UCLouvain
FUCaM MONS

LOUVAIN
School of Management

Louvain School of Management

Disclosure Quality, Legitimacy, and Investor Resilience: A Replication and Extension of Zahller, Arnold, and Roberts (2015)

Evidence that the accuracy–legitimacy pathway travels to a post-CSRD semiconductor setting

Author : Antonin COUPLET
Supervisor(s): Rudy DE WINNE
Academic year 2025.-2026.
Master [120] in Management (International Business)

Abstract

Corporate sustainability reports have never been more detailed, or more distrusted. As disclosure volume grows under regulations like the EU Corporate Sustainability Reporting Directive (CSRD), the report itself stops signalling anything. What matters is its quality. Zahller, Arnold, and Roberts (2015) provided the only experimental evidence that CSR disclosure quality builds investor trust that protects firms during a crisis, but their study is a decade old, set in a voluntary disclosure regime, a low-materiality sector, and a narrower investor population than exists today.

This thesis retests that pathway under conditions Zahller et al. couldn't: a post-CSRD regulatory environment, the semiconductor sector, and a population of retail investors recruited via Prolific (N = 90). A 2×2 between-subjects experiment manipulated the accuracy and completeness of a fictional semiconductor firm's CSR report. Perceived organisational legitimacy was measured before an exogenous sector wide shock, and social resilience after it. Moral Awareness was tested as an individual level moderator.

Neither accuracy nor completeness affected the composite resilience measure directly. Both, however, raised perceived legitimacy, and legitimacy fully mediated their effect on resilience (replicating Zahller et al.'s pathway). The accuracy-legitimacy coefficient found here ($\beta = .31$) is nearly identical to their original estimate ($\beta = .335$) despite every surrounding condition having changed, evidence that the mechanism travels across contexts rather than being specific to the original study. Completeness's effect on legitimacy was stronger here than in 2015, consistent with disclosed shortfalls retaining signalling value even once broad coverage became a legal requirement. Moral Awareness couldn't be meaningfully tested due to a pronounced ceiling effect in the sample.

Exploratory item-level analysis suggests social resilience is not a single behaviour. Legitimacy fully explains why better disclosure reduces investors' intention to divest, exactly as the theory predicts. It explains nothing, however, about investor patience during a slow recovery: there, completeness has a direct, unexpected effect of its own, leaving investors less patient once a firm has admitted its own shortcomings. A pattern with some resonance to backfire effects documented elsewhere in the CSR literature. This difference, while exploratory, suggests disclosure quality protects some investor behaviours more than others, rather than resilience uniformly, and that "social resilience" may be better understood as several distinct investor responses than a single construct.

Responsible use of generative AI

For this thesis, Claude (Anthropic) and Deepl was used as a linguistic assistant to improve the wording and grammatical accuracy of the english text, and to assist with drafting and debugging R code for the statistical analyses. All ideas, analyses, arguments and conclusions presented are my own, generative AI was not used to generate content or results.

Statements on responsible use of generative AI	Your answer
The content of my/our master's thesis is my/our original output, even if I/we used generative AI to help prepare it.	☒Yes ☐No
I/we master the theories, tools and methods that I/we use for the thesis.	☒Yes ☐No
I/we verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.	☒Yes ☐No
I/we acknowledge that I/we master the final output and that I am/we are able to explain it and answer questions about it.	☒Yes ☐No
I/we made sure not to provide generative AI tools with access to confidential data or information	☒Yes ☐No

Signature

	Author last name, first name(s)	Date	Signature
1	Antonin Couplet	26-07-09	

Acknowledgements

I would like to thank my supervisor, M. Rudy De Winne, for his rigorous and consistently available guidance throughout this work. His precise feedback, his ability to identify where clarity was lacking, and his patience through my revisions were essential to the quality of this thesis. I am also grateful for his support in the administrative steps related to funding the data collection.

I would also like to thank the “Fonds de la Recherche Scientifique” FNRS for the funding granted (Grant n° 40013829), which made the experimental data collection underlying this thesis possible.

Finally, I want to thank my family and friends for their patience, especially those who sat through the 10 minutes experiment more than once, and did it in English, which certainly didn't make it easier.

Table of contents:

1. Introduction 1
2. Literature review 2
2.1 Disclosure Quality as a Signal 2
2.2 Building Organizational Legitimacy 3
2.3 The Reputational Buffer: From Legitimacy to Resilience 4
2.4 The Missing Link: Moral Awareness 5
2.5 The Zahller et al. (2015) Study and the Present Replication 6
3. Methodology 8
3.1 Research Design 8
3.2 Participants and Recruitment 9
3.3 Experimental Procedure 10
3.4 Manipulations 12
3.5 Summary of Measures 12
3.6 Testing H1 and H2: Direct Effects on Resilience 13
3.7 Testing H3: Does Legitimacy Mediate the Relationship? 15
3.8 Testing H4: Does Moral Awareness Moderate the Effect? 17
4. Empirical Analysis and Results 20
4.1 Sample Description 20
4.2 Manipulation Checks 21
4.3 Descriptive Statistics and Correlations 21
4.4 Hypothesis Testing 23
4.4.1 H1: Does accuracy increase resilience? 23
4.4.2 H2: Does completeness increase resilience? 23
4.4.3 H3: Does legitimacy explain why disclosure quality matters? 24
4.4.4 H4: Does moral sensitivity change how strongly disclosure quality matters? 25
4.5 Exploratory Item-Level Analyses 25
5. Discussion 28
5.1 The Mediating Role of Legitimacy 28
5.2 Moral Awareness and the Limits of H4 29
5.3 Deconstructing Social Resilience 29
5.4 Limitations and Future Research 30
6. Conclusion 32
References 33

List of Tables:

Table 1: Summary of Hypotheses 10
Table 2: Summary of Measures 14
Table 3: Participant Demographics (N = 90) 22
Table 4: Means and Standard Deviations by Condition 23
Table 5: Zero-Order Correlations 23
Table 6: ANCOVA Results for Social Resilience 24
Table 7: Mediation Results (PROCESS Model 4) 26

List of Figures:

Figure 1: Experimental workflow 11
Figure 2: Mediation relationship 16

List of Appendices

Appendix A: Experimental Stimulus Materials 36
Appendix B: Exogenous Shock: Industry News Stimulus 38
Appendix C: Full Survey Instrument (Measurement Items) 39
Appendix D: Construct Validation: Full PCA and Reliability Statistics 41
Appendix E: Robustness Checks for H3 42
Appendix F: Item level mediation Models (Section 4.5) 43
Appendix G: Full R Code.....44

1. Introduction

Sustainability reports keep getting longer, but investor trust hasn't followed. When every firm publishes one, the report itself signals nothing. What matters is its quality: how precise its figures are, and how honest it is about what went wrong.

Zahller, Arnold, and Roberts (2015) showed, in a single experimental study, that this precision and honesty do more than inform investors. They build a reserve of trust that protects a firm when a crisis hits, even one it did not cause. A decade later, that finding has never been re-tested, and the world it described has changed. Disclosure that was once voluntary is now legally required across Europe. The retail investors reading these reports are a different, more heterogeneous population than the experienced US sample Zahller et al. studied. And the sectors where sustainability failures carry real financial consequences look nothing like the forestry industry the original study examined.

This raises a direct question: does disclosure quality still build the legitimacy that protects a firm today, under mandatory reporting, among the retail investors now driving European markets, and in a sector where the stakes are financial rather than reputational? And does that protection depend on who is doing the investing?

This thesis answers that question through a controlled experiment: a 2×2 design manipulating the accuracy and completeness of a semiconductor firm's CSR report, tested on retail investors before and after an exogenous sector-wide shock, with individual moral sensitivity tested as a possible moderator. Section 2 builds the theoretical case for why disclosure quality should operate through legitimacy rather than directly, and why each contextual shift (regulatory, sectoral, demographic) gives reason to doubt the original pathway. That reasoning becomes a testable design in Section 3. Section 4 reports the data and Section 5 asks where the pathway held up, and where it did not by setting the limitation.

2. Literature review

One claim organises this literature review, and it is worth stating up front. CSR disclosure quality does not change investor behaviour on its own. It matters because quality shapes something intermediate: how legitimate investors perceive the firm to be, and that perception is what protects the firm when a crisis later hits. This is the reading adopted here, rather than a direct effect account of disclosure quality. It is also a reading that rests on a single study, run a decade ago under conditions (a voluntary disclosure regime, a lower-materiality sector, a narrower investor population) that no longer describe the market this thesis studies. Whether the same mechanism holds once those conditions change, and whether it holds equally for every investor, is what the rest of this chapter works through, one link at a time.

2.1 Disclosure Quality as a Signal

Before asking why disclosure quality builds legitimacy (Section 2.2), a prior question needs answering. How does a CSR report reach investors judgment at all? Investors cannot observe a firm's actual sustainability practices, they only see what the firm chooses to disclose. This gap is where signalling theory becomes useful: it explains which kinds of disclosure investors can actually trust, and which they should treat as cheap talk.

Because investors cannot verify a firm's internal practices, the market risks a "lemons dynamic" (Akerlof, 1970), where superficial greenwashing looks the same as genuine commitment. Signalling theory (Spence, 1973) resolves this. A signal is credible only when it is costly to produce and hard for poor performers to fake. A CSR report with precise, externally verifiable data meets that bar. Producing it takes real effort, and the precision itself creates risk: specific, verifiable claims can later be checked and contradicted, which vague claims cannot (Connelly et al., 2011).

This matters specifically for retail investors, who lack the resources for deep fundamental analysis and rely instead on the affect heuristic, using their immediate emotional reaction to a disclosure (“does it feel credible or not”) as a substitute for the deeper financial analysis they lack the time or expertise to do themselves (Slovic et al., 2007; Elliott et al., 2014). A vague report leaves investors unable to verify anything the firm claims about itself, so the gap between what the firm knows and what investors can check remains, and triggers distrust. A precise report narrows that gap and builds a positive anchor that becomes legitimacy (Lock & Schulz-Knappe, 2019). The stakes are real. When disclosed information is later contradicted, the resulting collapse in credibility is costly for the firm (Backof et al., 2020).

This raises the operational question the present study is built around: what, concretely, makes a disclosure a credible signal rather than cheap talk? The literature converges on two answers. Accuracy is the use of specific, verifiable, quantitative metrics "hard to fake", reassuring for the same reason a costly signal is reassuring (Connelly et al., 2011); external assurance of such claims strengthens their credibility (Hoang & Trotman, 2021). Completeness is the breadth of what is disclosed, including negative results alongside positive ones, credible because a firm willing to admit shortfalls is not simply managing impressions (O'Dwyer et al., 2005). A report can be narrow but precise, or broad but vague (Hummel & Schlick, 2016), which is why the two dimensions are kept apart here rather than folded into one quality score.

But this reasoning assumes a voluntary disclosure world: it says accuracy and completeness are credible because they are costly to produce. Since 2024, EU firms have been legally required to report comprehensively under the Corporate Sustainability Reporting Directive (CSRD). A signal is only as credible as it is costly to fake, and a legal requirement is, by definition, no longer a costly choice. If completeness is now mandatory rather than discretionary, it may no longer signal much at all. While accuracy, which the law does not standardise, should keep its full signalling value. Section 4 tests this empirically.

Together, as credible signals, accuracy and completeness are what an investor has to go on when forming a judgment about the firm. The judgment (not the disclosure itself) is legitimacy, and it is what next sections examines.

2.2 Building Organizational Legitimacy

Accuracy and completeness give the investor a credible signal. The question now is what they do with it. The answer this thesis relies on is legitimacy, a report is not just believed or disbelieved. It shapes a broader judgment about whether the firm itself deserves trust.

Suchman (1995) defines organizational legitimacy as a "generalized perception that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms" (p. 574). A perception held by observers, not a property the firm possesses on its own. This perception can rest on different grounds, and two matter most for retail investors: does the firm serve their direct interests (what Suchman calls pragmatic legitimacy), and is it "doing the right thing" regardless of personal gain (moral legitimacy). This is why a credible CSR signal does more than inform: it also signals values. A firm whose report is accurate and complete looks like a firm that shares the investor's own norms around transparency, and it is this sense of shared values, not the information itself, that legitimacy is meant to capture.

This is also why legitimacy is measured before the shock in this study's design (Section 3.3): Suchman's own description of it as a "reserve of goodwill" implies it must already exist when a crisis hits, in order to act as a buffer against it, which is the mechanism the next section develops.

