Do

Before proceeding, a boundary condition. This report does not argue that all non-financial assurance is fraudulent or valueless. The Prajogo et al. (2020) data show that internal motives do lead to strict auditing, and strict auditing does lead to operational benefits (path coefficient 0.24, $p < 0.01$). Some firms use assurance properly. Some providers deliver rigorous audits. The problem is systemic, not universal. But systemic problems require systemic analysis. The existence of good actors does not negate the structural incentives that reward poor actors. The Enron scandal did not arise because Arthur Andersen lacked competent professionals. It arose because the paymaster relationship, in the absence of countervailing structural safeguards, predictably corrupts audit quality over time.

Structure of the Report

Section 2 develops the theoretical frameworks in depth, moving from agency theory through the information and insurance hypotheses to institutional legitimacy. Each framework receives granular attention, with explicit connections to the source papers' findings. Section 3 interrogates the epistemology of audit evidence—what counts as evidence, how strictness differs from market credibility, and how professional skepticism erodes under commercial pressure. Section 4 examines the operational mechanics of risk assessment and materiality, showing how these technical constructs become vehicles for symbolic compliance rather than substantive verification. Section 5 identifies current challenges, focusing on the commodification of assurance and the globalization of lenient practices. Section 6 provides the deep-dive case analysis of all three papers, walking through specific models, tables, and interview excerpts. Section 7 concludes with regulatory implications and a future research agenda.

A Note on Terminology

Throughout this report, "assurance provider," "auditor," and "certification body" are used according to each paper's original terminology. Prajogo et al. (2020) study ISO 9001 certification bodies. Perego and Kolk (2012) study sustainability assurance providers. Boiral et al. (2018) interview both accounting firms and consultants. The differences matter; a Big Four sustainability assurance engagement differs from a small consultancy's ISO certification audit. But the underlying ethical structure—paymaster selects, contracts, and pays provider—remains constant. That structural constant is the report's focus.

The Stakes

Why should a reader care about non-financial audit quality? Two reasons, one immediate and one systemic. First, the immediate: customers pay premiums for certified products and services. If the certification is fake or the audit superficial, those customers have been defrauded, however unintentionally. Heras-Saizarbitoria and Boiral (2019, cited in Prajogo et al., 2020, p. 2) documented the "widespread prevalence of fake ISO 9001 certificates." Each fake certificate represents a real transfer of wealth from a trusting buyer to a non-compliant seller. Second, the systemic: the audit function—financial and non-financial alike—derives its social license from the perception of impartiality. When that perception erodes across any domain, the erosion spreads. Regulators and standard-setters cannot cordon off sustainability assurance from financial audit credibility. The same firms (the Big Four, for instance) provide both services. Scandals in one arena infect trust in the other. The Enron collapse damaged faith in corporate reporting for a decade. A comparable scandal in sustainability assurance—say, a major carbon offset verification that systematically overstates emissions reductions—would have similarly corrosive effects.

The papers analyzed in this report suggest such a scandal is not a matter of if, but when. Boiral et al. (2018, p. 1118) quote one partner: "I think that the quality will continue to decline gradually until a huge scandal occurs and, then, it will be noted that these audits, in fact, do not reflect reality. That's what happened in finance and accounting." The warning is explicit. The question is whether anyone is listening.

2. Theoretical Frameworks of Auditing: From Economic Agency to Institutional Legitimacy

2.1. Agency Theory as the Foundational Lens

Agency theory provides the starting point for any serious discussion of auditing's economic rationale. The formulation is familiar but worth restating with precision. Jensen and Meckling (1976) defined the agency relationship as a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf, which involves delegating some decision-making authority to the agent. The problem arises from two assumptions. First, the principal and agent have divergent utility functions—the agent will not automatically act in the principal's best interests. Second, information asymmetry favors the agent, who possesses superior knowledge of their own efforts, the firm's operations, and the risks being taken. The agent can observe what the principal cannot.

Auditing enters this space as a monitoring mechanism. An independent third party examines evidence of the agent's actions and performance, then reports to the principal. The audit reduces information asymmetry by transforming private agent knowledge into verifiable public information. The principal gains confidence that the agent's representations (financial statements, sustainability reports, quality management claims) correspond to underlying reality. The agent gains credibility, which may lower the principal's required monitoring or reduce the agent's cost of capital. In theory, this is a Pareto-improving arrangement.

Now apply this model to the non-financial auditing context examined by Prajogo et al. (2020). Who is the principal? The conventional answer would be stakeholders: customers

relying on ISO 9001 certification, investors using sustainability reports, regulators monitoring environmental compliance. Who is the agent? Management, who operates the quality system or produces the sustainability data. Who is the auditor? The certification body or assurance provider. The neat three-part structure appears to hold.

It does not hold. The collapse occurs at the funding and selection mechanism. In the canonical agency model, the principal hires the auditor to monitor the agent. The auditor's loyalty runs to the principal; the agent cannot fire the auditor for delivering an unfavorable report. Financial auditing approximates this structure through audit committees, mandatory auditor rotation, and regulatory oversight of auditor changes. Even with these safeguards, the financial audit literature has documented persistent doubts about independence (Bazerman et al., 1997; Moore et al., 2006). The paymaster problem is not unique to non-financial auditing.

But non-financial auditing lacks even these imperfect safeguards. As Prajogo et al. (2020, p. 2) document, "organizations typically have discretion in selecting their assurance provider" and "audits are usually contracted and paid for by the company that wishes to become certified." The principal—the stakeholder who needs reliable information—does not hire the auditor. The agent hires the auditor. The agent pays the auditor. The agent can fire the auditor by simply not renewing the contract. The agent can select a different auditor if the first one proves too strict. The theoretical structure of agency theory has been inverted. The monitor is employed by the monitored.

The consequences of this inversion are not merely theoretical. Prajogo et al. (2020) operationalize them through the distinction between internal and external motives. Internal motives—building a solid quality management system, improving operational control—represent the agent's genuine desire to reduce information asymmetry and improve performance. External motives—meeting customer requirements, enhancing corporate image—represent the agent's desire to signal compliance without necessarily achieving it. When internal motives dominate, firms select assurance providers who offer "audit strictness" (substantive verification). When external motives dominate, firms select providers who offer "market credibility" (brand reputation). The path coefficients from their structural equation model (Figure 1, p. 9) tell a stark story: external motives drive market credibility selection at 0.39 ($p < 0.01$), significantly stronger than their effect on strict auditing selection at 0.23 ($p < 0.01$). The paymaster, given discretion, votes for the signal that maximizes market benefit, not the audit that maximizes operational truth.

This finding aligns with a broader critique of agency theory's applicability to auditing contexts. The theory assumes that contracts can be written to align incentives and that monitoring institutions function independently. But as Boiral et al. (2018) demonstrate through their interview data, independence is not a binary state that exists once a

contract is signed. It is a continuous, fragile accomplishment that commercial pressures systematically erode. Forty-five percent of their respondents mentioned commercial pressures explicitly. One consulting firm auditor stated flatly: "I think that, to begin with, the independence is very limited when you get paid by your client" (Boiral et al., 2018, p. 1117). This is not a practitioner failing to understand agency theory. It is a practitioner accurately describing the structural reality that agency theory papers often gloss over: the monitor cannot be truly independent when the monitored controls the monitor's access to resources.

Agency theory also struggles to account for the phenomena Perego and Kolk (2012) document. Why would a Fortune Global 250 firm obtain sustainability assurance from a provider that scores only 5.89 on a 27-point quality scale (the "Others" category, predominantly academic institutions in Japan)? The agency theory answer—to reduce information asymmetry for principals—makes no sense. A low-quality assurance statement does not credibly reduce information asymmetry. It may even increase it by creating a false appearance of verification. The institutional theory answer, developed in Section 2.3, fits better: the firm needs the *appearance* of independent verification to satisfy coercive pressures (Japanese government guidelines) or normative expectations (industry peers obtaining assurance). The actual quality of the verification is secondary. The principal—the stakeholder reading the report—may not be sophisticated enough to distinguish a high-quality assurance statement from a low-quality one. And the assurance provider, for their part, has no incentive to demand higher quality if the client only wants a signature.

This brings us to a deeper problem with agency theory in the non-financial context: the identity of the principal is contested. In financial auditing, the principal is reasonably clear: shareholders and debt holders who have a direct economic interest in accurate information. In sustainability assurance, who is the principal? The Global Reporting Initiative (GRI) framework lists multiple stakeholders: employees, customers, suppliers, local communities, NGOs, investors, regulators. These groups have different information needs, different capacities to enforce their demands, and different levels of sophistication in reading assurance statements. An investor might understand the difference between reasonable and limited assurance. A local community member likely does not. An NGO might spot a weak assurance statement. A retail customer buying a certified product almost certainly does not. The assurance provider, facing a diffuse and incapable principal, has little accountability to any specific stakeholder group. Their accountability runs, as a practical matter, to the paymaster who signs the check.

Perego and Kolk (2012, p. 185) capture this dynamic in their discussion of "managerial capture." The term refers to situations where the assurance process is controlled by the reporting organization's management, who shape the scope, methods, and conclusions

to serve their own interests rather than stakeholder needs. Smith et al. (2011, cited in Perego & Kolk, 2012, p. 175) develop this concept further, arguing that capture occurs when powerful interests appropriate the language and processes of auditing for commercial objectives. The assurance provider, embedded in a commercial relationship, may not resist this capture. Resistance risks losing the client. And losing clients has tangible consequences for professional service firms operating on thin margins in competitive markets.

The agency theory response to capture is typically structural: strengthen the principal's monitoring, align incentives better, increase transparency. But these responses assume the principal exists as an organized, capable entity. In non-financial auditing, the principal is often a dispersed, unorganized, low-information public. No audit committee represents local communities affected by a mining company's environmental practices. No shareholder vote approves the selection of a sustainability assurance provider. No regulator systematically reviews assurance statement quality across all certified firms. The institutional infrastructure that makes financial auditing functional—however imperfectly—does not exist for non-financial assurance. The paymaster operates in a vacuum of countervailing power.

This vacuum is not accidental. It is the logical consequence of applying a monitoring model originally designed for bilateral principal-agent relationships to a multilateral, public-good context. The firm's sustainability report, once assured, becomes a public document. Anyone can read it. But no one has the standing or resources to challenge its assurance quality except, perhaps, a well-funded NGO or a securities regulator with jurisdiction over misleading disclosures. The former are scarce; the latter have historically focused on financial, not sustainability, misstatements. The result is a market failure that agency theory, focused on bilateral contracting, does not adequately theorize.

We need additional frameworks. The information hypothesis (Section 2.2) and the insurance hypothesis (Section 2.2) address different aspects of what assurance promises to deliver. Institutional and legitimacy theory (Section 2.3) explain why firms pursue assurance even when the economic benefits are questionable. Each framework fills gaps that agency theory leaves open. But none, taken alone, fully explains the patterns observed across the three source papers. The phenomenon of non-financial assurance is theoretically overdetermined—multiple causal pathways lead to the same observed behavior. That is not a weakness in the analysis. It is a reflection of the empirical reality that firms have multiple, sometimes conflicting, motives for seeking assurance. Internal and external motives coexist (Nair & Prajogo, 2009, cited in Prajogo et al., 2020, p. 5). Firms want both operational improvement and market credibility, even though the path coefficients suggest these goals pull toward different provider characteristics. The task

of theory is not to reduce this complexity to a single driver. It is to map the terrain accurately, showing where different mechanisms dominate and how they interact.

2.2. The Information Hypothesis vs. The Insurance Hypothesis

Agency theory explains *why* principals demand monitoring, but it does not fully explain *what* auditors actually provide or *how* that provision creates value. Two competing hypotheses address these gaps. The information hypothesis posits that auditing improves the decision-usefulness of information. Principals receive more reliable data, make better resource allocation decisions, and the overall market for information functions more efficiently. The insurance hypothesis, by contrast, posits that auditing serves primarily as a risk-sharing mechanism. Principals value the audit not because it improves information per se, but because the auditor provides a "deep pocket" that can be sued if the information proves false. The audit's value lies in the assurance provider's reputation and capital, not in the incremental accuracy of the audited information.

These hypotheses are not mutually exclusive. Auditing can simultaneously provide information benefits and insurance benefits. But they have different implications for how we evaluate audit quality, auditor independence, and the paymaster problem. The distinction becomes particularly sharp in non-financial auditing contexts, where standards are voluntary, liability regimes are uncertain, and the "deep pocket" may be shallow or nonexistent.

The Information Hypothesis: Assurance as Signal Improvement

The information hypothesis traces to the accounting literature on voluntary disclosure. Managers possess private information about firm value. They disclose some of that information voluntarily, but investors discount voluntary disclosures because managers have incentives to bias information upward. An independent audit—particularly one conducted by a reputable provider—credibly signals that the disclosed information has been verified against a defined standard. The signal is credible because the auditor's reputation is at stake; a false or misleading audit would damage the auditor's brand and expose them to litigation risk.

