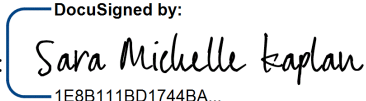


Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature: 1E8B111BD1744BA...

Sara Michelle Kaplan

Name

5/21/2024 | 10:20 PM EDT

Date

Title Sustainability certifiers: Diversification, perceived trustworthiness, and performance

Author Sara Michelle Kaplan

Degree Doctor of Philosophy

Program Business

Approved by the Committee

DocuSigned by:
Giacomo Negro
2D17167AD7F743F...

Giacomo Negro
Advisor

DocuSigned by:
Andrea Dittmann
14CFE4709CFA406...

Andrea Dittmann
Committee Member

DocuSigned by:
Ozgecan Kocak
36FE43F565AF4AE...

Ozgecan Kocak
Committee Member

DocuSigned by:
Weihua An
128E78B2DB5E473...

Weihua An
Committee Member

Committee Member

Committee Member

Accepted by the Laney Graduate School:

Kimberly Jacob Arriola, Ph.D, MPH
Dean, James T. Laney Graduate School

Date

Sustainability certifiers:

Diversification, perceived trustworthiness, and performance

Sara M. Kaplan, Ph.D. Candidate

May 2024

A dissertation submitted to the Faculty of the Laney Graduate School of Emory University in
partial fulfillment of the requirement for the degree of Doctor of Philosophy in Organization &

Management at Goizueta Business School

Abstract

Sustainability certifiers: Diversification, perceived trustworthiness, and performance

This dissertation consists of two papers related to the perceived trustworthiness of certifiers.

Certifiers are information intermediaries that define quality standards and ensure offerings within their domain meet them. In doing so, they play a pivotal role in shaping markets and industries.

Sustainability certifiers, in particular, have an ever-increasing impact due to growing awareness of the social and environmental risks organizations face. The first paper details the development and validation of a new measure for the perceived trustworthiness of certifiers. Certifier

trustworthiness is developed as a construct based on seminal research on interpersonal trustworthiness. The final certifier trustworthiness scale (CTS) consists of three dimensions: perceived ability, integrity, and authenticity. The validation of the CTS provides an invaluable tool for researchers to delve into the intricate ways audiences evaluate and establish trust in these unique organizations. The second paper develops and tests the theory that diversified sustainability certifiers are perceived as less trustworthy across all three certifier trustworthiness dimensions compared to their specialized counterparts, contributing to a diversification discount for certifiers. This paper uses a multi-method research approach pairing experimental and archival evidence to test the hypotheses. The findings support that specialist sustainability certifiers outperform unrelatedly diversified sustainability certifiers and that perceived trustworthiness mediates the relationship between certifier specialization and financial performance. Further, the experimental evidence supports the hypotheses that product category diversification negatively affects the three certifier trustworthiness dimensions - perceived ability, integrity, and authenticity—leading to decreased appeal and, ultimately, reduced performance.

Keywords: sustainability certifiers, trustworthiness, authenticity, diversification

Sustainability certifiers:

Diversification, perceived trustworthiness, and performance

Sara M. Kaplan, Ph.D. Candidate

May 2024

A dissertation submitted to the Faculty of the Laney Graduate School of Emory University in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Organization & Management at Goizueta Business School

Acknowledgements

I am grateful for the support from my advisor, Giacomo Negro, and for the guidance of Özgecan Koçak, Andrea Dittmann, and Weihua An. Any errors are my own.