2.3 The Reputational Buffer: From Legitimacy to Resilience

The idea that legitimacy protects a firm during a crisis does not originate in accounting or finance. Zahller et al. (2015) borrow it from sociology, adapting Adger's (2000) concept of “Social Resilience”. Which he defines as the capacity of communities to withstand external shocks without disrupting their social structures and institutions. Under Adger's framework, legitimate institutions are granted more social and economic slack by their community when trouble hits, precisely because that legitimacy lowers the perceived risk of continuing to support them. Zahller et al. transpose this directly into financial markets: the "community" becomes investors, and resilience becomes their willingness to extend patience. Investors, as a result, are more likely to give a legitimate firm the benefit of the doubt during market turbulence instead of exiting at the first sign of trouble. Because no such measure existed in the CSR literature before their study, they had to build this construct from scratch, which is itself one reason a replication of it is worth attempting.

A reserve is only worth the name if it pays out when needed. Does pre-crisis legitimacy actually reduce the damage once a shock hits? The archival evidence points that way. Firms with strong pre-crisis CSR engagement suffered smaller market penalties after the Bhopal disaster (Blacconiere & Patten, 1994), posted better returns during the 2008 financial crisis (Lins et al., 2017), and repeated the pattern through COVID-19 market volatility (Lu et al., 2024). Krüger (2015) adds a finding worth noting here: a negative CSR revelation destroys more value than an equivalent positive action creates. Good and bad news do not weigh the same. This is why the reserve legitimacy builds matters most as protection against a sudden drop, not simply as an accumulation of small gains.

However this protective effect is not automatic. Hickman and Wong-On-Wing (2024) show that CSR can work against a firm when a crisis strikes precisely the area where it had made the strongest claims. Investors read the prior disclosure as hypocritical rather than reassuring, and penalise the firm more harshly than if no disclosure had been made at all. The buffer this thesis tests is therefore conditional, not guaranteed. Legitimacy protects a firm from a shock it did not

cause, but does not protect a firm from being seen to have failed at what it claimed to be good at.

But archival data cannot show why a shock spreads to firms not directly responsible for it in the first place. This matters directly for the present design (the fictional firm and shock scenario are detailed in Section 3.3). The firm at the centre of that scenario, TechCore, is explicitly cleared of wrongdoing yet still loses value. The reason lies in contagion: reputational damage from a sector-wide scandal spreads to associated firms regardless of their actual involvement, simply because they operate in the same industry (Jonsson et al., 2009). The design in Section 3.3 rely exactly on this: TechCore is cleared of wrongdoing but its stock still falls, which separates sector-wide contagion from actual culpability, the one condition under which legitimacy's protective role can be tested cleanly.

Zahller et al. (2015) connect these pieces into a single behavioural chain: a firm with accurate, complete pre-crisis disclosure builds legitimacy, and that legitimacy determines whether investors read a later shock as situational (something that happened to a trustworthy firm) or as evidence of deeper managerial failure. Investors who judge the firm legitimate are more likely to extend patience, tolerate a slower recovery, and hold their position rather than sell. This perception-to-behaviour chain (quality → legitimacy → resilience) is the mechanism this thesis tests.

Testing it well, however, requires a shock that actually threatens the firm. Which is why this study moves away from the forestry sector Zahller et al. (2015) examined, toward the semiconductor industry. Long, opaque supply chains expose semiconductor firms to significant ESG risks (Oyetade et al., 2025), including forced labour and conflict minerals concerns that carry direct financial consequences, not just reputational discomfort. If the shock is too mild, investors may stay simply because there is little reason to leave, and legitimacy's protective role would never really be tested. A financially consequential sector like semiconductors avoids that problem, it puts real pressure on investors to leave, which is exactly the pressure legitimacy is meant to resist.

2.4 The Missing Link: Moral Awareness

Sections 2.1 and 2.2 treat investors as a single, uniform audience: the same disclosure produces the same shift in legitimacy for everyone who reads it. That assumption is where Zahller et al.'s (2015) model stops, and it is worth questioning directly, because the population of retail investors it was built for no longer looks the way it did a decade ago.

European retail trading has grown sharply since 2020: retail-specific transactions nearly tripled in market share, from 1.5% in January 2019 to 4.5% at their post-COVID peak (Aramian & Comerton-Forde, 2023), driven by mobile trading apps and zero-commission platforms. This matters for the pathway tested here because this expanded population is not homogeneous in how it weighs ESG information. Some retail investors are strongly motivated by stakeholder-capitalism values and willing to trade return for stronger environmental and financial disclosure, though not for social disclosure specifically (De Villiers et al., 2023), while others process the same information primarily through financial heuristics. If a single average effect of disclosure quality on legitimacy conceals this variation, the model needs a way to test for it directly, rather than assuming it away.

The candidate this thesis uses is Moral Awareness: an individual's sensitivity to the ethical and social consequences of their own decisions, rooted in Schwartz's (1977) Norm Activation Model. An investor attuned to the ethical stakes of where their money goes has more reason to weigh a report's precision and honesty than one who reads the same report purely as a financial-risk signal. Nafisa et al. (2023), who developed and validated the scale used here, show that retail investors' baseline moral sensitivity shapes how strongly they respond to a firm's ESG performance and practices, moderating this relationship differently across environmental, social, and governance dimensions.

The accuracy/completeness-legitimacy relationships should not be uniform across investors, but stronger for those with higher Moral Awareness. The population for whom a precise, honest report says something that matters to their evaluation of the firm. Testing this also answers a gap Zahller et al. (2015) themselves flagged, their model had no way to account for investor characteristics at all.

2.5 The Zahller et al. (2015) Study and the Present Replication

Zahller, Arnold, and Roberts (2015) are the only ones to have tested this mechanism experimentally. Using a 2×2 design with 100 US investors, they manipulated CSR disclosure accuracy and completeness and traced their effects through legitimacy to social resilience after an exogenous shock. Accuracy showed a strong path to legitimacy, completeness's effect was weaker and did not reach significance. Misiuda and Lachmann (2022), reviewing 27 experimental studies on investor perception of sustainability reporting, confirm that this remains the only experimental test of disclosure precision to date. Meaning the entire

mechanism built in Sections 2.1 to 2.3 rests on a single study, run in one country, one sector, and one regulatory regime, a decade ago.

Three conditions from that study no longer hold. The regulatory environment has shifted from voluntary to mandatory disclosure, which is exactly what section 2.1 argued should change what completeness signals. The sector (forestry) is one where CSR failures are reputationally damaging but not always financially material. By contrast, the semiconductor context described in section 2.3 makes the test more robust. The underlying risk of this sector carries direct financial consequences, giving legitimacy's protective role a stronger challenge to prove itself against. Finally, the population (experienced US investors under a voluntary regime) does not reflect the more heterogeneous retail investor base that Section 2.4 shows has since emerged in Europe, nor does the original model have any way to account for that heterogeneity. Since it treats every investor as reacting identically to a given disclosure.

Each gap points to the same question, in the order the mechanism was built. First, whether disclosure quality still builds legitimacy once broad disclosure is legally required rather than a discretionary signal. Then whether that legitimacy still buffers behaviour when the underlying risk is financially real, not just reputational. And finally, whether the buffer's strength depends on how morally sensitive the investor is to what is being disclosed. Section 3 turns each of these questions into a specific, testable prediction, with a formal null and alternative hypothesis for every relationship this chapter has argued should hold.

3. Methodology

3.1 Research Design

The three questions posed previously cannot be answered by argument alone. They need to be tested, under the specific conditions the literature review argued should matter (mandatory disclosure, a financially material sector, and a heterogeneous investor population). This is what the design below is built to do. A 2×2 between-participant factorial experiment, crossing Accuracy and Completeness at two levels (High vs. Low) to produce 4 experimental conditions, with legitimacy measured before an exogenous shock and resilience measured after it.

A between-participant design was chosen so that each participant sees only one version of the CSR report, preventing any carryover or comparison bias across conditions. Participants were randomly assigned to one of the four conditions. Financial performance and the nature of the exogenous shock were held constant across all groups, isolating the causal effect of disclosure quality on investor perceptions.

Like any controlled experiment, the design involves a trade-off between internal and external validity. Ecological validity is preserved by grounding the scenario in a Financial Times-style vignette and by recruiting actual non-professional investors rather than a student convenience sample.

Two methodological adaptations from Zahller et al. (2015) are introduced. First, the analytical strategy replaces the original SEM with composite scores combined with ANCOVA and Hayes' PROCESS macro. The shift is required by the introduction of a continuous moderator (SEM does not easily accommodate continuous moderators without a much larger sample) and follows Hickman and Wong-On-Wing (2024), who use the same 2×2 between-participant design and PROCESS macro with bootstrapping in their extension of the Zahller framework. Second, the strict one-strike exclusion protocol applied by Zahller et al. (2015) is replaced by a two-strike design, for three reasons: academic rigour (the memory-first question still tests true encoding), ecological validity (a real investor retains access to their documents), and platform compliance (Prolific's Attention and Comprehension Check Policy requires at least two opportunities to answer correctly).

Table 1 maps the four hypotheses tested in this design, with their null and alternative statements alongside the chosen statistical methods. Detailed breakdowns of the constructs and tests follow in Sections 3.6 to 3.8.

Table 1: Summary of Hypotheses

Hyp.	Statement	H_0 / H_1	Test
H1	Accuracy has a positive effect on resilience	$H_0 : \beta_1 = 0$ $H_1 : \beta_1 > 0$	ANCOVA
H2	Completeness has a positive effect on resilience	$H_0 : \beta_2 = 0$ $H_1 : \beta_2 > 0$	ANCOVA
H3a	Disclosure quality raises legitimacy (path a)	$H_0 : a_1 = 0$ $H_1 : a_1 > 0$	Mediation analysis (path model)
H3b	Legitimacy raises resilience (path b)	$H_0 : b_1 = 0$ $H_1 : b_1 > 0$	Mediation analysis (path model)
H3	Legitimacy mediates disclosure quality → resilience	$H_0 : a_1 \times b_1 = 0$ $H_1 : a_1 \times b_1 > 0$	Bootstrapped indirect effect (5,000 resamples)
H4a	Moral Awareness moderates accuracy → legitimacy	$H_0 : \beta_3 = 0$ $H_1 : \beta_3 > 0$	Regression (interaction term)
H4b	Moral Awareness moderates completeness → legitimacy	$H_0 : \beta_3 = 0$ $H_1 : \beta_3 > 0$	Regression (interaction term)

Note. H1 and H2 are tested via a two-way ANCOVA controlling for pre-shock investment rating (Section 3.6). H3a and H3b are tested jointly via a mediation path model estimated with Hayes' PROCESS macro (Model 4), H3 refers to the significance of the bootstrapped indirect effect $a_1 \times b_1$ (Section 3.7). H4a and H4b are tested in two separate regression models with an interaction term, for accuracy and completeness respectively (Section 3.8).

3.2 Participants and Recruitment

Participants were recruited through Prolific, a platform well suited for reaching non-student, investment active populations. Eligibility was restricted to individuals who actively manage personal investment or retirement accounts but hold no professional role in finance, consistent with the non-professional investor framework of Zahller et al. (2015). Combined with random assignment, this criterion limits sampling bias and strengthens external validity. While UK-based and outside the direct jurisdictional scope of the EU's CSRD, these participants operate within a parallel market field shaped by equivalent mandatory norms (UK SDR).

A pilot study (N = 25) was conducted prior to the main data collection to stabilise the questionnaire, verify the estimated completion time (6 minutes), and check the manipulation and attention.

Sample size was determined through two a priori power analyses conducted in G*Power. For H1 and H2 (Table 1), a two-way ANCOVA with one covariate, medium effect size (Cohen's f = 0.25), α = .05, and power = .80 yielded a minimum required sample of 128 participants. For H4, a linear multiple regression with one tested interaction term and three total predictors, assuming a small-to-moderate effect size ($f^2 = 0.06$), α = .05, and power = .80, yielded a minimum of 133 participants. The more conservative threshold of 133 was therefore adopted

as the target. To account for expected attrition from the two-stage screening protocol, 140 participants will be recruited through Prolific.

Data quality is enforced through a two-stage screening protocol: a one-strike attention check (immediate exclusion on failure) and a two-strike design for the three manipulation checks (warning and stimulus reload on a first error, exclusion on a second). The final usable sample after exclusions was N = 90.