Simnett, Vanstraelen, and Chua (2009, cited in Perego & Kolk, 2012, p. 185) apply this logic to sustainability assurance. They find that firms in stakeholder-oriented legal environments (e.g., continental Europe) are more likely to obtain assurance, and more likely to choose accounting firms as assurance providers, because the demand for credible sustainability information is higher in those contexts. The information

hypothesis predicts exactly this pattern: where stakeholders care about sustainability data, firms will invest in higher-quality assurance to differentiate themselves from non-assured competitors.

Prajogo et al. (2020) provide a more granular test of the information hypothesis through their distinction between operational and market benefits. Operational benefits—improved product quality, reduced operating costs—represent the information hypothesis at work. A strict audit verifies that the quality management system actually functions as designed. The firm receives not just a certificate but actionable feedback on system weaknesses. The path coefficient from strict auditing to operational benefits (0.24, $p < 0.01$) supports this prediction. Strict auditing delivers real, measurable improvements in firm operations. The information hypothesis holds in this channel.

But the path from strict auditing to market benefits tells a different story. That coefficient is 0.06 and not statistically significant (Prajogo et al., 2020, p. 9). Strict auditing does not directly improve market outcomes. Market benefits flow instead from the market credibility characteristic of assurance providers (path coefficient 0.12, $p < 0.05$) and from operational benefits indirectly (operational benefits to market benefits at 0.36, $p < 0.01$). This pattern suggests a two-step mechanism: strict auditing improves operations, improved operations eventually yield market rewards. Market credibility, by contrast, provides a direct market signal even without operational improvement.

The information hypothesis alone does not explain why market credibility can operate independently of strictness. If markets were fully informed about audit quality, they would discount signals from high-credibility but low-strictness providers. Customers and investors would recognize that a brand-name assurance provider might be delivering superficial audits. But the data suggest otherwise. External motives drive firms toward market credibility providers, and those providers deliver market benefits despite having no significant effect on operational benefits (the coefficient from market credibility to operational benefits is -0.03, non-significant). This is a failure of the information hypothesis in its strongest form. The market does not perfectly discriminate between substantive and symbolic assurance. Brand reputation substitutes for audit quality in the minds of at least some stakeholders.

Perego and Kolk (2012) provide complementary evidence from the sustainability domain. Their content analysis scores vary systematically by provider type but also by country. Japanese firms, despite high rates of assurance adoption, rely heavily on "Others" category providers (academic institutions, NGOs) that score only 5.89 on the 27-point quality scale. If the information hypothesis held strongly, these low-quality statements would be discounted by stakeholders, reducing any market benefit to Japanese firms from obtaining assurance. Yet Japanese firms continue to obtain

assurance at increasing rates (Table 2, Perego & Kolk, 2012, p. 7). The rational explanation is that stakeholders—perhaps including customers and regulators in Japan—are not penalizing low-quality assurance. They may not be capable of distinguishing quality levels, or they may value the mere presence of an assurance statement regardless of its content.

The information hypothesis requires sophisticated information users who can assess assurance quality. The evidence suggests such sophistication is not universal. As Boiral et al. (2018, p. 1117) note, assurance statements often use technical jargon and negative phrasing ("nothing has come to our attention") that obscures rather than clarifies the scope and limitations of the verification. A stakeholder reading such a statement may come away reassured without understanding that only limited assurance was provided, or that material topics were excluded from the scope. The information hypothesis assumes clarity and transparency. The practice of sustainability assurance delivers ambiguity and liability protection.

The Insurance Hypothesis: Assurance as Risk Transfer

The insurance hypothesis shifts focus from information to liability. Under this view, the primary value of an audit is not the accuracy it provides but the capital it makes available in the event of a misstatement. Investors and other stakeholders can sue the auditor for failing to detect material misrepresentations. The auditor's resources—professional indemnity insurance, firm capital, partner assets—function as a guarantee that stakeholders will recover losses if the information proves false. Auditing is, in this sense, a form of insurance. The premium is the audit fee. The coverage is the auditor's liability exposure.

This hypothesis has been tested primarily in financial auditing contexts. Research suggests that auditor reputation and size (the "Big N" premium) correlate with lower costs of capital, consistent with an insurance effect. Investors value the assurance that a deep-pocketed auditor will be available to compensate them if things go wrong. The auditor's independence and competence remain relevant, but they matter instrumentally—they reduce the probability of a claim. The insurance value is distinct from the information value, though the two are correlated in practice.

Now apply this hypothesis to non-financial auditing. The insurance mechanism largely does not exist. Boiral et al. (2018, p. 1117) document this explicitly. Assurance providers include liability limitation clauses in their statements. The responsibility for the report's content remains with the reporting company. The assurance provider explicitly disclaims responsibility to stakeholders. One respondent quoted in Boiral et al. (2018, p. 1117) described the standard formulation: "responsibility for the verification process is limited

to the managers of the company and not to the stakeholders." This is not insurance. It is the opposite of insurance. The assurance provider shields itself from liability while collecting fees for providing a stamp of approval.

The absence of a functioning insurance mechanism has profound implications for audit quality. In financial auditing, the threat of litigation constrains auditor behavior. Auditors who cut corners risk ruinous lawsuits. Partners who sign off on fraudulent statements may lose their licenses, their savings, and their freedom. The liability regime, though imperfect, creates a countervailing pressure against the paymaster relationship. Even if the client wants a lax audit, the auditor must weigh the short-term benefit of keeping the client against the long-term risk of being sued by stakeholders who relied on the audit.

Non-financial auditing lacks this discipline. As Boiral et al. (2018, p. 1123) note, "the voluntary and relatively unregulated nature of sustainability assurance" means that assurance providers face minimal liability risk. The ISAE 3000 and AA1000 standards do not create private rights of action for stakeholders. A community affected by a mining company's environmental damage cannot sue the assurance provider who verified the company's sustainability report as "accurate" if the report omitted material information about pollution incidents. The assurance provider's statement explicitly says they are not responsible to stakeholders. The courts have not, to date, systematically extended liability to sustainability assurance providers. The result is a moral hazard of significant proportions. Assurance providers can collect fees, provide superficial verification, and face no consequences when the verified information proves false.

Perego and Kolk (2012, p. 184) touch on this issue through their discussion of "capture." When assurance providers are not accountable to stakeholders, they become responsive only to the paymaster. The provider's commercial objectives—retaining the client, selling additional consulting services—dominate any professional obligation to the public. The partner who admitted that "assurance is a supplementary service" and "the actual audit is the big carrot that they wield in order to get the other services" (Boiral et al., 2018, p. 1119) was not describing an anomaly. They were describing the logical consequence of an insurance vacuum. If no one can sue the auditor for a bad audit, the auditor's only incentive is to please the client.

The information and insurance hypotheses, taken together, predict that audit quality should be higher where both mechanisms operate. Information benefits reward high-quality audits through market differentiation. Insurance benefits penalize low-quality audits through liability risk. Financial auditing, however imperfectly, approximates this combination. Non-financial auditing approximates neither. The information benefits are diluted by stakeholder inability to assess quality. The insurance benefits are absent

entirely. What remains is the paymaster relationship in its purest form: a commercial transaction between a firm wanting legitimacy and a provider willing to sell it, unconstrained by either sophisticated demand or credible threat of liability.

The Interaction Between Hypotheses and Provider Selection

The distinction between information and insurance benefits also clarifies why different provider types dominate different market segments. Accounting firms, with their established brands and (in financial audit contexts) deep pockets, might be expected to offer both information and insurance value. Yet even accounting firms, when providing sustainability assurance, deliver statements that cap out at 22 out of 27 quality points (Perego & Kolk, 2012, p. 12). The ceiling is not 27. The maximum possible quality exists. But no provider in the sample consistently achieved it. This suggests a systemic constraint: the paymaster relationship and the absence of liability risk create a race to the middle, not to the top. Providers compete on price and client relationships, not on demonstrable verification rigor.

Specialists (consulting firms) score even lower, at 10.80 mean quality. Certification bodies score similarly to accounting firms (13.50) but with wider variation (standard deviation 6.80 versus 4.25 for accountants). The "Others" category drags the average down dramatically. These differences are not random. They reflect underlying business models. Accounting firms bring financial audit discipline to sustainability work, but their statements remain constrained by legal caution (the "nothing has come to our attention" formulation). Certification bodies focus on management system standards (ISO 14001, ISO 9001) where verification procedures are more established, but they may lack expertise in sustainability reporting's broader stakeholder accountability demands. Specialists may have deep sustainability knowledge but lack audit rigor. The "Others" may have neither.

The information hypothesis would predict that firms seeking to signal high credibility would select the highest-quality providers (accounting firms or certification bodies) and pay for reasonable assurance, not limited. The data show the opposite: limited assurance predominates, and even high-quality providers deliver statements that fall far short of maximum quality. The insurance hypothesis would predict that firms seeking to transfer risk would select deep-pocket providers. But since liability is largely absent, this prediction cannot be tested. What can be tested is whether firms select providers based on the *perceived* insurance value—the brand reputation that might, in a stakeholder's mind, signal deeper pockets or greater credibility. The market credibility coefficient (0.39 from external motives) suggests they do.

The deeper point is that neither hypothesis, alone or together, fully explains the empirical patterns. Something else is at work. That something, developed in Section 2.3, is institutional legitimacy. Firms pursue assurance not because they want better information or because they want to transfer risk. They pursue assurance because their institutional environment—customers, regulators, industry associations, competitors—expects them to. The assurance statement is a ritual that confers legitimacy regardless of its substantive content. The paymaster relationship, far from being a bug in this system, is a feature. It allows firms to purchase legitimacy at the lowest possible cost, selecting providers who will deliver the necessary stamp without inconvenient questions or expensive verification procedures.

2.3. The Institutional & Legitimacy Theory Challenge

Agency theory explains why monitoring matters. The information and insurance hypotheses explain what value monitoring provides. Neither fully explains why firms would seek assurance that delivers no operational benefit, provides no insurance, and signals only a brand name. Yet the evidence across all three papers shows precisely this behavior. Firms select market credibility providers. Those providers deliver no operational improvement. The statements they produce score poorly on quality metrics. Firms continue to hire them. The market does not punish the behavior. Something else must explain this pattern. That something is institutional legitimacy.

Legitimacy theory, rooted in the work of Suchman (1995) and Meyer and Rowan (1977), offers a fundamentally different account of organizational behavior. Organizations do not only pursue economic efficiency. They also pursue social acceptability. They adopt practices, structures, and certifications because those practices are perceived as legitimate within their institutional environment, not because they improve performance. Legitimacy, as Suchman (1995, p. 574) defines it, is "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions." The key phrase is "generalized perception." Legitimacy exists in the eyes of beholders. It does not require that the legitimated practice actually work. It only requires that relevant audiences believe it should exist.

Prajogo et al. (2020, p. 4) apply this logic directly to ISO 9001 certification. They cite Boiral (2003) and Christmann and Taylor (2006) to distinguish substantive implementation (actually building a functioning quality management system) from symbolic implementation (adopting the standard's appearance without its substance). The distinction is not binary. Organizations can fall anywhere along a continuum. But the critical insight is that symbolic implementation can confer legitimacy gains even when

substantive implementation is absent. A customer seeing an ISO 9001 certificate on a supplier's wall may assume quality processes exist, regardless of whether they actually do. The certificate itself—the rational myth—carries legitimacy independent of its referent.

Decoupling as a Normal Organizational Response

The gap between symbolic adoption and substantive implementation is not a failure of individual organizations. It is a structural feature of institutionalized environments. Meyer and Rowan (1977) introduced the concept of decoupling to describe this phenomenon. Organizations adopt formal structures and policies to satisfy external expectations, then decouple those structures from day-to-day operations to maintain efficiency. The quality manual sits on a shelf. The environmental policy is posted on a wall. Employees continue working as they always have. The decoupling persists because external audiences rarely inspect operations closely. They accept the formal structures as sufficient evidence of compliance.

Perego and Kolk (2012, p. 174) explicitly invoke decoupling in their analysis of sustainability assurance. They cite Boiral and Gendron (2011) on "rational myths" and note that sustainability assurance practices are often "decoupled from organizational practices." The assurance statement may claim that the report has been verified against GRI guidelines and AA1000 principles. Yet the verification may have been limited in scope, conducted by a provider with minimal competence, and concluded with negative phrasing that provides no positive assurance. The statement is decoupled from the actual sustainability performance it purports to verify. The firm obtains legitimacy without substantive accountability.