Table of Contents

ABSTRACT	1
CERTIFYING ORGANIZATIONS: DIVERSIFICATION, PERCEIVED TRUSTWORTHINESS, AND PERFORMANCE ... ERROR!	
BOOKMARK NOT DEFINED.	
PAPER 1: THE CERTIFIER TRUSTWORTHINESS SCALE	9
THEORETICAL BACKGROUND	9
VALIDATION STRATEGY	13
<i>Conceptualization</i>	13
<i>Scale Development</i>	19
<i>Content Validity</i>	20
SCALE EVALUATION AND REFINEMENT	24
<i>Pretest</i>	24
<i>Final Test</i>	30
DISCUSSION	38
LIMITATIONS AND FUTURE DIRECTIONS	40
PAPER 1 APPENDICES	41
TABLE 12: CTS DESCRIPTIVE STATISTICS FOR NORM DEVELOPMENT	52
PAPER 2: SUSTAINABILITY CERTIFIER DIVERSIFICATION: EFFECTS ON PERCEIVED TRUSTWORTHINESS AND PERFORMANCE	63
THEORETICAL BACKGROUND	63
<i>Certifier Diversification and Performance</i>	63
<i>Certifier Diversification and Perceived Trustworthiness</i>	66
<i>Perceived Trustworthiness as a Mediator</i>	69
CONTEXT OF SUSTAINABILITY CERTIFIERS	70
DATA AND METHODS	71
<i>Study 1 – Experiment, Mechanism of Perceived Trustworthiness</i>	71
<i>Study 2a – Survey, Identifying Unrelated Diversification</i>	78
<i>Study 2b – Observational, Ecolabels</i>	82
DISCUSSION	94
LIMITATIONS AND FUTURE DIRECTIONS	96
CONCLUSION	98
PAPER 2 APPENDICES	99

Sustainability certifiers:

Diversification, perceived trustworthiness, and performance

Amidst growing public environmental and social consciousness, the pressure for producers to certify their offerings has increased (Rothenberg et al., 2019). Certifying organizations (henceforth, “certifiers”) vary significantly in the products and services they cover, the quality issues they represent, and their organizational structure. According to the market research firm Mordor Intelligence (2023), the global market for testing, inspection, and certification was worth over \$236 billion in 2023 (about 0.2% of total GDP); it will likely reach a value of \$328 billion by 2028. With an increase in the size of the certification market comes added complexity in terms of the number of involved actors and increased options for producer organizations. The emergence of accreditation bodies such as the ISEAL Alliance, founded in 2016 to improve the credibility and impact of “sustainability systems,” highlights demand from the organizations they certify for reliable insight into the trustworthiness of sustainability certifiers (ISEAL, 2023).

Third-party certifiers play a crucial role in reducing information asymmetry between producers and consumers. They use their reputations to support attributes of goods or services that are not easily evaluated by users even after purchase (Barnett, 2019). They are information intermediaries, focusing on “the gathering and provision of data relevant to consumer choice” (Sharkey et al., 2023). In practice, the term “certifier” labels various types of organizations. In the present research, the term refers to organizations that currently manage the standards and verification process for at least one certification offering and own trademark rights to the certification label (if trademarked). In some cases, certifiers oversee the verification process but outsource parts of the process (e.g., audits) to another organization (Gorton et al., 2021).

The primary consumers of certifications are individuals or groups of people; they use certifications to signal that they, their group, or their products or services meet desired standards. For the present research, these primary consumers are referred to as “producers.” Producers often use the certification they earn to promote themselves, their product, or their offerings. Some research indicates that end consumers are willing to pay higher prices for certified products (e.g., Prell et al., 2020). Thus, the secondary consumers of certifications are those who purchase or use the services or products offered by the producers. For example, Vegan Action offers a certification for products made without animal products or byproducts (Get certified, 2024). The primary users (producers) for the Vegan Action certification are mainly food and beverage suppliers and retailers such as Ben & Jerry’s (Non-dairy pints, 2024). The secondary users are individuals who purchase food and beverage products for consumption. Thus, the certifier intermediates the relationship between the producer and the end consumer, influencing consumer behavior and producer practices.