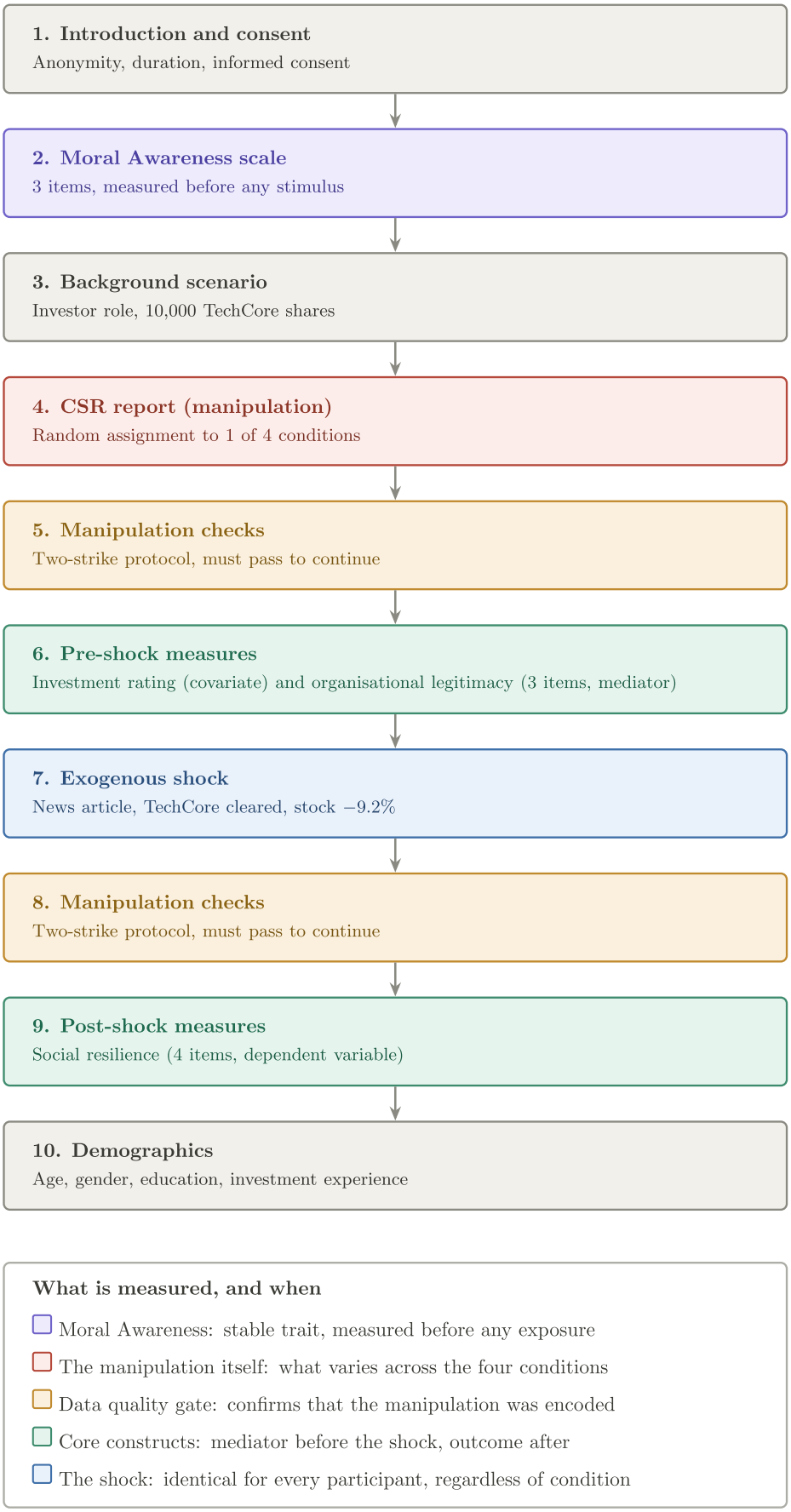
3.3 Experimental Procedure

The experiment was programmed and hosted using oTree, a Python-based platform designed for behavioural experiments, and closely follows the sequential structure used by ZAR. Figure 1 summarises the ten steps participants went through, from initial consent to the final demographic questions, and indicates at which point each key measure was taken relative to the manipulation and the shock.

Two points of this sequence matter for the analyses that follow. First, Moral Awareness is measured before participants see any information about TechCore, so that it reflects a stable personal trait rather than a reaction to the experimental material (this is discussed further in section 3.8). Second, the manipulation check in steps 5 and 8 acts as a quality gate rather than a measure itself. Only participants who correctly identified the content of the report they had just read were allowed to continue, following the two-strike protocol (described in section 3.2).

Three constructs are used to test H1 through H4: the pre-shock investment rating, organisational legitimacy, and social resilience. Each is described in detail where it is first used (sections 3.6, 3.7, and 3.8 respectively).

Figure 1: Experimental workflow



3.4 Manipulations

Crossing Accuracy and Completeness produces four report versions in a 2×2 between-subjects design (Figure 1, step 4): participants are randomly assigned to read one of High-Accuracy/High-Completeness, High-Accuracy/Low-Completeness, Low-Accuracy/High-Completeness, or Low-Accuracy/Low-Completeness. Each version is adapted from the original CSR report materials used by Zahller et al. (2015), updated to a semiconductor-sector, post-CSRD context. The full text of all four report versions is provided in Appendix A.

Accuracy is manipulated through the precision and verifiability of the report's language, independently of which topics are covered. In the High Accuracy condition, claims are backed by concrete, quantitative metrics ("audited 87% of Tier-1 suppliers in 2025, with 12 non-conformities identified and corrected"). In the Low Accuracy condition, the same underlying claim is stated in vague, unverifiable terms ("significantly increased supplier audits"), with no figure a reader could check. Both versions cover the same topics; only the verifiability of the language differs.

Completeness is manipulated through the breadth of topics disclosed and whether negative information is included. In the High Completeness condition, the report addresses a wide range of non-financial areas and explicitly acknowledges shortfalls alongside successes, such as a “missed diversity target” or “a minor safety-compliance lapse”. In the Low Completeness condition, the report is selectively positive. It covers fewer topics and omits negative indicator. Both versions use language of comparable specificity (only the scope and valence of the content differ.

Whether participants actually perceived these manipulations as intended is verified through the manipulation-check items and the two-strike protocol.

3.5 Summary of Measures

Each construct is described in detail in the section where it is tested. Table 2 provides a summary of all measured variables, including their timing, scales, and response anchors.

Table 2: Summary of Measures

Construct	Role	Items → Variables	Scale	Aggregation
Moral Awareness	Moderator	3 → 3	7-pt Likert (Strongly Disagree – Strongly Agree)	Mean (mean-centred)
Pre-Shock Investment Rating	Covariate	1 → 1	7-pt (Extremely Poor – Excellent Investment)	Raw score
Organisational Legitimacy	Mediator	3 → 3	7-pt, anchors vary by item[a]	Mean
Social Resilience	Dependent Variable	6 → 4[b]	Mixed (7-pt, %, time categories)[c]	Mean of z-scores

[a] Item-specific anchors (credibility, stability, integrity) are detailed in Section 3.7.
[b] The six raw items are combined into four analysis variables; see Section 3.6 for the construction procedure.
[c] Full scale definitions and coding are provided in Section 3.6.

3.6 Testing H1 and H2: Direct Effects on Resilience

H1 and H2, summarised in Table 1, ask a simple question: does reading a higher-quality CSR report make investors more resilient once a crisis hits, regardless of the mechanism behind it? H1 concerns accuracy, H2 concerns completeness.

What resilience means here, and how it is measured

After the shock (step 7), the experiment needs to capture how investors react to it (step 9), not just how they feel. Four concrete behaviours are asked about, each targeting a different part of what "staying resilient" could mean in practice.

- Item 1: Post-shock investment rating. "How would you now rate TechCore Industries as an investment opportunity?" (1 = Very Unfavourable to 7 = Very Favourable). This captures whether the participant's overall evaluation of TechCore has changed after the shock.
- Item 2: Disinvestment intention. "What percentage of your TechCore shares would you move into a safe S&P 500 index fund?" (slider: 0% = Keep all shares in TechCore to 100% = Move everything to the index fund). Because a higher percentage means more divestment, and the construct is built so that higher values always mean more resilience, this item is reversed before use (100% minus the percentage divested).
- Item 3: Partial recovery buffer. Participants are asked two questions about a scenario in which TechCore recovers at least 10% of the decrease in Net Income following the shock: (a) "How long do you expect this recovery to take?" and (b) "What is the

maximum time you would be willing to wait before divesting?". The buffer is the difference between the maximum waiting time and the expected recovery time.

- Item 4: Full recovery buffer. The same two questions are asked for a complete recovery to the pre scandal level. The buffer is computed identically.

For the two buffer items, positive values indicate remaining patience (the investor is willing to wait longer than they expect recovery to take), and negative values indicate that expected recovery exceeds the investor's willingness to wait.

These four items are measured on different scales (a 7-point rating, a percentage, and two buffer scores derived from categorical time responses) so they cannot be combined directly. Each item is therefore standardised into a z-score, indicating how far a participant's response lies above or below the sample mean in standard deviation units rather than in its original metric. The Social Resilience score was computed as the average of the four items.

Testing whether disclosure quality changes resilience

The most direct way to test H1 and H2 is to compare average social resilience scores across the four experimental conditions. If accuracy matters, participants exposed to high-accuracy CSR reports should report higher resilience than those exposed to low-accuracy reports. The same logic applies to completeness.

However, Social Resilience is measured only after the shock. Post-shock differences may therefore reflect not only participants' reaction to the shock, but also their prior tendency to evaluate TechCore favourably or unfavourably.

To account for this individual variation, the pre-shock investment rating (step 6) is included as a covariate. This covariate is not a perfect control. Following ZAR, it was measured after participants had already been exposed to the manipulated CSR report, although before the shock. The measure is therefore pre-shock but not pre-treatment. It may capture not only stable individual attitudes towards TechCore, but also participants’ immediate reaction to the CSR information. This creates a trade-off. Including it may remove part of the treatment effect that this thesis aims to estimate, excluding it risks leaving unexplained individual differences in the model. We chose to include the covariate, so its coefficients can be compared directly against ZAR's original estimates. But as a robustness check, the same analysis is re-estimated without the covariate in Appendix E, to verify that the results do not depend on this modelling choice.

The test used is therefore an ANCOVA (analysis of covariance). It works like a standard comparison of group averages (an ANOVA) but first removes the variation in resilience that can be predicted from the pre-shock rating (the covariate), and only then compares what remains across the four conditions. This reduces the noise against which the group differences are evaluated. This is what makes it more precise than a simple group comparison: by accounting for individual variation in starting position

Formally, the model estimated is:

$$Social\ Resilience\ = \beta_0 + \beta_1(Accuracy) + \beta_2(Completeness) + \beta_3(Accuracy \times Completeness) + \beta_4(Pre\text{-}Shock\ Rating) + \varepsilon$$

In this equation, β_1 captures the effect of accuracy on resilience once completeness and the covariate are held constant, and β_2 captures the effect of completeness in the same way. β_3 captures whether the two dimensions interact (for instance, whether a report needs to be strong on both dimensions at once to make a real difference). β_4 is the covariate adjustment described above.

H1 predicts a positive effect of accuracy: H_0: $\beta_1 = 0$ is tested against H_1: $\beta_1 > 0$. H2 predicts a positive effect of completeness: H_0: $\beta_2 = 0$ is tested against H_1: $\beta_2 > 0$. Significance is assessed at the conventional threshold of $p < .05$.

For the ANCOVA model to be valid, the relationship between the pre-shock rating and social resilience must be the same across all four conditions. This assumption is tested before interpreting the main results.

3.7 Testing H3: Does Legitimacy Mediate the Relationship?

H1 and H2 test whether disclosure quality affects resilience. H3, along with its two sub-components H3a and H3b (Table 1), asks whether this effect operates through legitimacy rather than directly. To test this, legitimacy must be measured at a specific point in the sequence (step 6), after participants have read the CSR report, but before the shock occurs. This timing is not arbitrary. If legitimacy is what creates the reputational buffer, it must exist before the shock hit. Measuring it afterwards would make it impossible to separate cause from consequence.

Measuring legitimacy

Organisational Legitimacy is measured through three adapted items on a 7-point scale. Zahller et al. draw these items from the broader trust and CSR literature:

- Item 1: Report credibility. "How believable do you find TechCore's sustainability report?" (1 = Very Doubtful to 7 = Very Believable). This captures pragmatic legitimacy: does the disclosure reduce the investor's uncertainty?
- Item 2: Holding comfort. "How comfortable would you be holding TechCore stock without monitoring it constantly?" (1 = Very Uncomfortable to 7 = Very Comfortable). This captures the reliability dimension: does the investor trust the firm enough to look away?
- Item 3: Management integrity. "How confident are you in the integrity of TechCore's senior management?" (1 = Not at all Confident to 7 = Very Confident). This captures moral legitimacy: does the investor believe management is acting ethically?

The composite Organisational Legitimacy score is the mean of the three items.