Boiral et al. (2018) provide the most granular evidence of decoupling in practice. Their interview data reveal how assurance providers rationalize superficial verification. One respondent explained that the quality and scope of the audit depend on the client: "the reporting company largely determines the scope and methods of the audit" (Boiral et al., 2018, p. 1117). Another noted that "the technical jargon used by assurance providers, including the proliferation of negative formulae such as 'nothing has come to our attention,' can be confusing and even misleading for stakeholders" (p. 1117). A partner with a consulting firm offered a striking analogy: "On the ethical level, there is a real problem because the report indicates that it has been assured and verified [...] but only a part of the report has actually been verified [...] Would it be possible to have a financial audit based only on a small department and leaving the rest aside just because we think it is not important?" (p. 1118). The rhetorical question answers itself. No financial auditor would accept such a scope limitation. Yet sustainability assurance providers routinely do.

Why does decoupling persist? Because no one with power to change the system has an incentive to enforce coupling. Assurance providers benefit from the ambiguity. They collect fees without assuming liability. Clients benefit from the legitimacy without incurring the cost of substantive implementation. Stakeholders, in principle the losers from decoupling, lack the information, sophistication, and standing to detect or remedy the gap. A customer purchasing an ISO 9001-certified product cannot easily determine whether the certification reflects a functioning quality system or a ceremonial manual. A reader of a sustainability report cannot verify whether the assurance statement's claims match the actual verification procedures. The decoupling is invisible to all but the most determined and expert investigators.

Legitimacy as a Driver of Assurance Adoption

Legitimacy theory predicts that organizations will adopt practices that enhance their perceived appropriateness, regardless of technical efficiency. The adoption of ISO 9001 certification, sustainability reporting, and assurance services fits this pattern. Prajogo et al. (2020) operationalize legitimacy-seeking through their external motives construct. External motives include meeting customer requirements, enhancing company image and reputation, and responding to market pressures. These are legitimacy concerns, not efficiency concerns. The path coefficient from external motives to market credibility provider selection (0.39, $p < 0.01$) substantially exceeds the coefficient to strict auditing selection (0.23, $p < 0.01$). Firms seeking legitimacy preferentially select providers offering brand reputation over providers offering rigorous verification.

The country-level patterns in Perego and Kolk (2012) reinforce this interpretation. Japan, with its high rate of assurance adoption but low-quality "Others" providers, exemplifies legitimacy-driven behavior. Japanese firms faced coercive pressure from government guidelines on environmental reporting. They responded by obtaining assurance statements. They did not necessarily respond by demanding high-quality assurance. The presence of a statement, any statement, satisfied the regulatory expectation. The quality of the statement was secondary. The legitimacy gain from having any assurance exceeded the gain from having high-quality assurance, especially if high-quality assurance cost more or required more organizational disruption.

France provides a complementary case. French legislation mandated environmental and social reporting for publicly traded firms, effective for the 2005 reporting cycle. Perego and Kolk (2012, Table 2, p. 7) show that French assurance statements increased from 5 in 2002 to 17 in 2005, then to 18 in 2008. The coercive pressure drove adoption. But did it drive quality? The content analysis scores for France show modest improvement over time, from 11.40 in 2005 to 12.75 in 2008 (Table 10, p. 11). These scores are lower than the United Kingdom (18.57 in 2008) and Germany (14.90 in 2008). Coercive pressure

drove quantity, not quality. Firms did what was necessary to comply with the law, not what was necessary to produce credible, informative assurance statements.

The United States presents the opposite pattern. Low assurance adoption (only 12 statements across the entire panel, 11.1% of reports) but relatively high quality when assurance is obtained (mean score 9.83, though with high variance). The litigious US environment, as Perego and Kolk (2012, p. 184) note, may deter firms from obtaining assurance unless they are willing to invest in high-quality verification that would withstand legal scrutiny. A superficial assurance statement, in the US context, might create more liability risk than it mitigates. Better to have no statement than a weak one that plaintiffs could use against the firm. This is legitimacy of a different kind—legitimacy through avoidance of legal exposure rather than through conformity to stakeholder expectations.

Three Mechanisms of Institutional Pressure

DiMaggio and Powell (1983) identified three mechanisms through which institutional pressures produce isomorphism: coercive, mimetic, and normative. Each appears in the source papers.

Coercive pressures arise from formal authority, legal requirements, or resource dependence. The French legislation on sustainability reporting is coercive. The Japanese government guidelines on environmental reporting, though not strictly mandatory, carried sufficient authority to shape firm behavior. Perego and Kolk (2012, p. 184) note that "the promulgation of a more stringent legislation on social and environmental reporting increased regulatory pressure" in Japan and France. Coercive pressure drove adoption rates up. It did not necessarily drive quality up, as the content analysis scores show.

Mimetic pressures arise from uncertainty. When goals are ambiguous or cause-effect relationships unclear, organizations imitate successful peers. The adoption of ISO 9001 certification across supply chains, documented by Corbett (2006, cited in Prajogo et al., 2020), exemplifies mimetic pressure. A firm's major customers are certified. Its competitors are certified. The firm adopts certification not because it has independently verified the standard's benefits but because the institutional environment has made certification the taken-for-granted way of doing business. The "institutionalization of ISO 9001" that Lo and Yeung (2018, cited in Prajogo et al., 2020, p. 11) describe—where customers take the certificate for granted—reflects mimetic pressure operating over time. Once a practice becomes institutionalized, firms adopt it without question.

Normative pressures arise from professionalization. Professional standards, training, and networks create shared expectations about appropriate practice. In sustainability

assurance, normative pressure comes from two directions. Accounting firms bring the norms of financial auditing—independence, professional skepticism, evidence-based conclusions—into the sustainability domain. Perego and Kolk (2012, p. 184) describe this as "a process of normative isomorphism, implying a transposition of professional rules and procedures from financial into non-financial auditing." The AA1000 standard, developed by AccountAbility, represents a different normative source, emphasizing stakeholder engagement and inclusivity. Assurance providers choose which normative framework to follow, or combine elements from both.

The interaction among these pressures creates the complex adoption patterns Perego and Kolk (2012) document. Early adopters (the 12 consistent firms that obtained assurance in all four periods) were likely responding to a combination of normative pressure (professional commitment to transparency) and mimetic pressure (industry leadership position). Late adopters and laggards may have responded primarily to coercive pressure (regulatory requirements) or mimetic pressure (catching up with peers). The "non-adopters" (118 firms, more than half the panel) either faced no coercive pressure, operated in institutional environments where assurance was not normative, or calculated that the costs of assurance exceeded the legitimacy benefits.

Symbolic vs. Substantive Legitimacy

The legitimacy literature distinguishes between symbolic and substantive legitimacy. Symbolic legitimacy arises from the appearance of conformity. The organization adopts the right structures, displays the right certificates, issues the right reports. Substantive legitimacy arises from actual performance. The organization reduces its environmental footprint, improves product quality, treats stakeholders fairly. The distinction maps directly onto Prajogo et al.'s (2020) internal/external motives and strictness/credibility characteristics. Internal motives and strict auditing aim for substantive legitimacy—deserving the certificate by actually meeting the standard's requirements. External motives and market credibility aim for symbolic legitimacy—appearing to deserve the certificate regardless of actual performance.

The problem, from an ethical perspective, is that symbolic legitimacy can be purchased. A firm can select a market credibility provider, pay the fee, receive the certificate, and display it to stakeholders. The provider's brand reputation confers legitimacy even if the underlying audit was superficial. Stakeholders, unable to observe audit quality directly, rely on the provider's brand as a heuristic. The heuristic works when brand reputation correlates with audit quality. The evidence suggests the correlation is weaker than stakeholders assume. Market credibility providers in the Prajogo et al. (2020) sample delivered no operational benefit. Their brand did not predict audit strictness. Yet the market rewarded them anyway.

Perego and Kolk (2012, p. 177) express concern about this dynamic: "the credibility of the certification is therefore linked to the reputation of the assurance provider, which is not necessarily tied to the quality of its auditing services." The gap between reputation and quality is the central ethical vulnerability of the paymaster system. Firms can capture the legitimacy benefits of assurance without incurring the costs of substantive implementation. They merely need to select the right provider—the one whose brand signal overwhelms any inquiry into actual audit rigor.

Boiral et al. (2018, p. 1120) provide a final piece of evidence. They document the "techniques of neutralization" that assurance providers use to justify ethically questionable behavior. Familiarity with clients is justified as necessary for organizational learning. Superficial verification is justified as meeting client requests. Overlap between auditing and consulting is justified as serving client needs. The neutralizations are not lies. They contain partial truths. Familiarity does enable learning. Clients do request limited scopes. Consulting does serve client needs. But the partial truths obscure a deeper reality: the assurance provider has structurally aligned interests with the paymaster, not with the stakeholders who rely on the assurance statement. Legitimacy theory explains why this alignment persists. The institutional environment rewards the appearance of verification, not its substance. Until that changes—until stakeholders demand and can enforce substantive assurance—the paymaster will continue to select providers who deliver exactly what they pay for: a certificate, not a truth.

3. The Epistemology of Evidence and the Meaning of Assurance

Theoretical frameworks explain why firms seek assurance and why the paymaster relationship creates conflicts. But a deeper question remains: what counts as evidence in a non-financial audit, and how does that evidence transform into assurance? Epistemology—the theory of knowledge—rarely appears in auditing textbooks. Yet every audit is an epistemological exercise. The auditor gathers evidence, evaluates its reliability, and draws conclusions about the correspondence between representations and reality. The quality of the audit depends on the quality of this epistemological chain. The paymaster relationship corrupts the chain at every link.

3.1. Substantive vs. Symbolic Auditing: A Conceptual Continuum

Prajogo et al. (2020) operationalize audit quality through two characteristics: strictness and market credibility. These are not opposite ends of a single dimension. They are conceptually distinct attributes that can and do vary independently. An assurance provider can be both strict and reputable. A provider can be reputable but not strict. A provider can be strict but unknown. A provider can be neither. The empirical distribution across 597 firms shows all four combinations exist. The question is which combination dominates under different motivational conditions.

Substantive auditing, proxied by the strictness characteristic, involves rigorous verification of actual practices against a reference point's requirements. An ISO 9001 audit that is substantive would include: review of documented procedures, observation of operations, interviews with employees at multiple levels, sampling of records, testing of outputs, and follow-up on previous non-conformities. The auditor would not accept a quality manual as sufficient evidence. They would verify that the manual reflects actual practice. They would probe for gaps between rhetoric and reality. They would issue non-conformities when gaps exceed materiality thresholds. A substantive audit provides reasonable assurance that the certified system actually functions as described.

Symbolic auditing, proxied by market credibility in the absence of strictness, involves verification that is superficial, ceremonial, or outright fraudulent. A symbolic ISO 9001 audit might include: review of documentation without site observation, interviews only with management, sampling limited to favorable records, no testing of outputs, and acceptance of corrective action plans without verification. The auditor may be competent but constrained by commercial pressures. Or the auditor may be incompetent, lacking the skills to detect non-conformities. Or the auditor may be complicit, knowingly issuing certificates to undeserving firms. The spectrum from superficial to fraudulent is continuous. All are symbolic in the sense that the certificate signals compliance that does not exist substantively.

The distinction matters because stakeholders rely on certification as if it provided substantive assurance. When a customer buys an ISO 9001-certified product, they assume the supplier has a functioning quality management system. When an investor reads a sustainability report with an assurance statement, they assume the information has been independently verified. Symbolic auditing violates these assumptions. The certificate or statement provides a false signal. The stakeholder is misled. The market for certification becomes a market for lemons, where low-quality providers drive out high-quality providers because customers cannot distinguish between them.

Prajogo et al. (2020, p. 11) demonstrate the economic consequences of this distinction through their structural equation model. Operational benefits flow only from strict auditing (0.24, $p < 0.01$). Market credibility alone delivers no operational improvement.

Firms that select market credibility providers for symbolic reasons receive exactly what they pay for: a signal, not substance. The signal may be sufficient for legitimacy purposes if stakeholders do not investigate further. But the signal has no technical content. The firm's quality management system does not improve. The customer's product quality does not increase. The certificate is a simulacrum—a copy without an original.

3.2. Levels of Assurance and the Problem of Scope

The assurance literature distinguishes between reasonable assurance and limited assurance. Reasonable assurance (sometimes called "positive assurance") means the auditor has performed sufficient procedures to conclude that the subject matter is free from material misstatement. The conclusion is expressed positively: "In our opinion, the sustainability report presents fairly, in all material respects..." Limited assurance (sometimes called "negative assurance") means the auditor has performed fewer procedures and concludes only that nothing has come to their attention indicating material misstatement. The conclusion is expressed negatively: "Nothing has come to our attention that causes us to believe that the sustainability report is not presented fairly..."

The difference is not merely semantic. Reasonable assurance requires more evidence, more testing, more skepticism, more documentation. It requires the auditor to actively affirm the report's accuracy. Limited assurance requires only that the auditor report what they did not find. The bar is lower. The cost is lower. The legal exposure is lower. It is unsurprising, then, that limited assurance predominates in sustainability reporting. Perego and Kolk (2012, p. 177) note that "a high or reasonable level of assurance for the whole sustainability report is very rarely issued, as opposed to a moderate or limited level of assurance." Their content analysis does not directly track assurance level, but the predominance of limited assurance is well-documented elsewhere.