Trustworthiness is particularly relevant to certifiers. Certifiers assure audiences that an organization or its products and services meet a certain level of quality regarding select criteria. Audiences are generally aware that producers are self-interested and may not be transparent about the weaknesses of their offerings. Certifiers and other gatekeepers help bridge the knowledge gap between producers and external audiences such as investors and consumers. Often, no other mechanisms are in place to ensure producers produce what they claim and how they claim. Governmental agencies do not have the resources nor authority to oversee production within every industry. For example, the dietary supplement industry is largely unregulated except for post-market recalls by the Food and Drug Administration. Thus, audiences, particularly consumers, often rely upon certifiers to ensure products and services meet set standards.

The first paper of this two-paper dissertation focuses on the development and validation of the Certifier Trustworthiness Scale (CTS), aimed at measuring the perceived trustworthiness of certifiers. Drawing on seminal research on interpersonal trustworthiness, I conceptualize perceived certifier trustworthiness as a tri-dimensional construct comprising perceived ability, authenticity, and integrity. In the second paper, I delve into select drivers and outcomes of perceived certifier trustworthiness. Specifically, I theorize and provide empirical support for the notion that unrelated diversification by certifiers across business categories negatively impacts perceived trustworthiness across all three dimensions, leading to a diversification discount for certifiers over time.

Paper 1: The Certifier Trustworthiness Scale

The idea that public audiences evaluate the trustworthiness of organizations is well-established. However, little research within the discipline of organizational theory has explicitly tested how external audiences perceive the trustworthiness of organizations (see for exceptions Pirson et al., 2019; Pirson & Malhorta, 2011). Exploring trustworthiness promises to illuminate pathways between organizational strategies and success.

Trustworthiness is especially relevant to certifiers, as they transfer claims of their credibility to the offerings or entities they certify. However, we know surprisingly little about the factors driving certifiers' perceived trustworthiness. The research in this first paper integrates psychological and sociological models of trustworthiness to create a complete and concise measure of individuals' judgments of certifier trustworthiness. The validation of a psychometrically valid scale of certifier trustworthiness will allow researchers to explore how audiences determine which organizations to trust.

Theoretical Background

Certification, a crucial process in which an authoritative entity formally acknowledges that a venture meets established standards, is deeply intertwined with the concept of trust (Sine, David, & Mitsuhashi, 2007). Commonly certified ventures include individuals, organizations, and products. Certification serves as a signal that mostly unobservable processes or characteristics possess desired qualities. Third-party certifiers are often responsible for establishing the quality standards used to assess certification candidates and verifying that the candidates meet the standards.

Trustworthiness is a perceiver's (the truster's) judgment about whether an entity (the trustee) can be trusted generally or within a particular domain. A popular definition of trust comes from Mayer, Davis, and Schoorman (1995), who state that trust is:

(T)he willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. (p. 712)

In Mayer and colleagues' (1995) model, perceived trustworthiness is a composite of three dimensions: ability, integrity, and benevolence. These three dimensions positively predict trust, and the truster's propensity to trust moderates these relationships. Thus, trust results from the characteristics of the perceiver and the perceived.

Despite the extensive use of a dimensional trustworthiness construct within organizational behavior and related disciplines, the dimensionality of trustworthiness at the organizational level has rarely been examined (see for exception Peifer & Newman, 2020; Pirson & Malhorta, 2011; Pirson et al., 2019). While the trustee is a higher-order construct in the organizational context than in an interpersonal context, the truster often remains an individual. Thus, many of the same psychological mechanisms will likely apply (Rousseau et al., 1998). This is especially reasonable since a wide range of research demonstrates that audiences are prone to anthropomorphize organizations (e.g., Ashforth, Schinoff, & Brickson, 2020; Coyle-Shapiro & Kessler, 2000; Martin, 2021).