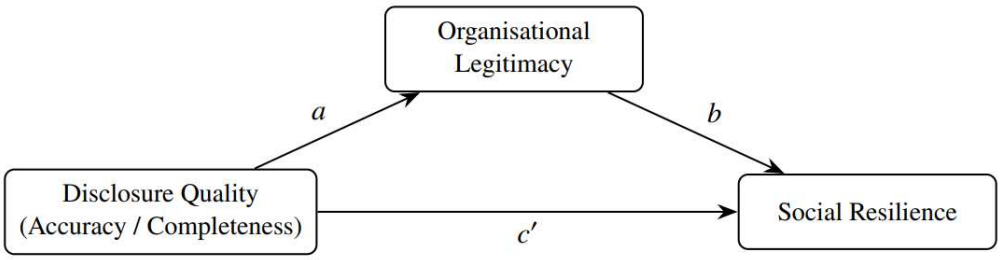
The mediation test

The ANCOVA in Section 3.6 can only say whether disclosure quality has an effect on resilience overall (not whether that effect passes through legitimacy). Testing that requires a model that puts legitimacy explicitly between disclosure quality and resilience, and asks whether it soaks up their relationship.

The mediation model estimates three relationships at once. Does disclosure quality predict legitimacy? Does legitimacy predict resilience, independently of which report was read? And does disclosure quality still predict resilience directly once legitimacy is accounted for? If this third relationship is not significant, legitimacy fully explains the link, this is called full mediation, and it is the pattern this thesis expects to find.

Figure 2 illustrates this structure: disclosure quality can reach resilience through an indirect path via legitimacy (a × b) or a direct path bypassing it (c').

Figure 2: mediation relationship



This structure translates into two equations estimated simultaneously:

$$Equation\ 1: Legitimacy = a_0 + a_1(Disclosure) + a_2(Pre - Shock\ Rating)$$

$$Equation\ 2: Resilience = b_0 + b_1(Legitimacy) + c'(Disclosure) + b_2(Pre - Shock\ Rating)$$

In Equation 1, a_1 captures whether disclosure quality predicts legitimacy (path a). In Equation 2, b_1 captures whether legitimacy predicts resilience once disclosure quality is held constant (path b), and c' captures any remaining direct effect. The indirect effect is the product $a_1 \times b_1$.

H3a predicts $a_1 > 0$ (H_0: $a_1 = 0$ vs. H_1: $a_1 > 0$), H3b predicts $b_1 > 0$ (H_0: $b_1 = 0$ vs. H_1: $b_1 > 0$). And H3 overall predicts a non-zero indirect effect (H_0: $a_1 \times b_1 = 0$ vs. H_1: $a_1 \times b_1 > 0$).

Testing the indirect effect requires a choice of method. ZAR used SEM (structured equation modelling), but SEM demands a larger sample than this study's design doesn't allows once a continuous moderator is introduced, an alternative is needed here. An older alternative, the Sobel test, approximates the significance of the indirect effect by assuming its sampling distribution follows a normal curve, an assumption known to be inaccurate in small samples. The approach used here, Hayes' PROCESS macro (Model 4), avoids this assumption entirely. The indirect effect is recalculated on 5,000 bootstrap resamples of the data, producing a range of plausible values instead of a single estimate. If this range excludes zero (meaning that even its lowest plausible value is still different from zero) the indirect effect is considered statistically different from zero, and H3 is supported. This is consistent with Hickman and Wong-On-Wing (2024), who use the same method (ANOVA & Hayes) in their extension of ZAR's framework.

The pre-shock investment rating is included as a covariate in both equations. It captures individual differences in how participants evaluated TechCore as an investment before either measure was taken, so that the estimated effect of disclosure quality is not confounded with these prior differences. As in Section 3.6, this covariate is not a perfect control, the same robustness check (re-estimating the model without it) is reported in Appendix E.

3.8 Testing H4: Does Moral Awareness Moderate the Effect?

H4, tested here as two separate models (H4a and H4b, Table 1), asks whether the first stage of the mediation model (the path from disclosure quality to legitimacy) depends on investor moral sensitivity. If so, the same high-quality report would produce a larger shift in perceived legitimacy for investors who care more about the ethical consequences of their decisions than for those who do not.

Measuring Moral Awareness

Moral Awareness is measured before participants see any information about TechCore (step 2), before even the financial summary. This timing is deliberate: it introduces a temporal and proximal separation between the moderator and the experimental stimuli, ensuring that scores capture a stable predisposition rather than a situational reaction (Podsakoff et al., 2012).

Three items are asked on a 7-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree), adapted from Nafisa et al. (2023), who developed and validated this scale to measure the moral sensitivity of retail investors towards the ESG practices of the companies they hold, grounded in the Norm Activation Model (Schwartz, 1977)

- Item 1: Awareness of consequences. "When companies in my portfolio are linked to human rights violations in their supply chains, I believe real people are harmed."
- Item 2: Ascription of responsibility. "As an investor, I bear some personal responsibility for the social practices of the companies I hold."
- Item 3: Personal moral norm. "I have a moral duty to consider ESG issues before making an investment decision."

The composite Moral Awareness score is the mean of the three items.

The moderation test

Unlike Accuracy and Completeness, which each take only two values, Moral Awareness is measured on a 7-point Likert scale and averaged across three items, producing a much wider range of possible values. Although derived from Likert items, the three-item mean is treated as continuous for analysis, as is standard practice for multi-item scales, avoiding the loss of information that splitting participants into artificial high/low groups would cause.

Before computing the interaction, Moral Awareness is mean-centred (the sample average is subtracted from each score), so that zero represents an average level of moral sensitivity rather than an absence of it. As a result, β_1 (Disclosure) is read as the effect of disclosure quality on legitimacy for a participant with average moral sensitivity, not zero.

The model estimated is:

$$Legitimacy = \beta_0 + \beta_1(Disclosure) + \beta_2(Moral\ Awareness, centred) + \beta_3(Disclosure \times Moral\ Awareness) + \beta_4(Pre - Shock\ Rating) + \varepsilon$$

H4 predicts that β_3 is positive. Formally, $H_0: \beta_3 = 0$ (the effect of disclosure quality on legitimacy is the same regardless of Moral Awareness) is tested against $H_1: \beta_3 > 0$ (the effect

is stronger for investors with higher Moral Awareness). The gap in perceived legitimacy between high- and low-quality disclosure grows as moral sensitivity increases ($p < .05$). Two versions of the model are estimated, one with Disclosure representing accuracy (H4a) and one for completeness (H4b). The pre-shock investment rating is included as a covariate in both for consistency.

4. Empirical Analysis and Results

4.1 Sample Description

Data were collected in two batches through Prolific in June 2026. Of the 102 participants who accessed the study, 90 completed it and were approved (88.2%): 5 were rejected (3 for failing the manipulation checks twice, 2 for abnormally short completion times), 6 returned the study voluntarily, and 1 timed out. No further exclusions were applied, yielding a final analytic sample of N = 90.

The pre-specified target of N = 133 was not reached. The decision not to launch a third recruitment batch deserves transparent discussion. Three considerations informed this decision.

First, the Moral Awareness ceiling effect with lack of variance (M = 5.70, SD = 1.29, Mdn = 6) reflected a distributional constraint that additional participants were unlikely to resolve, the issue was variance, not volume. Second, the study's central theoretical prediction (H3) was already supported at N = 90, with bootstrapped indirect effects statistically significant. Third, the composite Social Resilience score exhibited lower internal consistency than expected (Cronbach's α = .61, full details in Appendix D). This means the four items don't move together consistently: they capture genuinely different things (an investment rating, a divestment percentage, two recovery buffers) rather than one unified construct. Averaging them into a single score, as H1 and H2 require, can therefore hide real effects that exist at the item level rather than at the composite level (a possibility examined directly in Section 4.5).

Recruitment carried a financial cost: £227.90 in total for N=90 (£162.80 in participant rewards, £54.25 in Prolific platform fees, £10.85 in VAT)[1]. Launching a third batch to reach the pre-specified N = 133 would have added a proportional cost for a return that the three considerations suggest would be low. It would not resolve the ceiling effect, would not strengthen an already-robust H3, and would not fix a construct problem. The right response to α = .61 is construct refinement, not more participants.

Table 3 reports the full sample characteristics. The sample was predominantly male (77.8%), White (84.4%), and employed full-time (85.6%), with high educational level and moderate investment experience.

[1] This data collection was made possible thanks to funding from the FNRS

Table 3: Participant Demographics (N = 90)

	n	%		*n*	%
Age (years)			**Education**		
M (SD) = 35.2 (9.53)	–	–	GCSE or equivalent (CE1D)	7	7.8
Range = 19–61	–	–	A-Levels or equivalent (CESS)	19	21.1
			Bachelor's degree	40	44.4
Sex			Master's degree	21	23.3
Male	70	77.8	Doctorate or higher	3	3.3
Female	20	22.2			
			Investment Experience		
Ethnicity			Very little	17	18.9
White	76	84.4	Moderate	51	56.7
Black	6	6.7	Considerable	22	24.4
Asian	3	3.3			
Mixed	3	3.3	**Employment Status**		
Other	2	2.2	Full-Time	77	85.6
			Part-Time	11	12.2
			Other/Unemployed	2	2.2

Note. *n* = number of participants; % = percentage of total sample (N = 90).

4.2 Manipulation Checks

Manipulation check performance was satisfactory. On the accuracy check, 87% of participants correctly identified their condition on the first attempt; on the completeness check, 89%; and on the scenario check, 90%. Strike rates were largely independent across checks, only 3 participants failed both CSR checks simultaneously, suggesting that encoding difficulty was not condition dependent. Overall, 67 participants (74.4%) passed all three checks without any stimulus reload. A subsample of 71 participants who required no reload on either CSR check (a broader group than the 67 who also passed the scenario check) is used in Appendix D.

Following the two-strike protocol, all 90 retained participants had correctly encoded the manipulation prior to the pre-shock measurement. The attention check was passed by all 90 participants, indicating consistent engagement throughout the study.

4.3 Descriptive Statistics and Correlations

Before running any formal test, it is worth checking whether the raw data already point in the direction the hypotheses predict, simply by looking at how three average scores behave across the four conditions: legitimacy (how trustworthy the firm appears before the shock), resilience (how willing investors are to stay invested after it), and the pre-shock investment rating used as a covariate. Group sizes varied slightly across the four conditions (n = 20–27, Table 4), consistent with random assignment ($\chi^2(1) = 0.37$, $p = .544$).

H1 and H2 predict that a higher-quality report should raise resilience, condition by condition. Here the data don't cooperate, the average score on our four resilience items is nearly flat across the four conditions, with no clear ordering by accuracy or completeness.

H3a predicts that a higher-quality report should raise perceived legitimacy instead. This is exactly what is observed: legitimacy decreases step by step as the report becomes less accurate and less complete (Table 4). Legitimacy moves with disclosure quality, resilience does not.

Table 4: Means and Standard Deviations by Condition

Condition	n	Legitimacy M (SD)	Resilience M (SD)	Initial Quality M (SD)
High Accuracy / High Completeness	27	5.42 (0.74)	0.11 (0.46)	5.37 (0.63)
High Accuracy / Low Completeness	20	5.12 (0.73)	0.17 (0.75)	5.65 (0.75)
Low Accuracy / High Completeness	21	4.70 (1.39)	−0.16 (0.52)	5.33 (1.06)
Low Accuracy / Low Completeness	22	4.11 (1.16)	−0.14 (0.92)	5.05 (1.00)

Note. M = mean. SD = standard deviation. Initial Quality = pre-shock investment rating (covariate).

H3b predicts that legitimacy, once formed, should translate into resilience. Table 5 supports this: legitimacy and resilience are positively correlated ($r = .40$, $p < .001$), participants who found the firm more legitimate before the shock also proved more resilient after it.

H4 predicts that Moral Awareness should shape how strongly disclosure quality affects legitimacy. Here Table 5 shows nothing: Moral Awareness is essentially uncorrelated with both legitimacy and resilience. This flat pattern likely reflects the ceiling effect discussed in Section 4.1.

Table 5: Zero-Order Correlations

Variable	1	2	3	4
1. Initial Quality	_			
2. Legitimacy	.54***	_		
3. Resilience	.27*	.40***	_	
4. Moral Awareness	.13	−.09	−.01	_

**p < .05, ***p < .001*

4.4 Hypothesis Testing

4.4.1 H1: Does accuracy increase resilience?

Participants who read a more accurate CSR report are expected to show greater resilience after the shock than those who read a vague one. In the model from Section 3.6:

$$Social\ Resilience = \beta_0 + \beta_1 \cdot Accuracy + \beta_2 \cdot Completeness + \beta_3 \cdot Accuracy \times Completeness + \beta_4 \cdot Pre - Shock\ Rating + \varepsilon$$

β_1 is the coefficient of interest: it captures how much the (standardised) resilience score shifts, on average, when moving from a low to a high-accuracy report. Formally, H_0: $\beta_1 = 0$ is tested against H_1: $\beta_1 > 0$.

If accuracy matters, β_1 should be positive and reach significance at $p < .05$.