Boiral et al. (2018, p. 1117) provide practitioner testimony on the implications. One respondent explained: "Negative statements such as 'we did not see anything that would cause us not to believe that...' are used because it's only a moderate level of assurance. I think that, in some way, it degrades the value of the assurance statement for a lot of users of those reports, because it's almost like saying: Well, maybe the auditors didn't do enough to find out if there was a problem, so what's the value of the assurance?" The respondent, a former auditor, answers their own question: the value is limited. Stakeholders who understand the distinction between reasonable and limited assurance will discount limited assurance statements accordingly. Stakeholders who do not understand the distinction may be systematically misled.

The problem of scope compounds the problem of level. An assurance engagement can be scoped to cover only part of a report. The auditor might verify financial indicators but not social indicators. They might verify greenhouse gas emissions from direct operations but not from the supply chain. They might verify statements about environmental policy but not statements about compliance with specific regulations. The scope is defined in the engagement letter, not always clearly disclosed in the assurance statement. A stakeholder reading only the statement may not realize that material portions of the report were not verified at all.

Boiral et al. (2018, p. 1118) quote a partner who found this practice indefensible: "the report indicates that it has been assured and verified [...] but only a part of the report has actually been verified [...] There are no specific rules requiring us to verify the entire report. However, would it be possible to have a financial audit based only on a small department and leaving the rest aside just because we think it is not important?" The financial audit analogy cuts both ways. In financial auditing, scope limitations that exclude material accounts would result in a disclaimer of opinion or a qualified opinion. The auditor could not issue an unqualified opinion while disclosing in small print that half the balance sheet was not audited. Yet sustainability assurance routinely accepts such limitations without equivalent qualification. The stakeholder sees an assurance statement and assumes full coverage. The assumption is often false.

3.3. Professional Skepticism vs. Managerial Capture

Professional skepticism, as defined in ISA 200, includes a questioning mind, alertness to conditions that may indicate possible misstatement, and a critical assessment of audit evidence. The auditor does not assume management is honest. The auditor does not assume management is dishonest. The auditor gathers evidence and evaluates it without bias. Skepticism is an attitude, not a procedure. It informs every aspect of the audit, from planning through evidence gathering to conclusion.

The paymaster relationship is structurally corrosive to skepticism. An auditor who questions management too aggressively risks losing the client. An auditor who demands access to uncomfortable evidence may be told the evidence does not exist. An auditor who issues non-conformities may find their firm not invited to bid on next year's contract. The threat need not be explicit. Auditors internalize commercial pressures. They learn, through socialization, which questions are safe and which are dangerous. They learn to scope their procedures to avoid inconvenient discoveries. They learn to write statements that protect the firm's relationship with the client.

Boiral et al. (2018) document this internalization extensively. Forty-five percent of respondents explicitly mentioned commercial pressures. One consulting firm auditor noted: "If it's really a client that the auditing company is dependent on for income, it might be subconsciously more supportive of the organisation because they don't really want to get them offside" (p. 1117). The word "subconsciously" is revealing. The auditor does not consciously decide to compromise their principles. The pressure operates below the level of deliberate choice. The auditor's perceptions, judgments, and conclusions are shaped by a background awareness of the client's importance to the firm's revenue. The resulting bias is not intentional, which makes it harder to detect and correct.

Managerial capture, as Perego and Kolk (2012, p. 185) describe it, extends beyond individual auditor bias to the structure of the engagement. Management controls access to information. Management defines the scope of the audit. Management selects the assurance standard. Management chooses the provider. The auditor works within constraints set by management. Even the most skeptical auditor cannot verify information they cannot access. Even the most rigorous auditor cannot expand the scope beyond what management has agreed to pay for. The capture is built into the contract.

Smith et al. (2011, cited in Perego & Kolk, 2012, p. 175) theorize capture as occurring when "senior company management or professional associations tend to take control of or 'capture' sustainability policy and practices by appropriating the language and processes of traditional financial auditing in order to meet their own commercial and professional objectives." The language of auditing—independence, materiality, verification—is appropriated and redeployed for legitimating purposes. The processes are followed ceremonially. The substance is hollow. But the appearance remains.

The Boiral et al. (2018) interview data provide a concrete example of how capture operates at the procedural level. One respondent explained: "We can encourage firms to release material information, but then we face the most critical and global problem of our profession: we are auditors but at the same time there is a client-provider relationship and it can be difficult to manage" (p. 1117). The respondent, an auditor with a consulting firm, acknowledges the difficulty but offers no solution. There is no solution within the current structural arrangement. The only solutions are structural: removing the client's discretion over provider selection, mandating provider rotation, prohibiting the joint provision of consulting and assurance, establishing regulatory oversight of assurance quality. None of these solutions are currently implemented at scale.

3.4. The "Pick and Mix" Approach to Assurance Standards

Perego and Kolk (2012, Table 7, p. 10) document the proliferation of assurance standards and combinations thereof. In 2005, 68.9% of assurance statements in their sample mentioned no standard at all. By 2008, that proportion had dropped to 48.8%—still nearly half. Among statements that did mention standards, the combinations varied widely: AA1000AS only, ISAE3000 only, GRI only, AA1000AS+GRI, ISAE3000+GRI, AA1000AS+ISAE3000, and all three together. No single standard or combination dominated.

This proliferation is not a sign of a mature, professionalized market. It is a sign of fragmentation and confusion. Assurance providers select standards (or combinations thereof) that suit their capabilities and their clients' preferences. An accounting firm familiar with ISAE3000 will use that standard. A consulting firm steeped in stakeholder engagement will use AA1000AS. A provider wanting to appear comprehensive will reference all three. The stakeholder reading the assurance statement sees a list of standards and assumes the verification was conducted according to rigorous, recognized criteria. The assumption may be unwarranted. Referencing a standard does not guarantee compliance with its requirements. An auditor can mention ISAE3000 in their statement without having followed its guidance for scoping, evidence gathering, documentation, and reporting.

Boiral et al. (2018, p. 1118) describe the consequences of this fragmentation: "The confusion arising from the plurality of standards undermines the credibility of assurance among stakeholders." The stakeholder cannot know which standard implies what level of rigor. They cannot know whether a provider using AA1000AS+ISAE3000+GRI is more thorough than a provider using GRI only. The proliferation creates opacity. Opacity benefits providers, who can claim compliance with recognized standards without meaningful verification. Opacity harms stakeholders, who cannot assess the quality of the assurance they receive.

The fragmentation also enables a form of standard shopping. A firm wanting a superficial audit can select a provider known to use low-rigor standards or combinations. A firm wanting a rigorous audit can select a provider with a reputation for thorough application of ISAE3000. But the market does not reliably signal which providers deliver rigor. The brand-name accounting firms, as Perego and Kolk (2012, Table 12, p. 12) show, deliver higher average quality than specialists or "Others." But the variance is substantial (standard deviation 4.25). A stakeholder cannot assume that any given Big Four assurance statement is high quality. Some are. Some are not. The statement itself provides insufficient information to distinguish them.

The epistemological conclusion is uncomfortable but unavoidable. The stakeholder reading an assurance statement cannot know, with confidence, what the statement

means. They cannot know whether the audit was substantive or symbolic. They cannot know whether reasonable or limited assurance was provided. They cannot know whether material topics were scoped out. They cannot know whether the standards referenced were actually followed. They cannot know whether the auditor maintained professional skepticism or succumbed to commercial pressure. The statement provides a signal, but the signal-to-noise ratio is low. The assurance market, as currently structured, produces more noise than signal. The paymaster relationship is the source of the noise. Until the structure changes, stakeholders should treat non-financial assurance statements with profound, justified skepticism.

4. The Mechanics of Risk, Materiality, and the Dilution of Professional Judgment

Auditing standards across both financial and non-financial domains rest on three technical pillars: risk assessment, materiality, and professional judgment. These constructs are meant to translate broad ethical principles (independence, skepticism, due care) into operational procedures. An auditor assesses the risk of material misstatement, determines a materiality threshold, and exercises judgment in designing procedures and evaluating evidence. In theory, these mechanics constrain auditor behavior and ensure consistent quality across engagements. In practice, as the source papers demonstrate, the paymaster relationship systematically dilutes each pillar.

4.1. The Audit Risk Model Reconfigured for Non-Financial Data

The standard audit risk model expresses the relationship as: $\text{Audit Risk} = \text{Inherent Risk} \times \text{Control Risk} \times \text{Detection Risk}$. Inherent risk is the susceptibility of an assertion to material misstatement before considering internal controls. Control risk is the risk that internal controls will not prevent or detect material misstatements. Detection risk is the risk that the auditor's procedures will not detect a material misstatement that exists. The auditor sets detection risk based on assessed inherent and control risks, then designs procedures to reduce detection risk to an acceptable level.

Apply this model to an ISO 9001 certification audit. Inherent risk for quality management assertions is moderate to high. A firm could have poorly documented processes, inadequately trained employees, or inconsistent application of procedures.

Control risk depends on the firm's internal quality management system—its documentation, training, monitoring, and corrective action processes. A firm with a mature quality culture will have lower control risk. Detection risk is the auditor's responsibility. The auditor must design sampling, testing, and observation procedures sufficient to detect non-conformities if they exist.

The paymaster relationship inflates detection risk in three ways. First, the auditor may scope procedures narrowly to reduce cost and maintain client satisfaction. A smaller sample size, fewer site visits, less documentation review—each reduction increases the probability that existing non-conformities will remain undetected. Second, the auditor may accept weak evidence as sufficient. A documented procedure may be accepted as evidence of implementation, even when observation would reveal the procedure is not followed. Third, the auditor may fail to pursue indications of non-conformity. Anomalies may be explained away rather than investigated. Each of these responses is rational for an auditor who prioritizes client retention over verification rigor. Each is irrational from the perspective of the stakeholder who relies on the audit.

Prajogo et al. (2020) do not directly measure risk assessment practices, but their findings imply differential risk calibration across provider types. Market credibility providers, selected by firms with external motives, deliver no operational benefit. Their audits, on average, fail to detect non-conformities that matter for performance. This suggests that detection risk is systematically higher for market credibility providers than for strict auditing providers. The brand-name provider may be more concerned with maintaining the client relationship than with performing rigorous procedures. The strict auditing provider, by contrast, may be willing to accept lower margins or higher client turnover in exchange for professional integrity. The market does not reward the latter behavior, which is why firms with external motives avoid strict providers.

Boiral et al. (2018, p. 1118) provide qualitative evidence of detection risk inflation. One respondent noted: "You actually have to be fairly social, especially for those who are doing it for the first time. They are scared and a bit ambivalent, so you need to be there and say: Let's do this together!" The language of collaboration and reassurance is not the language of skeptical inquiry. The auditor positions themselves as a coach, not an inspector. This role blurring increases detection risk. The auditor who is "doing it together" with the client is less likely to identify and report non-conformities. The relationship has become collegial, not adversarial. Professional skepticism requires a degree of adversarial distance that the collaborative model erodes.

4.2. Materiality as a Contested Terrain

Materiality is the threshold at which omitted or misstated information would influence the decisions of a reasonable user. In financial auditing, materiality is quantified as a percentage of net income, total assets, or revenue. The thresholds are not universal, but they are anchored in financial metrics that provide comparability across engagements. An auditor who set materiality at 10% of net income would be expected to justify why 10% rather than 5% or 2%. The concept is contestable but disciplined.

Non-financial auditing lacks comparable anchors. What is material in a sustainability report? A stakeholder-oriented definition might include any information that would affect a stakeholder's assessment of the firm's social or environmental impact. But which stakeholders? Employees, local communities, NGOs, investors, regulators, customers—each may have different materiality thresholds. The AA1000 standard emphasizes stakeholder inclusivity as the basis for materiality determination. The GRI guidelines require reporting organizations to explain their materiality determination process. But these requirements are procedural, not substantive. They do not prescribe quantitative thresholds. They do not require external verification of the materiality determination itself.

Perego and Kolk's (2012) content analysis coding framework (Appendix, p. 15-16) includes a three-level materiality scale:

- 0: No reference
- 1: Reference limited to a broad statement (e.g., "covers all material aspects") or a negative statement that no work was done to confirm material issues
- 2: Reference and explanation of materiality setting, or broad statement plus stakeholder perspective introduced (e.g., "issues material to stakeholders have been considered")
- 3: Reference, explanation, and stakeholder perspective introduced

The progression from level 1 to level 3 represents increasing transparency about how materiality was determined. A level 3 statement would explain the process for identifying material topics, the criteria used to prioritize them, and how stakeholder input was incorporated. A level 1 statement would merely assert that material aspects were covered, without explanation.

Across the 180 assurance statements in their sample, the average score on this item (and the related completeness and responsiveness items) was low. Boiral et al. (2018, p. 1120) confirm this pattern, noting that "when it comes to formulating a recommendation and providing positive assurance, Specialists and Certification bodies assure that to be more elaborate and informative." The implication is that assurance providers avoid explicit, positive statements about materiality. They prefer broad

assertions that cannot be falsified. A statement that the report "covers all material aspects" is unfalsifiable because the reader does not know what the auditor considered material. The statement provides reassurance without accountability.

The problem of materiality is not merely technical. It is ethical. When an assurance provider accepts a client's narrow materiality determination without scrutiny, they are complicit in omitting potentially significant information from the report. A mining company might determine that environmental impacts on local water sources are not material because they do not affect financial performance. An assurance provider who accepts this determination without question allows the omission to stand. Stakeholders who rely on the assurance statement assume the report covers what matters. Their assumption is false. The auditor has participated in the deception, passively if not actively.

Prajogo et al. (2020, p. 11) touch on materiality indirectly through their discussion of "performance portrayal gaps" (Adams, 2004). A gap arises when organizations that fail to meet reference point requirements are formally recognized, resulting in a positive portrayal of performance that is undeserved. The certification of undeserving firms creates a materiality problem at the market level. The signal from certification is no longer reliable. Customers cannot distinguish between firms that meet the standard and firms that do not. The information asymmetry that certification was meant to reduce has been replaced by a different asymmetry: customers know that some certified firms are undeserving but cannot identify which ones.

4.3. The Selection of Assurance Providers as a Risk Management Strategy

Firms facing uncertainty about audit outcomes can manage that uncertainty through provider selection. This is the insight at the core of Prajogo et al.'s (2020) H1 and H2. Firms with internal motives (seeking genuine quality improvement) select strict auditing providers. Firms with external motives (seeking market legitimacy) select market credibility providers. The selection is not random. It is a strategic response to the risks inherent in the audit process.

What are those risks? For a firm with internal motives, the risk is that a superficial audit will fail to identify non-conformities, leaving the quality management system weak and operational benefits unrealized. The firm therefore selects a strict provider, accepting that the audit may be more expensive, more time-consuming, and more likely to produce non-conformities. The trade-off is acceptable because the firm values the operational benefits that strict auditing delivers.

For a firm with external motives, the risks are different. The primary risk is not operational failure but signal failure. The firm needs the certificate to maintain customer relationships and market access. A failed audit (non-certification) would be disastrous. A strict audit that identifies numerous non-conformities might delay certification or result in qualified certification. The firm therefore selects a provider that balances market credibility with leniency. The provider's brand satisfies customer expectations. The provider's audit practices minimize the risk of non-certification. The firm achieves its objective—certification—without the operational improvements that strict auditing would require.

This is risk management, not fraud. The firm is not necessarily asking for a fraudulent certificate. They are selecting a provider whose reputation and practices are aligned with the firm's risk profile. The provider, in turn, understands the implicit bargain. They will conduct an audit that is technically compliant with standards but calibrated to client expectations. The audit will not be fraudulent—there will be some evidence gathering, some testing, some documentation. But it will not be rigorous. It will not detect non-conformities that the client does not want detected. The audit will deliver the certificate. That is what the client paid for.

Boiral et al. (2018, p. 1116) document the provider's side of this bargain. One respondent explained: "There are many auditors overlooking important issues just to please their clients and sometimes advising them to add things to improve corporate image. In the end, this is not necessarily greenwashing but... this is very similar to it" (p. 1117). The respondent struggles to distinguish legitimate client service from unethical behavior. The line is blurry. Advising a client on how to improve their report could be value-adding. Advising a client on how to appear compliant without actually becoming compliant crosses the line. Where exactly the line sits depends on judgment. The paymaster relationship pressures judgment toward the client's interests.

The path coefficients from Prajogo et al. (2020, Fig. 1, p. 9) quantify the risk management calculus. External motives drive selection of market credibility providers at 0.39 ($p < 0.01$). Internal motives drive selection of strict auditing providers at 0.24 ($p < 0.01$). The difference between these coefficients (0.15) represents the strength of the preference alignment. Firms with external motives are more than 1.6 times as likely to prioritize market credibility over strictness, compared to the baseline effect of internal motives. The paymaster uses their discretion to select providers who will deliver the outcome they desire.

The ethical implication is uncomfortable. The paymaster system does not merely permit firms to select lenient providers. It incentivizes them to do so. A firm that selected a strict provider for a certification audit would incur higher costs, face higher risk of non-

certification or qualification, and receive no additional market benefit because customers cannot distinguish strict audits from lenient ones. The rational choice—the choice that maximizes expected utility for the firm's managers—is to select the least strict provider that carries sufficient market credibility to satisfy customers. The market selects against rigor. This is not a failure of individual ethics. It is a failure of market design. The paymaster relationship, combined with stakeholder inability to assess audit quality, creates a classic adverse selection spiral. Strict providers lose market share to lenient providers. Lenient providers proliferate. Audit quality declines. The spiral continues until a scandal forces structural change.

5. Current Challenges: The Commodification, Globalization, and Regulatory Gaps in Non-Financial Assurance

The theoretical frameworks developed in Sections 2 through 4 reveal a consistent pattern: the paymaster relationship, combined with stakeholder inability to assess audit quality, creates systematic incentives for symbolic over substantive assurance. This section synthesizes the empirical evidence from all three papers on three interconnected challenges that flow from this pattern: the commodification of assurance services, the globalization of lenient practices through cross-border accreditation shopping, and the regulatory gaps that allow these practices to persist.

5.1. The Commodification of Assurance Services

Assurance, in its ideal form, is a professional service governed by ethical principles that prioritize the public interest over commercial gain. The auditor serves stakeholders, not the client. The audit fee compensates for professional work but does not buy a particular outcome. This ideal is never fully realized in any domain, but the distance between ideal and reality is particularly vast in non-financial assurance.

Boiral et al. (2018, p. 1116) provide direct evidence of commodification. One of their key findings, reflected in Figure 2 (p. 1120), shows that commercial pressures were mentioned by 45% of interviewees as a significant ethical issue. These pressures manifest in three ways, as documented on page 1116: negotiation of assurance contracts with pressure to lower prices, dependence on important clients for economic

survival, and the shaping of assurance statements to avoid damaging client relationships.

The commodification dynamic extends beyond price competition to the very structure of assurance engagements. As one respondent quoted by Boiral et al. (2018, p. 1119) explained: "One of the problems is that, for the Big Four accounting firms, assurance is a supplementary service and not their big business [...] So the actual audit is the big carrot that they wield in order to get the other services in there to make money. In terms of total revenue, audits don't make a profit. The consulting and advisory services are the ones that make a profit." The assurance engagement is not a standalone professional service. It is a loss leader. The firm provides assurance at low margins, or even at a loss, to establish a relationship that will generate profitable consulting revenue.

This business model creates a fundamental conflict of interest. The assurance provider who views the audit as a "carrot" cannot afford to antagonize the client. A strict audit that identifies numerous non-conformities or qualifies the assurance conclusion would jeopardize not only the assurance fee but the much larger consulting opportunity. The rational response, from a commercial perspective, is to keep the client satisfied. Keep the audit superficial. Issue the certificate. Move on to selling consulting services.

The Mexican case described by Boiral et al. (2018, p. 1119) represents the extreme endpoint of this logic. A former consultant recalled: "There were two different phone numbers and addresses in our firm. For example, I could work on the sustainability report or the integrated report of one of our clients, and when the client asked our advice about the assurance of the report, we had to answer that we knew a very good assurance provider... actually it was ourselves! The worst part is that we had to convince the client to make some changes and, obviously, the auditors who were from our firm were saying exactly the same thing." The firm maintained two separate identities—consultant and auditor—to create the appearance of independence while maintaining actual integration. When an unexpected client visit threatened to expose the arrangement, "they stayed in a nearby Starbucks until the client left the office."

This anecdote is not presented as representative of all assurance providers. Boiral et al. (2018) interviewed 38 providers across multiple countries, and the Mexican case was one of three interviews conducted in Mexico. But the anecdote is revealing of what becomes possible when commercial pressures go unchecked. The structural incentives that produce such extreme behavior in some contexts produce milder but still problematic behavior in others. The Big Four firm that uses assurance as a "carrot" is not committing fraud. They are not maintaining two separate identities. But they are structurally compromised. Their assurance conclusions are shaped by their desire to sell consulting services. The bias is subtle, deniable, and pervasive.

Perego and Kolk (2012, Table 6, p. 9) provide longitudinal data on provider market shares that reflect commodification dynamics. Accounting firms dominated the market in 1999 (55.6% of statements). Their share remained stable through 2005 (53.3%) but declined to 45.1% by 2008. Specialists (consulting firms) declined from 38.9% in 1999 to 13.4% in 2008. Certification bodies remained a small player. The "Others" category grew from 0% in 1999 to 30.5% in 2008. This shift toward non-traditional providers—academic institutions, NGOs, stakeholder panels—may reflect a search for lower-cost assurance or assurance perceived as more independent. But as Table 12 (p. 12) shows, the "Others" category delivered the lowest quality scores (mean 5.89 on the 27-point scale). The market shift toward low-cost, low-quality providers is exactly what commodification theory would predict. When price becomes the primary differentiator, quality declines.

5.2. Globalization of Lenient Assurance and Cross-Border Shopping

The paymaster problem is not confined within national borders. Firms can, and do, select assurance providers from jurisdictions with weaker oversight, lower standards, or more lenient enforcement. Prajogo et al. (2020, p. 13) note that "countries such as Australia and New Zealand have seen an influx of certifications by assurance providers with international accreditation that are informally known for their lenient practices. Such certifiers undermine the credibility of national conformity assessment infrastructure." A firm dissatisfied with a strict domestic certifier can simply hire an international certifier known to be more accommodating. The certifier's accreditation—perhaps from a jurisdiction with minimal oversight—provides a veneer of legitimacy. The client obtains the certificate. The national accreditation body has limited authority to intervene.

Cross-border shopping is enabled by the mutual recognition agreements that govern international trade in services. A certification issued by an accredited body in one country is generally recognized in others. This is a feature of the global trading system, not a bug. But the feature becomes a vulnerability when accreditation bodies vary in rigor. A lenient accreditor produces lenient certifiers. Those certifiers then compete with stricter certifiers in other markets. The stricter certifiers lose business because they are more expensive and more likely to fail clients. The market selects for leniency. The race is to the bottom.

Perego and Kolk (2012, p. 184) provide evidence of regional variation that is consistent with cross-border shopping pressures. Japan shows a distinct pattern: high adoption of assurance (39 statements across the panel, more than any country except the UK) but

heavy reliance on the "Others" category of providers (academic institutions, NGOs) that score poorly on quality metrics (mean 5.89). Japanese firms are obtaining assurance, but they are not obtaining high-quality assurance from professional accounting firms or certification bodies. They are obtaining low-cost, low-quality assurance from providers who may lack the competence or independence to perform rigorous verification. The institutional pressure to have an assurance statement is satisfied. The substantive value of that statement is questionable.

The United States presents the opposite pattern. Low adoption (only 12 assurance statements across the panel, 11.1% of reports) but relatively higher quality when assurance is obtained (mean 9.83, though with high variance). As Perego and Kolk (2012, p. 184) explain, "the litigious tradition of the U.S. market, characterised by a formal legal approach and contestation, appears to have stimulated a compliance orientation, in which voluntary reporting and assurance has consequently made much less headway." A US firm that obtains assurance may be motivated by a specific litigation concern—for example, a securities law filing that requires verified information. The assurance provider, facing potential liability, may be more rigorous. The result is a smaller but higher-quality market.

The United Kingdom represents a different equilibrium. High adoption (37 statements, the most of any country) and high quality (mean 18.57 in 2008, the highest of any country). The UK combines a stakeholder-oriented corporate governance tradition with a well-developed profession of sustainability assurance. The equilibrium is fragile, however. As Boiral et al. (2018, p. 1118) note, the quality of assurance "will continue to decline gradually until a huge scandal occurs." The UK market may be closer to that scandal than current quality scores suggest.

Prajogo et al. (2020, p. 13) raise a critical question about cross-border certification: "Should such scandals lead to removal of certifications of these firms? How should accreditation firms handle such cases?" The questions are unanswered. No international mechanism exists to revoke a certification issued by a lenient certifier in another jurisdiction. A firm caught with a fake ISO 9001 certificate may face reputational damage, but the certificate itself remains valid until the certifier revokes it. The certifier, if they are the source of the leniency, has no incentive to revoke. The system lacks self-correcting mechanisms.