$\beta_1 = 0.23$, meaning that moving from a low to a high accuracy report shifts the standardised resilience score by 0.23 standard deviations, on average positive as expected, but not significant ($F(1, 85) = 2.64$, $p = .108$). The coefficient points in the right direction but is too weak to reject H_0. H1 is not supported.

4.4.2 H2: Does completeness increase resilience?

H2 predicts the same pattern for completeness: a positive β_2. Results show no significant effect ($\beta_2 = -0.04$, $F(1, 85) = 0.08$, $p = .776$), unlike accuracy, the coefficient does not even point in the expected direction. H2 is not supported.

The interaction between accuracy and completeness ($\beta_3 = 0.08$, $p = .780$) was also non-significant, indicating the two dimensions do not combine to produce a disproportionate effect. Full ANCOVA output is reported in Table 6.

Table 6: ANCOVA Results for Social Resilience

Source	β	SS	df	F	p	η2p
Accuracy	0.23	1.15	1	2.64	.108	.05
Completeness	−0.04	0.04	1	0.08	.776	<.001
Accuracy × Completeness	0.08	0.03	1	0.08	.780	<.001
Initial Quality (covariate)	0.19	2.29	1	5.26	.024	.06
Residuals	_	36.98	85			

4.4.3 H3: Does legitimacy explain why disclosure quality matters?

H1 and H2 asked whether disclosure quality affects resilience directly. H3 asks a different question: does it work indirectly, by first shaping how legitimate investors perceive the firm to be? In the two-equation model from Section 3.7:

$$Legitimacy = a_0 + a_1 \cdot Disclosure + a_2 \cdot Pre - Shock\ Rating$$

$$Resilience = b_0 + b_1 \cdot Legitimacy + c' \cdot Disclosure + b_2 \cdot Pre - Shock\ Rating$$

Three coefficients matter here. a_1 asks whether disclosure (accuracy or completeness) raises legitimacy (H3a: $a_1 > 0$). b_1 asks whether legitimacy then raises resilience (H3b: $b_1 > 0$). c' asks whether disclosure still has a direct effect once legitimacy is accounted for. If not, legitimacy fully explains the link, the pattern (full mediation) this thesis expects to find.

Accurate disclosure raises legitimacy ($a_1 = 0.70$, $p < .001$). H3a is supported. And legitimacy raises resilience ($b_1 = 0.20$, $p = .011$). H3b is supported. Once legitimacy is accounted for, accuracy no longer has any direct effect of its own ($c' = 0.09$, $p = .530$). The whole effect runs through the indirect effect ($a_1 \times b_1 = 0.14$), whose bootstrapped confidence interval, [0.02, 0.32], excludes zero. H3 is supported for accuracy, via full mediation: accuracy does not move resilience directly, it moves resilience by first building legitimacy.

For completeness, the same pattern holds, complete disclosure raises legitimacy ($a_1 = 0.50$, $p = .012$) and legitimacy raises resilience ($b_1 = 0.24$, $p = .002$). H3a and H3b are both supported. Again, no direct effect survives ($c' = -0.14$, $p = .324$), and the indirect effect (0.12, CI [0.02, 0.26]) excludes zero. H3 is supported for completeness, via full mediation.

This accuracy–legitimacy coefficient is the single most important number in this thesis, so it is worth comparing directly to Zahller et al. (2015), but only once it is on the same scale. A standardised β expresses an effect in standard-deviation units rather than raw scale points. The $a_1 = 0.70$ reported above is in raw units (points on legitimacy's original 7-point scale), standardised, it becomes β = .31. The figure comparable to ZAR's own reporting.

Zahller et al. (2015) found β = .335[2] for accuracy (nearly identical to the β = .31 found here). Completeness, by contrast, is stronger in this study (β = .22) than in their original estimate (β = .152). What this stability and this divergence imply is discussed in Section 5.1.

[2] One difference remains. ZAR estimated their path via SEM, which corrects for measurement error and tends to inflate coefficients relative to the regression used here. If anything, this makes the closeness of the two estimates below more striking, not less.

Table 7: Mediation Results (PROCESS Model 4)

		Accuracy model		Completeness model	
Path	**Coef.**	**value**	p	**value**	p
Disclosure → Legitimacy	a_1	0.70	<.001	0.50	.012
Legitimacy → Resilience	b_1	0.20	.011	0.24	.002
Direct effect	c'	0.09	.530	−0.14	.324
Indirect effect	$a_1 \times b_1$	0.14	CI [0.02, 0.32]	0.12	CI [0.02, 0.26]

4.4.4 H4: Does moral sensitivity change how strongly disclosure quality matters?

H4 predicts that the accuracy-legitimacy and completeness-legitimacy relationships tested above should be stronger for investors with higher Moral Awareness. From section 3.8:

$$Legitimacy = \beta_0 + \beta_1 \cdot Disclosure + \beta_2 \cdot Moral\ Awareness\ (centred) + \beta_3 \cdot Disclosure \times Moral\ Awareness + \beta_4 \cdot Pre - Shock\ Rating$$

β_3 is the coefficient of interest. If moral sensitivity matters, the legitimacy gap between a good and a poor report should widen among the investors who care most, that widening is what β_3 compute. Formally, H_0: $\beta_3 = 0$ is tested against H_1: $\beta_3 > 0$. Separately for accuracy (H4a) and completeness (H4b), since each dimension may interact differently with Moral Awareness. If Moral Awareness moderates the effect, β_3 should be positive and reach significance at $p < .05$.

Accuracy (H4a). $\beta_3 = 0.27$, not significant ($p = .135$), the coefficient points in the expected direction but falls well short of the threshold. H4a is not supported.

Completeness (H4b). $\beta_3 = -0.06$, not significant ($p = .710$), the coefficient does not point in the expected direction at all. H4b is not supported. Neither result should be read as firm evidence against moderation.

As discussed in Section 4.1, Moral Awareness showed a pronounced ceiling effect, leaving too little variance to detect an interaction even if one exists, the final sample ($N = 90$) also provided only about 60% power for this test, against a pre-specified target of $N = 133$. H4 is best read as untestable with the present data, rather than rejected, a distinction discussed in section 5.3.

4.5 Exploratory Item-Level Analyses

The composite Social Resilience score found nothing for H1 and H2. But that score averages four different things a participant does or reports (rate the investment, decide how much to divest, and estimate two recovery timeframes) and a null result on the average can raises an

obvious question. Is nothing really happening, or is something happening in only one or two of the four items, cancelled out once they are averaged together? This analyse look for that signal.

The same ANCOVA used for the composite score is re-run separately on each of the four items of Social Resilience. Two of the four items show something the composite did not. Accuracy affects Item 2 (disinvestment) in the expected direction: participants in the High Accuracy condition divested 34.5% of their position on average, versus 46.5% in the Low Accuracy condition (a 12-point difference), though not statistically significant at the conventional threshold ($p = .085$). Completeness affects Item 4 (the long-term recovery buffer) significantly ($p = .031$), but backwards: participants in the High Completeness condition showed a narrower buffer ($M = 0.29$)[3] than those in the Low Completeness condition ($M = 0.62$). Meaning less patience than a selectively positive report. Items 1 and 3 show no effect of either dimension.

That finding is followed up with the same mediation model used in Section 4.5.3 (PROCESS Model 4), now applied item by item rather than to the composite score. For Item 2, this is a natural extension of what H3 already established at the composite level. Since legitimacy fully mediates the effect of disclosure quality on Social Resilience overall, the same pattern should reappear on Item 2 specifically. Its own total effect fell short of significance (accuracy, $p = .085$), but this is not a problem for the mediation test: a mediator can still carry a real effect even when the total effect alone is too weak to detect. For Item 4, where completeness's total effect was already significant, it instead asks how that effect operates: whether it runs through legitimacy also, as the theory predicts, or bypasses it entirely.

For Item 2, the pattern holds exactly as the theory predicts: legitimacy's path to the outcome (path b) is significant in both the accuracy model ($b_1 = -10.52$, $p = .003$) and the completeness model ($b_1 = -10.99$, $p = .002$; full breakdown in Appendix F), no direct effect (path c') of disclosure quality survives once legitimacy is in the model, and the indirect effect excludes zero for both accuracy (−7.40, CI [−15.12, −1.84]) and completeness (−5.52, CI [−11.58, −0.94]).

Item 4 does not follow the same rule: legitimacy's path to the outcome (path b) is not significant in either model (accuracy model: $b_1 = -0.01$, $p = .90$; completeness model: $b_1 = 0.07$, $p = .44$). Accuracy has nothing further to test here, since it never had a total effect on Item 4 to begin with. Completeness is different, its direct effect (path c'), once legitimacy is included in the

[3] The buffer is measured in scale categories (Section 3.6 & Appendix C), not real time units. Because categories are not evenly spaced (e.g., moving from "one month" to "six months" is a different time gap than moving from "3–5 years" to "5+ years"), a difference of 0.29 vs. 0.62 categories should be read as a difference in patience level, not a precise duration.

model, remains significant ($c' = -0.36$, $p = .029$; Appendix F). Meaning completeness bypasses legitimacy entirely rather than working through it. So, two of the four items that make up "social resilience" are explained differently: disinvestment (Item 2) through the legitimacy pathway this thesis was built to test, the long-term recovery buffer (Item 4) through something else entirely. That split is explored further in Section 5.3.

These patterns are exploratory and were not planned. They point to questions worth testing properly in future research, not conclusions this thesis can draw on its own.

5. Discussion

5.1 The Mediating Role of Legitimacy

The central finding of this thesis is that the quality-legitimacy-resilience pathway identified by Zahller et al. (2015) replicates in the present context. Accuracy and completeness did not directly affect Social Resilience on the composite measure (H1, H2 not supported), but both increased perceived legitimacy, and legitimacy in turn increased resilience. Once legitimacy was included in the model, the direct effect of disclosure quality dropped to non-significance (H3 supported via full mediation). Far from contradicting each other, the two findings describe one mechanism from two angles. Disclosure quality does not move investor behaviour on its own. It matters because it builds the legitimacy that stabilises investor responses when a crisis occurs.

This absence of a direct effect should not, on its own, be read as evidence that disclosure quality does not matter. It should be read alongside two further considerations. First, Social Resilience was measured as an average of four behaviours that respond to disclosure quality through different mechanisms rather than a single one (Section 4.5), averaging can flatten real item-level effects into an apparently null composite result. Second, the final sample fell below the pre-specified target for the direct-effect tests, reducing the chance of detecting a modest total effect. So the safe reading is narrower: no direct effect showed up on the composite measure used here, which is not the same as none existing. Section 5.3 develops the distinction.

The replication is especially notable for accuracy. The accuracy-legitimacy coefficient found here ($\beta = .31$) is nearly identical to ZAR estimate a decade earlier ($\beta = .335$), despite a different sector, sample, and regulatory regime. This estimate also remained stable across the two design choices discussed in Section 3.7. The indirect effect holds whether or not the pre-shock covariate is included, and in a joint path model estimated with lavaan (Appendix E). That stability, across every contextual factor that changed and across these internal robustness checks, is the clearest evidence that the accuracy-legitimacy link is not an artefact of the original study's setting, but a mechanism that travels across contexts.

Completeness is more interesting. It is a stronger predictor of legitimacy here ($\beta = .22$) than in ZAR ($\beta = .152$). One plausible reading is that mandatory reporting changes what completeness signals. When broad disclosure is legally expected, completeness may matter less as a signal of coverage alone and more as a sign that the firm is willing to admit its own failures, especially when firms acknowledge weak or disappointing performance rather than presenting selectively

positive information. This study cannot separate the two empirically. Completeness here bundles topical breadth and the disclosure of negative information into a single manipulation, so this remains one plausible explanation among others for why completeness's effect has strengthened, not a finding this thesis can confirm on its own.

This is not, however, the whole story for completeness. Section 4.5 shows that completeness also has a direct effect, entirely outside the legitimacy pathway just described and, unexpectedly, in the opposite direction to what a simple "more disclosure builds more trust" account would predict.

5.2 Moral Awareness and the Limits of H4

H4 could not be meaningfully evaluated. Moral Awareness showed a pronounced ceiling effect, leaving too little variation to test whether it moderated the relationship between disclosure quality and legitimacy, a difficulty compounded by a sample underpowered for this specific test. The right conclusion is not that H4 was disproven, but that the study could not provide a fair test of it.

Two explanations remain possible, though this thesis cannot tell which is right. The first explanation is methodological. A self-reported scale about moral sensitivity is subject to socially desirable answers. Participants may simply present themselves as more ethically attentive than they actually are, regardless of what the study is officially about. The second is more speculative: if sustainability disclosure has become increasingly normalised in the post-CSRD environment, moral sensitivity to it may simply vary less across individuals than earlier theory assumed, consistent with Suchman's (1995) account of cognitive legitimacy. Where a practice becomes taken for granted rather than a differentiator. This second explanation is theoretically suggestive and not tested here.