5.3. Regulatory Gaps and Limited Oversight

Financial auditing operates under a regulatory architecture that includes: mandatory auditor rotation or tendering (in many jurisdictions), audit committee oversight of

auditor selection and compensation, restrictions on non-audit services provided to audit clients (e.g., Sarbanes-Oxley Section 201), public disclosure of audit fees and non-audit service fees, and regulatory inspections of audit firms (e.g., PCAOB inspections). This architecture is imperfect. It did not prevent Enron, WorldCom, or more recent scandals. But it provides a baseline of oversight that the non-financial assurance market lacks.

Prajogo et al. (2020, p. 13) explicitly note the oversight gap: "the findings underscore the consequences of the relatively limited oversight national standards bodies exercise over assurance providers. This is particularly magnified in countries that allow assurance providers to seek accreditation abroad." National standards bodies (e.g., JAS-ANZ in Australia and New Zealand) accredit certification bodies. But their resources are limited. Their enforcement powers are limited. Their jurisdiction over internationally accredited certifiers is unclear.

Boiral et al. (2018, p. 1118) provide a practitioner's perspective on regulatory gaps. One respondent noted: "On the ethical level, there is a real problem because the report indicates that it has been assured and verified [...] but only a part of the report has actually been verified [...] There are no specific rules requiring us to verify the entire report. However, would it be possible to have a financial audit based only on a small department and leaving the rest aside just because we think it is not important?" The answer is no. Financial auditing standards (ISA 320, ISA 450) require the auditor to consider materiality across the entire set of financial statements. A scope limitation that excluded a material account would result in a qualified opinion or disclaimer. Sustainability assurance has no equivalent requirement. Scope is negotiated with the client. The auditor may accept a scope that excludes material topics. The statement may not disclose the exclusion clearly.

The absence of a Sarbanes-Oxley equivalent for non-financial assurance is particularly significant. SOX Section 201 prohibits a registered public accounting firm from providing certain non-audit services (including consulting) to an audit client. The prohibition is not absolute—some services are permitted with pre-approval. But the law creates a bright line that firms must navigate. No similar prohibition exists for sustainability assurance. A firm can provide sustainability consulting and sustainability assurance to the same client. The same engagement team can perform both services. The conflict of interest is not merely potential. It is structural.

Boiral et al. (2018, p. 1119) document how this structural conflict operates in practice. The respondent who described assurance as "the big carrot" went on to explain: "Most of the accounting firms don't even significantly charge for the service, they keep the cost down as low as possible and give their assurance teams the mandate to do everything they can to make sure they don't piss off the audit client. And so, they're sort

of hamstring in their ability to actual criticize effectively the content of the report." The assurance team is not independent. They are operating under a mandate to preserve the client relationship for consulting sales. The conflict is not hidden. It is openly acknowledged within firms. The only question is whether it influences audit outcomes. The evidence suggests it does.

Prajogo et al. (2020, p. 14) propose a potential solution: "further research could explore mechanisms to reduce this financial dependence, such as the feasibility of shifting selection of the assurance providers from the auditee to the national standards organization." This proposal—removing provider selection discretion from the paymaster—directly addresses the structural cause of the ethical issues documented across all three papers. If a national accreditation body assigned assurance providers to firms, or if firms could only select from a pool of providers approved for their rigor, the paymaster's ability to shop for leniency would be constrained. The provider's dependence on the client would be reduced because the provider would not fear losing future assignments if they issued non-conformities.

The proposal faces practical challenges. National accreditation bodies would need significant additional resources. Assignment mechanisms would need to be transparent and fair. Firms would resist losing their discretion over provider selection. But the alternative—continuing the current system of paymaster-controlled, commercially pressured, symbolically oriented assurance—is untenable. As Boiral et al. (2018, p. 1118) warn: "I think that the quality will continue to decline gradually until a huge scandal occurs and, then, it will be noted that these audits, in fact, do not reflect reality. That's what happened in finance and accounting." The question is whether the non-financial assurance market will wait for that scandal before acting, or whether regulators, standard-setters, and professional bodies will implement structural reforms preemptively. The history of financial auditing suggests the former. The warnings from practitioners suggest the latter is necessary.

6. Deep Dive Case Studies: Granular Analysis of All Three Papers

Sections 2 through 5 integrated findings from the three source papers at a thematic level. This section descends from themes to specifics. Each paper receives its own deep dive, walking through the research design, key tables, statistical results, and quoted passages that support the report's central arguments. The goal is not summary but

interrogation: what exactly do the numbers show, and what limits should we place on their interpretation?

6.1. Case Study 1: Prajogo et al. (2020) — The SEM Model and Hypothesis Testing

Research Design and Sample

Prajogo et al. (2020) surveyed ISO 9001-certified firms in Australia and New Zealand. The sampling frame was the Joint Accreditation System of Australia and New Zealand (JAS-ANZ) database, listing over 10,000 certified organizations. From this frame, 4,000 Australian and 1,000 New Zealand firms were randomly selected. The survey targeted managers with "knowledge and responsibility for the implementation and maintenance of the company's quality management system." Usable responses numbered 597, a 12% response rate.

Table 1 (p. 7) shows the industry distribution. Manufacturing dominates at 29.6%, followed by construction (17.4%) and property and business services (11.6%). Table 2 (p. 7) shows firm sizes: 28% small (fewer than 20 employees), 60% medium (20-199 employees), 12% large (200+ employees). The sample is predominantly small-to-medium enterprises (SMEs), which is representative of the ISO 9001-certified population but limits generalizability to large multinationals. This is a deliberate trade-off. Large firms behave differently from SMEs in certification decisions (they have more resources, more complex supply chain pressures, more sophisticated quality systems). The findings should be understood as primarily applicable to SME contexts, with cautious extension upward.

Construct Measurement and Validation

Table 3 (p. 8) presents the construct validity and reliability results. Internal motives (two items: "To build a foundation for systematic business management" and "To have better control of operations of the business") show a loading path of 0.87 and 0.91 respectively, with construct reliability of 0.88. External motives (two items: "To meet customer and market requirements" and "To enhance company's image and reputation") show loadings of 0.53 and 0.78, reliability of 0.61. The lower loading for the first external motive item suggests it is a weaker indicator of the construct. "Meeting customer and market requirements" may be driven by genuine customer pressure rather than symbolic legitimacy-seeking. The conflation is problematic. A firm facing genuine customer pressure for quality may be externally motivated in a substantive sense. The measure does not cleanly separate symbolic from substantive external motivation.

Operational benefits (two items: "Improved level of quality of our products and/or services" and "Reduced our operating costs") show loadings of 0.73 and 0.70, reliability 0.68. Market benefits (two items: "Keeping business with our major customers" and "Improved our market coverage and penetration") show loadings of 0.76 and 0.88, reliability 0.81. All reliability values exceed the minimum 0.6 threshold, though the Pearson r correlations (reported in parentheses) are used instead of Cronbach's alpha because each scale has only two items. The correlation for external motives (0.41) is notably lower than for internal motives (0.79), again suggesting that the two external motive items capture somewhat different underlying constructs.

The two characteristics of assurance providers—strict auditing and market credibility—are measured by single items. This is a limitation acknowledged by the authors. Single-item measures cannot capture the multidimensionality of complex constructs. A strict audit could be strict in documentation review but lax in site observation. Market credibility could derive from brand reputation, size, longevity, or marketing expenditure. The single-item approach forces respondents to make a holistic judgment that aggregates multiple dimensions. The authors defend this choice (p. 8) based on focus group feedback and the argument that single items reduce halo effects. The defense is plausible but not dispositive. Multi-item scales would have been preferable.

The Structural Equation Model

Figure 1 (p. 9) presents the research model with path coefficients. The model shows good fit: NFI = 0.957, CFI = 0.969, GFI = 0.969, RMSEA = 0.060, SRMR = 0.045. These indices exceed conventional thresholds for acceptable fit (RMSEA < 0.08, CFI > 0.90). The model fits the data reasonably well.

The direct paths from motives to benefits (not mediated by provider selection) are estimated to control for the possibility that motives influence benefits through other channels (e.g., internal motives may drive substantive implementation of ISO 9001 regardless of which provider is selected). Internal motives have a direct effect on operational benefits of 0.38 ($p < 0.01$). External motives have a direct effect on market benefits of 0.26 ($p < 0.01$). Operational benefits also have a positive effect on market benefits of 0.36 ($p < 0.01$). These direct effects are substantial. Provider selection partially mediates the relationship between motives and benefits, but the direct effects dominate.

Hypothesis Testing Results

Table 4 (p. 11) summarizes the comparative hypothesis tests. The testing procedure is rigorous. For each hypothesis, the authors created a competing model where two path coefficients were constrained to be equal, then compared the chi-square fit of the

competing model to the original model ($\chi^2 = 113.83$, $df = 36$). A significant increase in χ^2 indicates the two paths are significantly different.

H1: Internal motives will have a stronger relationship with strict auditing than with market credibility. The competing model (paths set equal) produced $\chi^2 = 119.95$ ($df = 37$), $\Delta\chi^2 = 6.12$, $\Delta df = 1$. Since $6.12 > 3.84$ (the critical value for $p < 0.05$), H1 is supported. The path coefficients confirm the direction: internal motives $\rightarrow$ strict auditing = 0.24 ($p < 0.01$); internal motives $\rightarrow$ market credibility = 0.12 ($p < 0.05$). The effect on strict auditing is twice the magnitude of the effect on market credibility.

H2: External motives will have a stronger relationship with market credibility than with strict auditing. The competing model produced $\chi^2 = 133.26$ ($df = 37$), $\Delta\chi^2 = 19.43$, $\Delta df = 1$. Since $19.43 > 6.63$ ($p < 0.01$), H2 is supported. Path coefficients: external motives $\rightarrow$ market credibility = 0.39 ($p < 0.01$); external motives $\rightarrow$ strict auditing = 0.23 ($p < 0.01$). The effect on market credibility is 70% larger than the effect on strict auditing.

H3: Strict auditing will have stronger relationships with operational benefits than with market benefits. The competing model (setting the path from strict auditing to operational benefits equal to the path from market credibility to operational benefits) produced $\chi^2 = 117.63$ ($df = 37$), $\Delta\chi^2 = 3.80$, $\Delta df = 1$. This exceeds the critical value for $p < 0.05$ (3.84) by a narrow margin. H3 is supported, though the support is marginal. Path coefficients: strict auditing $\rightarrow$ operational benefits = 0.24 ($p < 0.01$); market credibility $\rightarrow$ operational benefits = -0.03 (not significant). The non-significant negative coefficient for market credibility is striking. Market credibility providers do not merely deliver less operational benefit than strict providers. They deliver no detectable operational benefit at all.

H4: Market credibility will have stronger relationships with market benefits than with operational benefits. The competing model (setting strict auditing $\rightarrow$ operational benefits equal to market credibility $\rightarrow$ market benefits) produced $\chi^2 = 116.58$ ($df = 37$), $\Delta\chi^2 = 2.75$, $\Delta df = 1$. This exceeds the critical value for $p < 0.1$ (2.71) but not for $p < 0.05$ (3.84). H4 is supported at the 90% confidence level, a weaker standard than the other hypotheses. Path coefficients: market credibility $\rightarrow$ market benefits = 0.12 ($p < 0.05$); strict auditing $\rightarrow$ market benefits = 0.06 (not significant).

Critical Interpretation

The pattern of coefficients reveals a troubling asymmetry. Internal motives drive strict auditing (0.24), which drives operational benefits (0.24). External motives drive market credibility (0.39), which drives market benefits (0.12). The effect of external motives on market benefits is mediated through provider selection, but the indirect effect ($0.39 \times 0.12 = 0.0468$) is much smaller than the direct effect of external motives on market

benefits (0.26). Most of the market benefit that externally motivated firms receive comes directly from pursuing certification, regardless of which provider they select. The additional benefit from selecting a market credibility provider is marginal.

This finding undermines a simple rational-choice interpretation. If market credibility providers delivered substantial incremental market benefit, the indirect effect would be larger. The data show otherwise. So why do externally motivated firms select market credibility providers? Two possibilities. First, the survey measure of market benefits may not capture the full range of benefits that firms seek. A market credibility provider might help a firm gain access to a specific customer that requires certification from an approved provider list. This benefit would not appear in the general market benefits scale. Second, firms may select market credibility providers because they reduce the risk of certification failure, not because they increase market benefits conditional on certification. The model does not include a measure of certification success or failure. The risk-reduction hypothesis cannot be tested with the available data.

Limitations Acknowledged by the Authors

Prajogo et al. (2020, p. 14) identify several limitations. The single-item measures for provider characteristics are noted. The cross-sectional design cannot establish causality; the direction of effects could be reversed (firms that achieve operational benefits may attribute them to strict auditing). The sample is limited to Australia and New Zealand, countries with relatively low corruption, which may understate the paymaster problem compared to jurisdictions with weaker institutions. The response rate of 12% raises non-response bias concerns, though the authors note this is "generally acceptable for research where the unit of analysis is at the firm's level."