5.3 Deconstructing Social Resilience

The most revealing result in this thesis may not be one of the four confirmatory hypotheses, but the exploratory item-level pattern in Section 4.5. Disinvestment (Item 2) behaves exactly as the theoretical model predicts: legitimacy fully mediates the effect of both accuracy and completeness on it, mirroring the composite-level result in every respect. On this behavioural outcome, the theory works cleanly.

The long-term recovery buffer (Item 4) breaks the pattern. Legitimacy has no effect on it at all, yet completeness still affects it directly, and in the opposite direction than expected: investors

exposed to fuller disclosure appear less, not more, willing to wait through a slow recovery. This unexpected direction is consistent with the backfire mechanism described by Hickman and Wong-On-Wing (2024) in Section 2.3, where investors read a firm's disclosure of its own shortcomings as evidence against it rather than as a mitigating signal. This parallel should be read with caution: Hickman and Wong-On-Wing describe investors reinterpreting a disclosure as hypocritical once a crisis exposes exactly the area a firm had claimed to manage well, while the completeness manipulation used here is not deceptive (it's simply a broader and more honest report). The two mechanisms may not be identical, and this thesis cannot confirm that they are. What they share is the broader idea that naming a weakness can cost a firm investor patience instead of earning it, and it is this shared idea, not a shared mechanism, that makes the parallel worth noticing. This result also fits the weak factor structure reported in Appendix D, it is one more sign that the four Social Resilience items do not move together as a single construct.

The distinction between the two items matters for how "social resilience" should be understood going forward. Exiting a position and tolerating a slow recovery are not the same response, even though both are framed as resilience-related in this thesis. Disinvestment depends on the reputational buffer that legitimacy provides. The long-term recovery buffer does not. Something else drives it, and legitimacy plays no part in that explanation.

5.4 Limitations and Future Research

Several limitations qualify the scope of these findings. The final sample (N = 90) fell short of the pre-specified targets for the direct-effect (N = 128) and moderation (N = 133) analyses, reducing power and limiting confidence in the null results. Participants were recruited exclusively through Prolific and based in the UK (more digitally engaged and financially literate than the average retail investor) and only indirectly capturing the institutional environment of the CSRD. The Social Resilience composite itself combines four behaviourally distinct items ($\alpha = .61$, Appendix D). Rather than a pure measurement flaw, this reflects what section 5.3 already suggest. The four items may capture genuinely different investor responses rather than interchangeable indicators of one construct. The completeness manipulation itself bundles two distinct signals (topical breadth and the disclosure of negative information) into a single factor, which means this thesis cannot say which of the two drives completeness's stronger-than-expected effect on legitimacy (Section 5.1) or its unexpected direct effect on long-term patience (Section 5.3). Finally, the design relies on a single fictional firm in a single sector, and measures stated behavioural intentions in a hypothetical scenario rather than observed behaviour during a real crisis.

Four directions follow naturally. Future studies should recruit a larger, adequately powered sample to properly test the full model, especially moderation by Moral Awareness. They should examine samples more directly embedded in the CSRD's jurisdiction. They should pre-register the item-level distinction identified in Section 5.3, to establish whether disinvestment and long-term patience really are governed by different mechanisms. A split that, if it replicates, would suggest the literature needs to speak less about social resilience in the singular and more about distinct resilience-related investor behaviours. And they should deconstruct completeness itself, manipulating topical breadth and the disclosure of negative information separately, to identify which component is actually responsible for the effects reported here.

6. Conclusion

This thesis asked whether the quality-legitimacy-resilience pathway identified by Zahller et al. (2015) still holds today, in a market shaped by mandatory disclosure, a higher materiality sector, and a different population of retail investors. On the whole, it does, with one important qualification the item-level results add below. Disclosure accuracy builds investor trust in a firm (perceived legitimacy) and that legitimacy, not the disclosure itself, is what protects the firm when a crisis hits. The strength of this pathway is nearly unchanged from a decade ago ($\beta = .31$ here versus $\beta = .335$ in the original study), despite every surrounding condition being different. That stability is the central finding of this replication.

The item-level analyses go somewhere the original study couldn't. What looked like a flat null on H1 and H2 hides two behaviours pulling in different directions. Legitimacy fully explains why better disclosure reduces the impulse to divest (exactly as the theory predicts). But it explains nothing about a second behaviour, investors' patience during a slow recovery. There, completeness has a direct effect of its own but backwards, it leaves investors less patient, not more, once a firm has admitted its own shortcomings. "Social resilience" may not be one construct at all, but a label stretched over behaviours that run on different mechanisms, a distinction that remains exploratory, but one worth testing directly in future work.

For corporate practice, the implication is not that comprehensiveness is worthless. Completeness does build legitimacy, and more strongly here than a decade ago, plausibly because admitting shortfalls is now one of the few costly, differentiating signals left under a mandatory reporting regime. But the item-level finding above suggests a limit to what comprehensiveness buys a firm in a crisis: a firm that has already named its own risks does not automatically earn more patience for it. If anything, in this study, it earned less. Precision (verifiable, auditable data) is what most reliably builds the reputational buffer that holds when a shock arrives. Under the CSRD, broad disclosure is now a legal floor that every firm clears, so it no longer sets anyone apart. Accuracy is the margin that firms still control, and with it, their legitimacy.

References

Adger, W. N. (2000). Social and ecological resilience: are they related?. *Progress in human geography*, *24*(3), 347-364. https://doi.org/10.1191/030913200701540465

Akerlof, G. A. (1978). The market for “lemons”: Quality uncertainty and the market mechanism. In Uncertainty in economics (pp. 235-251). Academic Press. https://doi.org/10.2307/1879431

Aramian, F., & Comerton-Forde, C. (2023). Retail trading in European equity markets. Available at SSRN 4428511. https://doi.org/10.2139/ssrn.4428511

Asay, H. S., Hales, J., Hinds, C., & Rupar, K. (2023). Nonprofessional investor judgments: Linking dependent measures to constructs. *The Accounting Review*, *98*(7), 1-32. https://doi.org/10.2308/TAR-2021-0551

Bachmann, P., & Ingenhoff, D. (2016). Legitimacy through CSR disclosures? The advantage outweighs the disadvantages. *Public Relations Review*, *42*(3), 386-394. https://doi.org/10.1016/j.pubrev.2016.02.008

Backof, A. G., Negangard, E. M., & Winchel, J. (2020). CSR restatements and the contagion effect: An experimental investigation of issue materiality and intent. *Available at SSRN 3388115*. http://dx.doi.org/10.2139/ssrn.3388115

Blacconiere, W. G., & Patten, D. M. (1994). Environmental disclosures, regulatory costs, and changes in firm value. Journal of accounting and economics, 18(3), 357-377. https://doi.org/10.1016/0165-4101(94)90026-4

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of management*, *37*(1), 39-67. https://doi.org/10.1177/0149206310388419

De Villiers, C., Cho, C. H., Turner, M. J., & Scarpa, R. (2023). Are shareholders willing to pay for financial, social and environmental disclosure? A choice-based experiment. *European Accounting Review*, *32*(1), 1-28. https://doi.org/10.1080/09638180.2021.1944890

Elliott, W. B., Jackson, K. E., Peecher, M. E., & White, B. J. (2014). The unintended effect of corporate social responsibility performance on investors' estimates of fundamental value. *The Accounting Review*, *89*(1), 275-302. https://doi.org/10.2308/accr-50577

Hickman, L. E., & Wong-On-Wing, B. (2024). CSR as insurance: Limitations and risk of backfiring. In K. E. Karim (Ed.), *Advances in Accounting Behavioral Research* (Vol. 27, pp. 73–93). Emerald Publishing Limited. https://doi.org/10.1108/S1475-148820240000027003

Hoang, H., & Trotman, K. T. (2021). The effect of CSR assurance and explicit assessment on investor valuation judgments. *Auditing: A Journal of Practice & Theory*, *40*(1), 19-33. https://doi.org/10.2308/AJPT-18-092

Hummel, K., & Schlick, C. (2016). The relationship between sustainability performance and sustainability disclosure–Reconciling voluntary disclosure theory and legitimacy

theory. *Journal of accounting and public policy*, *35*(5), 455-476. https://doi.org/10.1016/j.jaccpubpol.2016.06.001

Jonsson, S., Greve, H. R., & Fujiwara-Greve, T. (2009). Undeserved loss: The spread of legitimacy loss to innocent organizations in response to reported corporate deviance. *Administrative Science Quarterly*, 195-228. https://doi.org/10.2189/asqu.2009.54.2.195

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of financial economics*, *115*(2), 304-329. https://doi.org/10.1016/j.jfineco.2014.09.008

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, *72*(4), 1785-1824. https://doi.org/10.1111/jofi.12505

Lock, I., & Schulz-Knappe, C. (2019). Credible corporate social responsibility (CSR) communication predicts legitimacy: Evidence from an experimental study. *Corporate Communications: An International Journal*, *24*(1), 2-20. https://doi.org/10.1108/CCIJ-07-2018-0071

Lu, L., Li, X., & Huang, Y. (2024). Corporate social responsibility and stock resilience to COVID-19: A contract theory perspective. International Review of Economics & Finance, 89, 12–29. 10.1016/j.iref.2023.10.001

Misiuda, M., & Lachmann, M. (2022). Investors' perceptions of sustainability reporting—a review of the experimental literature. *Sustainability*, *14*(24), 16746. https://doi.org/10.3390/su142416746

Nafisa, R., Ashraful, A. M., & Qian, A. (2023). Corporate ESG issues and retail investorsâ€™ investment decision: a moral awareness perspective. *International Journal of Research in Business and Social Science (2147-4478)*, *12*(9), 113-125. https://doi.org/10.20525/ijrbs.v12i9.2990

O'Dwyer, B., Unerman, J., & Bradley, J. (2005). Perceptions on the emergence and future development of corporate social disclosure in Ireland: engaging the voices of non-governmental organisations. *Accounting, Auditing & Accountability Journal*, *18*(1), 14-43. https://doi.org/10.1108/09513570510584647

Oyetade, J. A., Igbalawole, O., & Johnson-Rokosu, S. F. (2025). From Shareholder Value to Stakeholder Impact: Integrating ESG Considerations into Financial Decisions. *Int. J. Res. Innov. Soc. Sci*, *9*, 239-255. http://dx.doi.org/10.2139/ssrn.5160191

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, *63*(1), 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in experimental social psychology* (Vol. 10, pp. 221-279). Academic Press. https://doi.org/10.1016/S0065-2601(08)60358-5

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, *177*(3), 1333-1352. 10.1016/j.ejor.2005.04.006

Spence, M. (1978). Job market signaling. In *Uncertainty in economics* (pp. 281-306). Academic Press. https://doi.org/10.1016/B978-0-12-214850-7.50025-5

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of management review*, *20*(3), 571-610. https://doi.org/10.2307/258788

Zahller, K. A., Arnold, V., & Roberts, R. W. (2015). Using CSR disclosure quality to develop social resilience to exogenous shocks: A test of investor perceptions. *Behavioral Research in Accounting*, *27*(2), 155-177. https://doi.org/10.2308/bria-51118

Appendix A: Experimental Stimulus Materials

The following four versions of the TechCore Industries Sustainability Report excerpt were used in the experiment. Each participant was randomly assigned to one condition only.

Condition 1: High Accuracy / High Completeness

Ethical Supplier Certification

We strive to ensure respect for human rights throughout our semiconductor supply chain. Our goal this year was to certify 100% of our Tier 1 suppliers according to the RBA (Responsible Business Alliance) standard. We achieved this goal with the full certification of our 47 direct suppliers.

Critical Minerals Traceability

We are implementing the necessary systems to ensure full traceability of lithium and rare earth elements. Our objective was to reach 85-90% blockchain traceability across our entire supply chain. Unfortunately, we only reached 76% traceability, primarily due to ERP integration delays affecting several Tier 2 facilities in Central Europe.

Working Conditions at Partner Sites

We are committed to regularly auditing our industrial partners. Our target was zero safety incidents across all certified sites. Unfortunately, we did not meet this objective. In February, a routine inspection at an assembly subcontractor in the United States identified inadequate ventilation standards affecting 23 production workers. We immediately suspended this supplier (representing 3% of our production), demanded corrective measures, and reinforced our audit protocols with mandatory quarterly unannounced visits.

Community Engagement

We support digital education in the regions where we operate. This year, we funded 12 programming training centers across three producing countries, training 847 young people. We met our goal of increasing community investment by 8% and exceeded our employee volunteering target of 15% with an actual participation rate of 18.3%.

Condition 2: High Accuracy / Low Completeness

Ethical Supplier Certification

We strive to ensure respect for human rights throughout our semiconductor supply chain. Our goal this year was to certify 100% of our Tier 1 suppliers according to the RBA (Responsible Business Alliance) standard. We achieved this goal with the full certification of our 47 direct suppliers.