6.2. Case Study 2: Perego and Kolk (2012) — Adoption Patterns and Content Analysis

Research Design and Sample

Perego and Kolk (2012) studied Fortune Global 250 firms over a 10-year period (1999-2008). The initial sample was drawn from the first half of the Fortune Global 500 list as of August 1998. Firms were contacted to request their most recent environmental, social, or sustainability report. Subsequent data collection occurred at three-year intervals (1999, 2002, 2005, 2008). The final panel consisted of 212 firms that survived the full period (accounting for mergers, acquisitions, and disappearances).

Table 1 (p. 6) provides aggregate statistics by country. The United States contributes the most firms (70, 33.0%), followed by Japan (48, 22.6%), Germany (21, 9.9%), France (18, 8.5%), and the United Kingdom (15, 6.1%). Sustainability reports are most frequent from Japan (125 reports), the US (108), Germany (59), the UK (50), and France (43). Assurance statements are most frequent from Japan (39 statements), the UK (37), Germany (18), the Netherlands (15), and France (13). The percentage of reports with assurance varies dramatically by country: UK (74.0%), Spain (83.3%), Italy (80.0%), Netherlands (60.0%), Japan (31.2%), France (30.2%), US (11.1%). The US is an outlier. Assurance is rare despite the large number of reporting firms.

Adoption Patterns Over Time

Table 2 (p. 7) shows the evolution of reporting and assurance by country and year. Aggregate reporting increased from 84 reports (39.6% of the panel) in 1999 to 147 reports (69.3%) in 2008. Assurance statements increased from 18 (21.4% of reports) to 82 (55.8% of reports). The growth in assurance outpaced the growth in reporting. By 2008, more than half of sustainability reports in this global panel included some form of external assurance.

Table 5 (p. 9) categorizes firms by adoption pattern. Only 12 firms (5.7% of the panel) were "consistent early adopters" (assurance in all four periods). These trendsetters are headquartered in Europe (except for one Australian firm) and belong mostly to Oil and Gas and Utilities sectors. The largest category is "non-adopters": 118 firms (55.7% of the panel) never obtained assurance across the four periods. Between these extremes are "consistent late adopters" (9 firms starting in 2002), "consistent followers" (13 firms starting in 2005), "laggards" (42 firms starting in 2008), and intermittent adopters (18 firms with irregular patterns).

The large proportion of non-adopters is striking given that the sample consists of Fortune Global 250 firms—the world's largest multinationals. If assurance were a genuinely valuable practice, these firms would have adopted it. That more than half never did suggests either that the perceived benefits do not justify the costs, or that institutional pressure for assurance varies significantly across countries and sectors. The sectoral distribution in Tables 3 and 4 (p. 8) shows that "more polluting" sectors (Oil and Gas, Utilities, Chemicals) have higher adoption rates, consistent with legitimacy theory: firms facing greater environmental scrutiny adopt assurance to manage stakeholder perceptions.

Assurance Provider Market Share

Table 6 (p. 9) tracks the market share of different provider types over time. The classification distinguishes Accounting firms, Specialists (broader and specialist

consultants), Certification bodies, and Others (academic institutions, NGOs, stakeholder panels, individual auditors). Accounting firms dominated the market in 1999 (55.6% of statements) but their share declined to 45.1% by 2008. Specialists declined even more sharply, from 38.9% to 13.4%. Certification bodies remained a small player (around 11% in 2008). The "Others" category grew from 0% to 30.5% over the decade.

The rise of "Others" is particularly notable in Japan. The authors note (p. 178) that "the role of academic institutions in providing assurance is overrepresented when compared to European-based firms where Accounting firms have been the traditional choice." This finding raises questions about the professionalization of assurance in different institutional contexts. Academic institutions may lack the audit methodology training and professional skepticism orientation that accounting firms develop through financial audit experience. The quality consequences are examined below.

Content Analysis Methodology and Results

The content analysis coding framework is provided in the Appendix (p. 15-16). Nineteen items are coded, with scores ranging from 0 to 27. The items cover formal aspects (title, addressee, assurer name and location, report date), responsibilities (reporter, assurer, independence, impartiality), scope and objective, competencies, criteria, standards, summary of work performed, materiality, completeness, responsiveness, and general conclusion. The materiality item is coded on a 4-point scale (0-3), with the highest score requiring "reference, explanation of materiality setting and stakeholder perspective introduced."

Table 9 (p. 10) shows the mean quality score by year. The average score increased from 9.72 in 1999 to 12.93 in 2008. The improvement is statistically significant but substantively modest. An average score of 12.93 out of 27 represents less than 48% of the maximum possible quality. Even in 2008, the average assurance statement omitted more than half of the desirable content items.

Table 10 (p. 11) shows quality scores by country. The United Kingdom achieves the highest mean (18.57 in 2008, 13.86 across the whole period), followed by Germany (14.90 in 2008, 13.33 across the period), France (12.75 in 2008, 12.23 across the period), and Japan (7.25 in 2008, 8.03 across the period). The US shows high variance but a mean of 10.71 in 2008 (though based on only 12 statements across the entire period, so small sample caution applies). The country-level variation is substantial. A stakeholder reading an assurance statement from a UK firm receives, on average, a much more informative statement than a stakeholder reading a statement from a Japanese firm. The variation is not explained by any single factor. Regulatory pressure, professional infrastructure, litigation risk, and cultural factors all likely contribute.

Table 11 (p. 11) shows quality scores by sector. Mining achieves the highest mean (14.75), followed by Food and Beverages (16.90—though this is driven by a small number of statements), Oil and Gas (13.39), and Banks and Insurance (12.50). The variation across sectors is less dramatic than across countries. High-impact sectors (mining, oil and gas) tend to have higher-quality assurance, consistent with legitimacy theory: firms with greater environmental and social risks invest more in credible assurance.

Table 12 (p. 12) shows quality scores by assurance provider type. This is the most consequential table for evaluating the paymaster problem. Accounting firms achieve the highest mean (13.71) but also the highest variance (standard deviation 4.25). Their maximum score is 22 out of 27—no accounting firm in the sample achieved a perfect or near-perfect statement. Certification bodies achieve a similar mean (13.50) with higher variance (6.80). Specialists achieve a lower mean (10.80). The "Others" category achieves a strikingly low mean (5.89).

The "Others" category finding is damning. Academic institutions, NGOs, and stakeholder panels—entities that might be expected to bring independence and stakeholder focus—instead deliver the lowest-quality assurance. The implication is that independence alone is insufficient. Assurance quality requires technical competence in verification methodologies, familiarity with assurance standards, and a professional culture that values rigor over accommodation. The "Others" providers may lack these attributes. Their clients—predominantly Japanese firms, based on the country-level data—receive assurance statements that provide minimal information and may even mislead stakeholders into believing a substantive verification occurred when none did.

The Coding Framework's Normative Assumptions

The content analysis framework reflects a particular normative view of what constitutes high-quality assurance. The emphasis on stakeholder inclusivity (items 16-18 on materiality, completeness, responsiveness) derives from the AA1000 standard's stakeholder accountability perspective. Firms and providers that prioritize financial statement users over broader stakeholders would score lower on these items, even if their assurance is rigorous in other respects. This is a limitation of the framework. A financial investor might not care whether the assurance statement addresses stakeholder responsiveness. They might care about the accuracy of reported emissions data. The framework does not weight items by their importance to different user groups. A statement could score poorly on stakeholder items but still serve the information needs of investors. The authors acknowledge this indirectly (p. 175) by noting that AA1000 "most closely aligns itself with the stakeholder accountability perspective" while

ISAE3000 reflects a professional accounting orientation. The coding framework privileges the former.

6.3. Case Study 3: Boiral et al. (2018) — Practitioner Perspectives on Ethical Dilemmas

Research Design and Sample

Boiral et al. (2018) conducted 38 semi-structured interviews with assurance providers. The sample included 23 respondents from accounting firms and 15 from consulting firms. Geographic distribution: Europe (31%), North America (31%), Asia and Oceania (22%), Africa (13%), South America (3%). Hierarchical levels: partners/directors (39%), auditors/consultants (61%). Selection was purposive, not random. The goal was not statistical representativeness but theoretical saturation—reaching a point where additional interviews would not yield new insights.

Table 1 (p. 1115) shows the distribution. The sample is weighted toward Europe and North America, with limited representation from developing economies. The Mexican case (three interviews) provides some insight into emerging market dynamics, but generalizability to other developing contexts is uncertain. The authors note that snowball sampling from an initial list of providers was used, which may introduce selection bias toward well-connected providers.

Interviews were conducted in English, French, or Spanish, by Skype or phone, lasting 45 minutes to 2 hours. The semi-structured script (Appendix, p. 1122-1123) covered four sections: information about the assurance process, methodology used, results of the assurance process, and professionalism/independence of assurance providers. The script includes open-ended questions (e.g., "Opinion about the general content of sustainability reports," "Distinctions between assurance and consulting works") that allow respondents to raise issues the researchers did not anticipate.

Key Quantitative Findings from the Interviews

Figure 2 (p. 1120) reports the percentage of interviewees who mentioned each ethical issue. Commercial pressures: 45%. Superficiality of the verification process: 45%. The line between auditing and consulting: 33%. Familiarity with audited companies: 21%. These percentages are not survey statistics in the conventional sense. They represent the proportion of interviewees who spontaneously raised each issue during semi-structured interviews. The percentages are likely lower bounds. An issue that 45% of interviewees

spontaneously mentioned is clearly salient. An issue that only 21% mentioned may still be important but less top-of-mind.

The Interconnectedness of Ethical Issues

Figure 1 (p. 1116) presents a conceptual model of the interconnectedness of ethical issues. The model shows how commercial pressures (at the center) drive superficial verification, blurring of consulting-auditing lines, and familiarity with clients. Each of these, in turn, reinforces the others. Familiarity increases commercial pressures (clients expect continued relationships). Superficial verification makes it easier to sell consulting services (the client is satisfied with a low-friction audit). The model is not empirically tested in the paper but is derived inductively from the interview data. The figure serves as a visual summary of the relationships respondents described.

Key Quotations and Their Implications

Several quotations from Boiral et al. (2018) are sufficiently revealing to warrant direct analysis.

On subconscious bias (p. 1117): "If it's really a client that the auditing company is dependent on for income, it might be subconsciously more supportive of the organisation because they don't really want to get them offside." The word "subconsciously" is critical. The respondent is not describing explicit corruption or intentional fraud. They are describing a psychological mechanism that operates below awareness. Even well-intentioned auditors may systematically favor clients on whom their firm depends. The bias is not a moral failing. It is a cognitive vulnerability. Structural safeguards (rotation, independent assignment, separation of consulting) are required to protect against it.

On the confusion of limited assurance (p. 1117): "Negative statements such as 'we did not see anything that would cause us not to believe that...' are used because it's only a moderate level of assurance. I think that, in some way, it degrades the value of the assurance statement for a lot of users of those reports, because it's almost like saying: Well, maybe the auditors didn't do enough to find out if there was a problem, so what's the value of the assurance?" The respondent, a former auditor, is critiquing their own profession. The negative phrasing that dominates limited assurance statements provides little positive information. Stakeholders who understand the distinction discount the statement appropriately. Stakeholders who do not understand the distinction may be actively misled. The respondent does not offer a solution, only a diagnosis.

On the Mexican dual-identity case (p. 1119): "There were two different phone numbers and addresses in our firm. For example, I could work on the sustainability report or the

integrated report of one of our clients, and when the client asked our advice about the assurance of the report, we had to answer that we knew a very good assurance provider... actually it was ourselves! The worst part is that we had to convince the client to make some changes and, obviously, the auditors who were from our firm were saying exactly the same thing." This passage describes a clear violation of ethical principles. The firm maintained two separate identities to create the appearance of independence. When an unexpected client visit threatened to expose the arrangement, "they stayed in a nearby Starbucks until the client left the office." The anecdote is extreme, but it illustrates what becomes possible when commercial pressures go entirely unchecked. The respondent is a former employee who left for ethical reasons. The fact that they are no longer at the firm may increase the credibility of their account (they have less incentive to protect the firm) but also introduces potential bias (they may exaggerate past misconduct to justify their departure).

On partner rotation and familiarity (p. 1119-1120): "It's amazing how quick you can become very familiar with the client, and get into their systems. So you understand their systems and start providing more guidance than is really reasonable in terms of neutrality and independence requirements. And so I rotate out every 5 years and I am the only one who does it." This respondent (a partner with a consulting firm) is describing a personal practice of rotation that is not mandated by any standard or regulation. They rotate every 5 years. Others do not. The quotation reveals that rotation is not standard practice in sustainability assurance, even though the familiarity threat is well-recognized in financial auditing. Another respondent (p. 1120) notes: "I don't think it's necessarily healthy for the same assurance provider to provide assurance year after year because I think you get too familiar, you get to know the people and that does reduce objectivity. So I think we need to change at least the lead assurance providers that work with the same organisation." The need is acknowledged. The practice is not widespread.