Community Engagement

We support digital education in the regions where we operate. This year, we funded 12 programming training centers across three producing countries, training 847 young people. We met our goal of increasing community investment by 8% and exceeded our employee volunteering target of 15% with an actual participation rate of 18.3%.

Condition 3: Low Accuracy / High Completeness

Ethical Supplier Certification

We strive to ensure respect for human rights throughout our semiconductor supply chain. We are working to have our direct suppliers certified according to recognized international standards. We have made significant progress this year with the certification of nearly all our key partners.

Critical Minerals Traceability

We are implementing the necessary systems to ensure better traceability of lithium and rare earth elements. Our goal was to significantly improve traceability throughout our entire supply chain. We have made progress but remain slightly below our target, primarily due to technical integration delays affecting several indirect facilities in Europe.

Working Conditions at Partner Sites

We are committed to regularly auditing our industrial partners. Our objective was to eliminate unacceptable practices. Unfortunately, a recent inspection at an assembly subcontractor in the United States revealed inadequate working conditions affecting production workers. We took immediate action by temporarily suspending this supplier and strengthening our audit protocols with more frequent visits.

Community Engagement

We support digital education in the regions where we operate. This year, we funded several programming training centers in various producing countries, training many young people. We increased our community investments and exceeded our employee volunteering target this year.

Condition 4: Low Accuracy / Low Completeness

Ethical Supplier Certification

We strive to ensure respect for human rights throughout our semiconductor supply chain. We are working to have our direct suppliers certified according to recognized international standards. We have made significant progress this year with the certification of nearly all our key partners.

Community Engagement

We support digital education in the regions where we operate. This year, we funded several programming training centers in various producing countries, training many young people. We increased our community investments and exceeded our employee volunteering target this year.

Appendix B: Exogenous Shock: Industry News Stimulus (Step 7)

The following article was presented to all participants identically, regardless of condition, six months after the CSR report excerpt. It constitutes the exogenous shock that triggers the social resilience measurement.

Financial Times: March 15, 2026

Lithium supply chain scandal: Tech sector hit by massive sell-off

(Jakarta/London): Technology stocks faced their worst sell-off in three years today following an Amnesty International investigation exposing systematic forced labour across 47 lithium extraction sites in Southeast Asia.

Scale of the Scandal

At least 12,000 workers are affected. The implicated sites supply lithium to all major global semiconductor manufacturers, creating shockwaves throughout the tech industry.

Market Reaction

Investors are selling indiscriminately amid fears of severe EU trade sanctions, dragging down the entire industry regardless of each company's specific exposure:

- PacificComponents (implicated in the report): fell 14.5%.
- TechCore Industries (NOT implicated in the report): fell 9.2%.
- MicroSystems SA (NOT implicated in the report): fell 8.8%.

Analyst Take

“The market is reacting to the fear of sweeping sanctions,” explains Marc Lesage-Moretti at Goldman Sachs. “Even companies like TechCore, which the report explicitly clears of sourcing from the controversial mines, are being punished by the market contagion. Investors are shooting first and asking questions later.”

Appendix C: Full Survey Instrument (Measurement Items)

The following reproduces the exact item wording, response scales, and answer choices for all measures, the attention check, the manipulation checks, and the demographic items used in the oTree code.

Moral Awareness (Moderator)

Three items. Scale: 1 = Strongly disagree to 7 = Strongly agree (higher = more moral awareness). Response options for all three items: 1 – Strongly disagree, 2 – Disagree, 3 – Somewhat disagree, 4 – Neither agree nor disagree, 5 – Somewhat agree, 6 – Agree, 7 – Strongly agree.

1. "I believe that human rights violations (such as forced labor) in corporate supply chains have severe negative consequences for society."
2. "As an investor, I feel a personal responsibility to avoid supporting companies that fail to monitor human rights in their supply chains."
3. "It goes against my personal moral principles to invest in a company that ignores unethical labor practices for the sake of higher financial returns."

Pre-Shock Measures

Scale: 1 = An extremely poor investment to 7 = An excellent investment (higher = better).

4. Initial Quality: "Overall, assuming you were not concerned about balancing your stock portfolio, how would you rate TechCore Industries as an investment?"

Organisational Legitimacy (measured before the shock, consistent with Zahller et al., 2015)

Three items. Scale: 1 = lowest to 7 = highest legitimacy (item-specific anchors below).

5. Credibility: "How believable did you find TechCore's Sustainability Report?" (1 – Very doubtful … 7 – Very believable)
6. Stability: "Under normal economic conditions, and without considering a need to balance your portfolio, how comfortable would you be holding this investment and not monitoring it regularly?" (1 – Very uncomfortable … 7 – Very comfortable)
7. Integrity: "If there was a crisis and you could not monitor management's actions, how comfortable would you be that top management would do what's in the best long-term interests of the company and its stakeholders, and not simply in their short-term interest?" (1 – Very uncomfortable … 7 – Very comfortable)

Social Resilience (Dependent Variable)

Six items on heterogeneous scales (see Appendix D for the PCA/reliability treatment). Higher values indicate more resilience unless noted.

8. Post-shock quality: "Overall, assuming you were not concerned about balancing your stock portfolio, how would you rate TechCore Industries as an investment?" (1 – An extremely poor investment … 7 – An excellent investment)
9. Disinvestment intention: "Without considering diversification and balance in your overall portfolio, and if you had the opportunity to do so without transaction costs, what

percentage of your original, inherited investment in TechCore would you shift to a market-indexed mutual fund?" (0%, 10%, 20% … 100%, in deciles; 0% = keep everything = high resilience)

10. Expected partial recovery: "All else being equal, how long would you expect it to take for TechCore stock to show a partial recovery of at least 10% of the recent decrease in Net Income?" (One month or less / Six months or less / One year or less / More than one year)
11. Maximum acceptable wait, partial recovery: "If you were to choose a 'wait and see' strategy, and assuming that all industry stocks recovered at the same rate, what is the MAXIMUM time you would be willing to wait to see a partial recovery in TechCore stock of at least 10% of the fall in Net Income?" (same four options)
12. Expected full recovery: "All else being equal, how long would you expect it to take for TechCore stock to show a full recovery of the recent decrease in Net Income?" (One month or less / Six months or less / One year or less / Between 1 and 3 years / More than 3 years and up to 5 years / Five years or more)
13. Maximum acceptable wait, full recovery: "If you were to choose a 'wait and see' strategy, and assuming that all industry stocks recovered at the same rate, what is the MAXIMUM time you would be willing to wait to see a full recovery to the level of Net Income prior to the industry scandal?" (same six options)

Attention Check

"This study includes a quality check. Researchers need to verify that participants read each question carefully. To confirm you are reading this, please select 'Neither agree nor disagree' (4)." (Same 7-point agreement scale as the Moral Awareness items.)

Manipulation Checks

14. Scenario: "A scandal involving forced labor in Southeast Asian lithium mines caused a severe market downturn for the entire industry:" — "TechCore was one of the companies directly involved in the forced labor" vs. "TechCore was NOT one of the companies directly involved in the forced labor"
15. Completeness: "TechCore's Sustainability Report disclosed:" — "Only areas of good performance" vs. "Areas where the company had both good and bad performance"
16. Accuracy: "TechCore's Sustainability Report used:" — "Specific language with numbers you can compare to prior years and other companies (e.g. exact percentages or specific financial amounts)" vs. "Vague language and non-verifiable or non-measurable amounts you cannot compare to prior years and other companies (e.g., 'significantly increased' or 'committed to')"

Demographics (survey items shown in the code excerpt)

Age, sex, ethnicity, and employment status (reported in Table 3) were also collected but are not shown in the code excerpt supplied for this appendix; add them here if you want the full set documented.

17. Education level: GCSE or equivalent / A-Levels or equivalent / Bachelor's degree / Master's degree / Doctorate or higher / Prefer not to say
18. Investment experience: Very little / Moderate / Considerable / Expert

Appendix D: Construct Validation: Full PCA and Reliability Statistics

Sections 4.1 and 4.5 refer to the reliability of the three multi item composites used in this thesis. This appendix explains how each was checked before any hypothesis was tested, using two standard tools.

A Principal Components Analysis (PCA) checks whether the items in a scale are really measuring one single underlying idea, or several different things bundled together. Items that all "load" strongly into one factor are evidence of this single underlying construct. Cronbach's alpha measures something related but distinct: how consistently participants answered across the items, a high alpha means participants who scored high on one item tended to score high on the others too.

Organisational Legitimacy performed well on both counts. All three items loaded strongly on a single factor (loadings: .86, .87, .87), explaining 75% of the variance. Internal consistency was excellent ($\alpha = .84$, 95% CI [.77, .89]).

Moral Awareness showed a similar pattern: loadings of .77, .90, and .93 confirmed a clean single factor structure (75% of variance explained), with the same internal consistency ($\alpha = .84$, 95% CI [.77, .89]).

Social Resilience was more complex. Given the heterogeneous scales of its four items (a 1–7 quality rating, a 0–100% divestment slider, and two temporal recovery buffers), all items were standardised before the PCA. All four loaded positively on a single factor (loadings: .76, .63, .66, .66), but the single-factor solution explained only 46.0% of the variance, compared to 64.01% in Zahller et al. (2015). And internal consistency fell below the conventional threshold ($\alpha = .61$, 95% CI [.46, .74]).

So common structure exists for Social Resilience, but it is comparatively weak. This is not merely a statistical artefact. Section 4.5 shows it at the mechanism level directly. Legitimacy fully mediates the effect of disclosure quality on disinvestment, but has no effect at all on the long-term recovery buffer, which follows a distinct, direct path instead. The weaker structure found here reflects a genuine multiplicity of investor responses bundled into one composite, not measurement noise alone.

This lower reliability is not simply a sign of poor scale construction. Each of the four items used maps into a category of investor judgment already established separately in the literature, rather than being randomly combined. Asay, Hales, Hinds, and Rupar (2023) identify three reliable factors in the investor-judgment literature (expectations of future firm performance, holistic firm perceptions, and investment risk evaluations). The four Social Resilience items map conceptually into the second and third of these (investment quality and divestment intent align with "holistic perceptions", the recovery estimates align with "investment risk"). Among the 71 participants who required no stimulus reload on either CSR manipulation check, α rose to .71, suggesting that experimental encoding noise also contributes to the lower overall figure, alongside the heterogeneity described above.

The composite score was nonetheless retained for the main analyses reported in Section 4.5, since these items reflect judgment dimensions well established in the literature. Their individual patterns are examined separately in Section 4.5.

Appendix E: Robustness Checks for H3

Section 4.4.3 rests on two modelling choices: testing accuracy and completeness in two separate models and controlling for the pre-shock investment rating. This appendix reports two checks verifying that neither choice drives the results reported there.

Separate models vs. one joint model. Accuracy and completeness were tested in two separate PROCESS models in the main analysis. To check that this did not distort the estimates, both were instead entered together into a single joint model (lavaan, MLR estimator). The coefficients barely moved (Accuracy: $b = 0.67$, $p < .001$; Completeness: $b = 0.45$, $p = .013$), testing them separately or together gives essentially the same answer.

With vs. without the covariate. The pre-shock investment rating was measured after participants had already read the CSR report (Section 3.7), so it could in principle be absorbing part of the effect it is meant to control for. Re-estimating the mediation model without it, the indirect effects held up (Accuracy: indirect effect = 0.20, 95% CI [0.07, 0.39]; Completeness: indirect effect = 0.13, 95% CI [0.01, 0.29]), both intervals excluding zero. Legitimacy still mediates the relationship whether or not the covariate is included.

Neither check involves an interaction term, so neither is affected by the effect-coding correction applied to the H1/H2 models (Sections 4.5.1/4.5.2).

Appendix F: Item level mediation Models (Section 4.5)

Section 4.5 tests whether legitimacy mediates the effect of disclosure quality on two individual items (disinvestment and the long-term recovery buffer) the same way it mediates the composite Social Resilience score in Section 4.5.3. As in Section 4.5.3, accuracy and completeness are each tested in a separate PROCESS Model 4 mediation, so each item has two legitimacy → outcome coefficients (b_1), one per model. Table F1 reports both.

Table F1: Legitimacy → Outcome Path (b_1), by Model

Outcome	**Predictor (X)**	**b_1**	**p**
Disinvestment	Accuracy	−10.52	.003
Disinvestment	Completeness	−10.99	.002
Full recovery buffer	Accuracy	−0.01	.900
Full recovery buffer	Completeness	0.07	.442

Note. Each row comes from a separate PROCESS Model 4 mediation, with the pre-shock investment rating included as a covariate.

For disinvestment, the two b_1 values (−10.52 and −10.99) are close and both significant: in either model, a more legitimate firm predicts less disinvestment. For the full recovery buffer, the two b_1 values are non-significant and do not agree in sign (−0.01 versus 0.07), consistent with Section 4.5's conclusion that legitimacy shows no detectable effect on this item in either model.

.

Appendix G: Full R Code

R script used to reproduce every statistic reported in Chapters 3 to 5. Packages are loaded at the top, the script follows the order of the thesis by sections:

Packages and data

```
library(dplyr)
library(car)
library(psych)
library(lavaan)
library(pwr)
source("PROCESS v5.0 for R/process.R")  # Hayes (2022),
    Model 4

data_source <- readRDS("data_propre.rds")
nrow(data_source)  # 90
```

Raw variable selection

```
data_brut <- data_source %>%
  select(
    participant.code,
    player.accuracy, player.completeness,
    player.excluded, player.exclusion_reason,
    player.strikes_scenario, player.strikes_
       completeness, player.strikes_accuracy,
    player.moral_awareness_1, player.moral_awareness_2,
        player.moral_awareness_3,
    player.initial_quality,
    player.legitimacy_credibility, player.legitimacy_
       stability, player.legitimacy_integrity,
    player.resilience_quality, player.resilience_
       disinvestment,
    player.recovery_partial_expected, player.recovery_
       partial_max,
    player.recovery_full_expected, player.recovery_full
       _max,
    player.attention_check,
    player.check_scenario, player.check_completeness,
       player.check_accuracy,
    player.education, player.investment_experience
  )

nrow(data_brut); ncol(data_brut)  # 90, 26
colSums(is.na(data_brut))         # should be 0
table(data_brut$player.excluded) # should be 0 (no
   excluded participant left)
```

Variable construction

```
data <- data_brut %>%
  mutate(
    # -0.5/+0.5 instead of 0/1, so each main effect
       reflects the average effect of that factor
       across both levels of the other one, not just
       its effect when the other factor happens to be
       at zero
    accuracy_num = ifelse(player.accuracy == "high", 1,
        0),

    completeness_num = ifelse(player.completeness == "
       high", 1, 0),

    accuracy_eff = ifelse(accuracy_num == 1, 0.5, -0.5)

    completeness_eff = ifelse(completeness_num == 1,
       0.5, -0.5),

    # (6) Covariate
    initial_quality = player.initial_quality,

    # (6) Organisational Legitimacy: mean of 3 items
    score_legitimacy = (player.legitimacy_credibility +
                        player.legitimacy_stability +
                        player.legitimacy_integrity)
                        / 3,

    # (2) Moral Awareness: mean of 3 items, centred "c"
    score_moral_awareness =
                        (player.moral_awareness_1 +
                         player.moral_awareness_2 +
                         player.moral_awareness_3)
                         / 3,

    moral_awareness_c = score_moral_awareness -mean(
       score_moral_awareness),

    # Social Resilience: components
    resilience_retained = 100 - player.resilience_
```

```
      disinvestment,  # reversed

    partial_buffer = player.recovery_partial_max -
      player.recovery_partial_expected,

    full_buffer = player.recovery_full_max - player.
      recovery_full_expected),

    # Standardised before averaging: components are on
      different scales (7pt, %, time categories)
    resilience_quality_z = as.numeric(scale(player.
      resilience_quality)),

    resilience_retained_z = as.numeric(scale(resilience
      _retained)),

    partial_buffer_z = as.numeric(scale(partial_buffer)
      ),

    full_buffer_z = as.numeric(scale(full_buffer)),

    # (9) Social Resilience : mean of 4 items
    score_resilience = (resilience_quality_z +
                        resilience_retained_z +
                        partial_buffer_z +
                        full_buffer_z)/ 4)
```

Sample description (Section 4.1)

```
table(data$player.education)
round(prop.table(table(data$player.education)) * 100,
  1)

table(data$player.investment_experience)
round(prop.table(table(data$player.investment_
  experience)) * 100, 1)

mean(data$score_moral_awareness)
sd(data$score_moral_awareness)
median(data$score_moral_awareness)
```

Manipulation checks (Section 4.2)

```r
data <- data %>%
  mutate(
    failed_accuracy_check = ifelse(player.strikes_
       accuracy > 0, 1, 0),

    failed_completeness_check = ifelse(player.strikes_
       completeness > 0, 1, 0),

    failed_scenario_check = ifelse(player.strikes_
       scenario > 0, 1, 0))

# first-attempt success rate, accuracy check
1 - mean(data$failed_accuracy_check)

# first-attempt success rate, completeness check
1 - mean(data$failed_completeness_check)

# first-attempt success rate, scenario check
1 - mean(data$failed_scenario_check)

# attention check: everyone should answer 4
table(data$player.attention_check == 4)

# balance of the 2x2 randomisation
table(data$player.accuracy, data$player.completeness)
chisq.test(table(data$player.accuracy, data$player.
   completeness))
```

Descriptive statistics and correlations (Section 4.4)

```r
data %>%
  group_by(player.accuracy, player.completeness) %>%
  summarise(
    n = n(),

    legit_mean = round(mean(score_legitimacy), 2),

    legit_sd = round(sd(score_legitimacy), 2),

    resil_mean = round(mean(score_resilience), 2),

    resil_sd = round(sd(score_resilience), 2),
```

```
    iq_mean = round(mean(initial_quality), 2),

    iq_sd = round(sd(initial_quality), 2),

    .groups = "drop")

corr.test(data[, c("initial_quality", "score_legitimacy
   ", "score_resilience", "score_moral_awareness")])
```

H1 and H2 – ANCOVA on Social Resilience (Section 4.5.1-4.5.2)

```
m_h1h2 <- lm(score_resilience ~ accuracy_eff *
   completeness_eff + initial_quality, data = data)

Anova(m_h1h2, type = 3)

summary(m_h1h2)$coefficients

effectsize::eta_squared(m_h1h2, partial = TRUE)

# same model without the covariate (robustness, see
   Section 3.6)
m_h1h2_nocov <- lm(score_resilience ~ accuracy_eff *
   completeness_eff, data = data)
Anova(m_h1h2_nocov, type = 3)

# "homogeneity of slopes" check: covariate x condition
   interactions shouldn't be significant
m_slopes <- lm(score_resilience ~ accuracy_eff *
   completeness_eff * initial_quality, data = data)
Anova(m_slopes, type = 3)

# diagnostics
vif(m_h1h2)
shapiro.test(residuals(m_h1h2))
ncvTest(m_h1h2)
```

H3 – Mediation through legitimacy (Section 4.5.3)

```
process(data = as.data.frame(data), y = "score_
   resilience", x = "accuracy_eff",
```

```
        m = "score_legitimacy", cov = "initial_quality"
            , model = 4, boot = 5000, seed = 12345)

process(data = as.data.frame(data), y = "score_
   resilience", x = "completeness_eff",
        m = "score_legitimacy", cov = "initial_quality"
            , model = 4, boot = 5000, seed = 12345)

# standardised a1 paths: turns the raw PROCESS
   coefficients (a1 = 0.70, 0.50) into betas, to
   compare to ZAR (2015)
data_z <- data %>%
  mutate(
    accuracy_eff_z = as.numeric(scale(accuracy_eff)),

    completeness_eff_z = as.numeric(scale(completeness_
       eff)),

    score_legitimacy_z = as.numeric(scale(score_
       legitimacy)),

    initial_quality_z = as.numeric(scale(initial_
       quality))
  )

summary(lm(score_legitimacy_z ~ accuracy_eff_z +
   initial_quality_z, data = data_z))$coefficients

summary(lm(score_legitimacy_z ~ completeness_eff_z +
   initial_quality_z, data = data_z))$coefficients
```

H4 – Moderation by moral awareness (Section 4.5.4)

```
m_h4a <- lm(score_legitimacy ~ accuracy_eff * moral_
   awareness_c + initial_quality, data = data)
summary(m_h4a)$coefficients
vif(m_h4a)

m_h4b <- lm(score_legitimacy ~ completeness_eff * moral
   _awareness_c + initial_quality, data = data)
summary(m_h4b)$coefficients
vif(m_h4b)
```

Exploratory item-level analyses (Section 4.6)

```
lm_quality <- lm(player.resilience_quality ~ accuracy_
   eff * completeness_eff + initial_quality, data =
   data)

lm_divest  <- lm(player.resilience_disinvestment ~
   accuracy_eff * completeness_eff + initial_quality,
   data = data)

lm_partial <- lm(partial_buffer ~ accuracy_eff *
   completeness_eff + initial_quality, data = data)

lm_full    <- lm(full_buffer ~ accuracy_eff *
   completeness_eff + initial_quality, data = data)

for (m in list(lm_quality, lm_divest, lm_partial, lm_
   full)) {print(Anova(m, type = 3))}

# direction of the two effects that reach/approach
   significance (full_buffer and disinvestment)
data %>% group_by(player.accuracy) %>%
  summarise(disinvest_mean = round(mean(player.
     resilience_disinvestment), 2))

data %>% group_by(player.completeness) %>%
  summarise(full_buffer_mean = round(mean(full_buffer),
      2))

# does legitimacy mediate these two item-level effects,
    or are they direct?
process(data = as.data.frame(data), y = "player.
   resilience_disinvestment",
        x = "accuracy_eff", m = "score_legitimacy", cov
            = "initial_quality",
        model = 4, boot = 5000, seed = 12345)

process(data = as.data.frame(data), y = "player.
   resilience_disinvestment",
        x = "completeness_eff", m = "score_legitimacy",
            cov = "initial_quality",
        model = 4, boot = 5000, seed = 12345)
```

```
process(data = as.data.frame(data), y = "full_buffer",
        x = "accuracy_eff", m = "score_legitimacy", cov
            = "initial_quality",
        model = 4, boot = 5000, seed = 12345)

process(data = as.data.frame(data), y = "full_buffer",
        x = "completeness_eff", m = "score_legitimacy",
            cov = "initial_quality",
        model = 4, boot = 5000, seed = 12345)
```

Construct validation – PCA and reliability (Appendix D)

```
# Organisational Legitimacy
items_legit <- data[, c("player.legitimacy_credibility"
                        "player.legitimacy_stability",
                        "player.legitimacy_integrity")]
alpha(items_legit)

principal(items_legit, nfactors = 1, rotate = "none")

# Moral Awareness
items_ma <- data[, c("player.moral_awareness_1",
                     "player.moral_awareness_2",
                     "player.moral_awareness_3")]
alpha(items_ma)

principal(items_ma, nfactors = 1, rotate = "none")

# Social Resilience (weaker structure, discussed in 4.1
   /5.3/5.4)
items_resilience <- data[, c("resilience_quality_z",
                             "resilience_retained_z",
                             "partial_buffer_z",
                             "full_buffer_z")]
alpha(items_resilience)

principal(items_resilience, nfactors = 1, rotate = "
   none")

# clean subsample: excludes participants who failed the
    accuracy or completeness manipulation check on the
   first attempt (N = 71). alpha rises to .71 on this
   subsample.
```

```
# Note: the scenario check is not part of this filter,
   so this is a different group from the N = 67 "passed
    all three checks" figure in Section 4.2.
data <- data %>%
  mutate(
  clean_subsample = ifelse( player.strikes_completeness
      == 0 & player.strikes_accuracy == 0, 1, 0))

table(data$clean_subsample)  # 71 vs 19

data_clean <- data %>% filter(clean_subsample == 1)

items_resilience_clean <- data_clean[, c("resilience_
   quality_z", "resilience_retained_z", "partial_buffer
   _z", "full_buffer_z")]

alpha(items_resilience_clean)  # 0.71
```

Robustness checks for H3 (Appendix E)

```
# accuracy and completeness estimated together instead
   of in two separate PROCESS models, to check neither
   estimate is distorted by testing them apart
model_joint <- '
  score_legitimacy ~ a1*accuracy_eff + a2*completeness_
     eff + a3*initial_quality
  score_resilience ~ b1*score_legitimacy + c1*accuracy_
     eff + c2*completeness_eff + b2*initial_quality
  indirect_acc := a1*b1
  indirect_comp := a2*b1
'
fit_joint <- sem(model_joint, data = data, estimator =
   "MLR")

summary(fit_joint, standardized = TRUE)

# same mediation model without the pre-shock rating as
   covariate
process(data = as.data.frame(data), y = "score_
   resilience", x = "accuracy_eff",
        m = "score_legitimacy", model = 4, boot = 5000,
            seed = 12345)
```

```
process(data = as.data.frame(data), y = "score_
   resilience", x = "completeness_eff",
        m = "score_legitimacy", model = 4, boot = 5000,
            seed = 12345)
```

Funding:

This work was supported by the Fonds de la Recherche Scientifique - FNRS under Grant n° 40013829.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Chaussée de Binche 151, 7000 Mons, Belgique | **www.uclouvain.be/lsm**