Techniques of Neutralization

Boiral et al. (2018, p. 1120) identify "techniques of neutralization" that assurance providers use to rationalize ethically questionable behavior. These include:

- Justifying familiarity as necessary for organizational learning
- Justifying superficial verification as meeting client requests
- Justifying consulting-auditing overlap as serving client needs

The concept derives from Sykes and Matza's (1957) criminological theory of delinquency, later applied to business ethics. Techniques of neutralization allow individuals to engage in behavior that violates their moral standards while maintaining a

self-image as ethical. The techniques are not lies. They contain elements of truth. Familiarity does enable learning. Clients do request limited scopes. Consulting does serve client needs. The neutralization works by emphasizing the partial truth while ignoring the countervailing ethical obligation. The auditor who says "the client wanted a limited scope" is not lying, but they are omitting that they had a professional obligation to explain the limitations of that scope to stakeholders.

The identification of neutralization techniques is a contribution of the Boiral et al. (2018) paper that the other two papers lack. Prajogo et al. (2020) and Perego and Kolk (2012) treat audit quality as a measurable outcome that varies systematically with provider characteristics and institutional pressures. Boiral et al. (2018) open the black box and show how individual auditors justify their own compromises. The neutralization perspective suggests that improving audit quality requires not just structural changes (rotating providers, separating consulting) but also changes in professional culture that make neutralization harder to sustain. Auditors need to be unable to say "the client wanted it" without feeling shame.

Limitations of the Boiral et al. (2018) Study

The sample size (38 interviews) is appropriate for qualitative research but cannot support statistical generalization. The authors do not claim generalizability. The geographic distribution is skewed toward Europe and North America. The Mexican findings are based on only three interviews; the Starbucks anecdote, while vivid, may not be representative of Mexican practice more broadly. The use of snowball sampling may have selected for providers who are more reflective or more critical; silent practitioners who do not discuss ethical issues may be underrepresented. The authors acknowledge (p. 1122) that "ethically questionable behavior in sustainability assurance is likely to be under-reported" due to social desirability bias. The 45% who spontaneously mentioned commercial pressures may be the tip of an iceberg. The actual prevalence could be higher.

7. Conclusion and Future Research Directions

7.1. Synthesis of Theoretical and Empirical Findings

This report began with a thesis: while traditional auditing theory assumes a linear, evidence-based path from client engagement to independent assurance to reduced

information asymmetry, the proliferation of non-financial auditing reveals a decoupled reality where symbolic legitimacy and market credibility systematically supersede substantive verification. The evidence from all three papers supports this thesis, though with important qualifications.

The paymaster problem is structural, not individual. Agency theory predicts that monitoring by an independent third party can reduce information asymmetry. But the model assumes the monitor is hired by the principal, not the agent. When the agent (the firm) hires and pays the monitor (the assurance provider), the theoretical foundation crumbles. The monitor's loyalty runs to the paymaster. The information asymmetry between principal and agent is not reduced. It is transformed into a different asymmetry: principals know that some certified firms are undeserving but cannot identify which ones.

The information and insurance hypotheses offer partial correctives. Information benefits accrue when assurance is substantive: strict auditing delivers operational improvements (Prajogo et al., 2020, coefficient 0.24). But market credibility providers deliver no such benefit. Insurance benefits are largely absent from non-financial assurance due to liability limitation clauses and the absence of a Sarbanes-Oxley equivalent. Stakeholders cannot sue assurance providers for superficial verification. The providers know this. The knowledge shapes their behavior.

Institutional and legitimacy theory provides the most complete explanation for the patterns observed across all three papers. Firms pursue assurance not primarily to improve information or transfer risk but to gain legitimacy in the eyes of stakeholders, regulators, and competitors. The assurance statement is a ritual. Its content matters less than its presence. The decoupling between symbolic assurance and substantive performance is not a bug. It is a feature of an institutional environment that rewards appearances over reality.

The empirical evidence for decoupling is strong. Prajogo et al. (2020) show that externally motivated firms select market credibility providers, who deliver no operational benefit. Perego and Kolk (2012) show that assurance quality scores average less than 50% of the maximum possible, even in the most recent period, and that "Others" category providers (academic institutions, NGOs) score only 5.89 out of 27. Boiral et al. (2018) show that 45% of assurance providers spontaneously acknowledge commercial pressures and superficial verification, and describe techniques of neutralization that justify ethically questionable behavior.

The information asymmetry that assurance was meant to reduce has not been eliminated. It has been compounded. Stakeholders now face two layers of uncertainty:

uncertainty about the firm's underlying performance, and uncertainty about whether the assurance statement provides any reliable information about that performance. A stakeholder reading an assurance statement from an "Others" provider in Japan cannot know whether the verification was substantive or symbolic. A stakeholder reading a statement from a Big Four accounting firm cannot be certain either, though the probability of substantive verification is higher. The market does not reliably signal assurance quality. The paymaster relationship ensures that it cannot.

7.2. Regulatory Implications

The voluntary model has failed. After two decades of sustainability reporting and ISO 9001 certification, the problems documented in these papers persist. Fake certificates continue to circulate (Prajogo et al., 2020, p. 2). Assurance quality remains low in many jurisdictions (Perego & Kolk, 2012, Table 10). Practitioners acknowledge ethical compromises (Boiral et al., 2018, Figure 2). Incremental improvements—better standards, more training, revised guidance—are unlikely to change the structural incentives that produce these outcomes. Structural problems require structural solutions.

Mandatory rotation of assurance providers. Boiral et al. (2018, p. 1119-1120) document the familiarity threat: auditors who work with the same client for years lose objectivity. One respondent rotates every 5 years and notes they are "the only one who does it." Mandatory rotation—of firms, not just lead partners—would reduce familiarity. The European Union's audit reform (2014) mandated auditor rotation for public-interest entities after a maximum of 10 years, with a 4-year cooling-off period. A similar requirement for non-financial assurance would align the two domains. The challenge is implementation: many non-financial assurance engagements are not with public-interest entities, and regulatory authority varies by jurisdiction.

Removing provider selection discretion from the paymaster. Prajogo et al. (2020, p. 14) suggest shifting provider selection to national standards bodies or accreditation organizations. The proposal directly addresses the source of the paymaster problem. If firms cannot select their auditor, they cannot shop for leniency. The auditor's loyalty would run to the assigning body, not to the client. The challenge is practical: national standards bodies lack the resources and infrastructure to assign assurance providers to thousands of firms. A hybrid model—where firms select from an approved panel of providers whose audits are subject to random quality inspections—might be more feasible. The key is removing the firm's ability to fire a provider for issuing non-conformities.

Prohibiting the joint provision of consulting and assurance. The Sarbanes-Oxley Act (2002) prohibits registered public accounting firms from providing certain non-audit services to audit clients. The prohibition was a direct response to the Enron scandal, where Arthur Andersen served as both auditor and consultant. No equivalent prohibition exists for non-financial assurance. Boiral et al. (2018, p. 1119) document the consequences: assurance is used as a "carrot" to sell consulting services, and providers are "hamstringing in their ability to actually criticize effectively the content of the report." Extending SOX-style restrictions to non-financial assurance would require legislative action in each jurisdiction or a coordinated international standard. The International Ethics Standards Board for Accountants (IESBA) could take the lead, but their authority over non-accounting providers (specialists, certification bodies, "Others") is limited.

Enhanced oversight and enforcement. National accreditation bodies need more resources and stronger enforcement powers. The problem of cross-border accreditation shopping (Prajogo et al., 2020, p. 13) requires international coordination. A mutual recognition framework that permits lenient certifiers from one jurisdiction to operate in another undermines quality everywhere. One solution is a global register of assurance providers with tiered recognition based on demonstrated quality. Another is mandatory disclosure of assurance provider inspection results, similar to the PCAOB's public inspection reports for audit firms. Stakeholders could then assess whether a provider has a history of quality problems.

7.3. Future Research Agenda

The papers in this report open several avenues for future research.

Quantifying the cost of symbolic certification. Prajogo et al. (2020, p. 14) note that customers may pay premiums for certified products without receiving commensurate quality. The magnitude of this wealth transfer is unknown. Research could compare prices of certified versus non-certified products in markets where certification is common, controlling for other quality signals. If certified products command a premium but do not deliver higher quality, customers are systematically overpaying. The welfare implications are significant.

Cross-jurisdictional enforcement studies. The problem of lenient certifiers moving across borders (Prajogo et al., 2020, p. 13) could be studied through tracking studies. Researchers could identify certifiers known for leniency in one jurisdiction, then track whether they gain market share in jurisdictions with stricter oversight. If the pattern holds, it would provide evidence for the "race to the bottom" hypothesis. If not, other factors (reputation concerns, customer preferences) may constrain the market.

Whistleblower studies in assurance. Boiral et al. (2018) interviewed practitioners who acknowledged ethical issues but did not report them outside the firm. Research could investigate the conditions under which assurance providers report misconduct to regulators or professional bodies. The Mexican case (dual identities, Starbucks) involved a former employee who left the firm. Current employees may be reluctant to speak. Anonymous surveys or confidential reporting mechanisms could yield more accurate estimates of misconduct prevalence.

AI and algorithmic auditing as a solution to human bias. The biases documented in these papers—commercial pressure, familiarity, subconscious client favoritism—are human biases. Artificial intelligence systems do not suffer from them. An algorithmic audit of sustainability data could be consistent, scalable, and independent of client relationships. The challenge is technical: many non-financial metrics (e.g., stakeholder engagement quality, supply chain labor practices) are not easily quantified. The challenge is also institutional: who would control the algorithm, and how would stakeholders verify its integrity? Research on algorithmic auditing is in early stages but could transform the field.

The role of financial analysts and institutional investors. Perego and Kolk (2012, p. 186) note that little is known about how sophisticated users of sustainability reports (e.g., financial analysts) react to assurance statements. Experimental studies (e.g., Pflugrath et al., 2012) have begun to address this gap. Further research could examine whether analysts discount assurance statements from low-quality providers, and whether their reactions vary by jurisdiction. If analysts do discriminate based on assurance quality, firms would have a financial incentive to invest in rigorous assurance.

Longitudinal studies of assurance quality trends. Perego and Kolk (2012) show modest quality improvements from 1999 to 2008. More recent data are needed. Has quality continued to improve, stagnated, or declined? The practitioner warning that "quality will continue to decline gradually until a huge scandal occurs" (Boiral et al., 2018, p. 1118) suggests stagnation or decline. Longitudinal content analysis of assurance statements from 2010 to 2025 would test this prediction.

7.4. Final Reflection

The audit function rests on trust. Stakeholders trust that the auditor's report means something—that evidence was gathered, procedures were followed, conclusions were honestly reached. The financial audit scandals of the early 2000s (Enron, WorldCom) damaged that trust. Regulation (Sarbanes-Oxley) and enforcement (PCAOB) partially restored it. Non-financial assurance has not yet had its Enron moment. The scandals so

far have been smaller: fake ISO 9001 certificates, superficial sustainability assurance statements that omitted material information, conflicts of interest between consulting and auditing. But the structural conditions for a larger scandal are in place.

The paymaster relationship creates incentives for leniency. The absence of liability protection removes constraints on leniency. The inability of stakeholders to assess audit quality makes leniency difficult to detect. The globalization of accreditation enables shopping for the most lenient provider. These conditions are not unique to any single country or sector. They are systemic features of the non-financial assurance market as currently structured.

The practitioners interviewed by Boiral et al. (2018) know this. One warned: "I think that the quality will continue to decline gradually until a huge scandal occurs and, then, it will be noted that these audits, in fact, do not reflect reality. That's what happened in finance and accounting." The warning is explicit. The question is whether regulators, standard-setters, and professional bodies will act before the scandal arrives, or wait until after, as they did with Enron. The history of auditing suggests the latter. The opportunity to learn from history and act preemptively remains open, if unlikely to be seized.

Source Articles:

1. Paymasters and Assurance Providers: Exploring Firms' Discretion in Selecting Non-financial Auditors

Daniel Prajogo¹ · Pavel Castka² · Cory Searcy³

Received: 6 August 2019 / Accepted: 16 May 2020 / Published online: 21 May 2020

2. Multinationals' Accountability on Sustainability: The Evolution of Third-party Assurance of Sustainability Reports

Paolo Perego · Ans Kolk

Published online: 16 August 2012

Springer Science+Business Media B.V. 2012

3. Ethical Issues in the Assurance of Sustainability Reports: Perspectives from Assurance Providers

Olivier Boiral¹ · Inaki Heras-Saizarbitoria² · Marie-Christine Brotherton¹ · Julie Bernard¹

Received: 22 September 2017 / Accepted: 14 March 2018 / Published online: 19 March 2018