Another reason the CTS is needed is that aspects of certifiers relevant to trustworthiness judgments are not all found in other organizational forms. "Organizational forms represent a specific kind of collective organizational identity" (Hsu & Hannan, 2005, citing Hannan, 2005; Pólos et al., 2002). Audiences understand organizations of a single form to share similar core

features and be distinctive from other forms of organizations (Fiol & Romanelli, 2011; Rao et al., 2000). A core activity of certifiers is assessing whether a venture meets the standards to which it ascribes. Another core activity of certifiers is publicizing the certification achievement of the venture. However, this publicization can vary greatly in scope. For example, one certifier may simply permit the venture to label itself as certified (using branded materials or otherwise), whereas another certifier may promote the venture on its marketing materials.

Another common feature of certifiers is that they set their standards. Exceptions occur when there is broader institutional intervention, such as from government bodies or religious sects. For example, the United States Department of Agriculture sets mandatory standards for organic certification, to which accredited certifiers must comply. However, each certifier has some leeway in interpreting and enacting the standards (The Cornucopia, 2019). Certifiers vary in legal and financial structure, extra-certification activities, size, accreditation status, and the number of certifications they offer.

Membership in an organizational form can meaningfully affect an organization's performance. For example, Zuckerman (1999) found that firms that were difficult to classify based on their activities suffered devaluation. Similarly to possessing any identity, belonging to a certain organizational form comes with expectations from audiences (Hsu & Hannan, 2005). Failure to comply with audiences' expectations due to organizational form membership can also lead to negative evaluations. Therefore, it is critical to have concept and empirical clarity surrounding unique organizational forms. Focusing on a single organizational form when conceptualizing constructs such as perceived organizational trustworthiness also allows for greater construct clarity and structural validity (Yaniv, 2011).

While trustworthiness constructs and measures exist for other organization types, the unique qualities of certifiers still need to be considered. Pirson and Malhorta (2011) developed a model of *public trust in business* based on the seminal model of Mayer and colleagues (1995). They theorized that business trustworthiness consists of technical competence, managerial competence, benevolence, integrity, transparency, and identification (value congruence). Additionally, due to their focus on audience attributes, they predicted that these dimensions would have varying importance depending on whether the audiences have shallow or deep relationships with the firm they are evaluating. In a vignette-style experiment, Pirson and colleagues (2019) found that benevolence, integrity, transparency, and ability all positively predicted trust in firms. While these studies provide valuable insight into the dimensionality of organizational trustworthiness, they have several limitations for the study of certifier trustworthiness.

First, as discussed above, certifiers are a unique organizational form, similar yet distinct from other information intermediaries such as rating agencies. At a minimum, certifiers are responsible for issuing certifications to individuals or other entities based on predetermined standards. Often, certifiers are also responsible for defining and managing the standards for which they assess adherence. Therefore, items should be specific to the functions of certifiers. Additionally, unlike many other information intermediaries, the nature of the relationship between certifiers and producers allows conflicts of interest to arise. Certifiers often receive a substantial portion of their revenue from the producers they certify (see Paper 2). Therefore, certifiers are motivated to provide a positive experience to producers to maintain their support. This feature of the certifier-producer relationship makes the certifiers' consistency in applying standards across producers and following applicable laws, regulations, and industry best

practices particularly critical. Finally, the scale developed by Pirson and colleagues (2019) did not appear to undergo a rigorous development and validation process, including psychometric analysis (Morgado et al., 2018). These limitations will be addressed below in the discussion of the current validation process for the CTS.

Validation Strategy

I predominantly followed the scale development and validation process described by MacKenzie and colleagues (2011). The validation process follows five phases. The first phase is the construct conceptualization phase, which includes clearly defining the construct and its dimensions. The second phase is the development of measures, including generating items to represent the construct and assessing the content validity of the items. An important consideration in this process is recognizing that certifier trustworthiness operates as a formative construct, where its indicators cause the construct. For instance, I posit that an organization's expertise serves as an indicator of certifier trustworthiness through the ability dimension. The third phase includes pretesting the scale's psychometric properties and evaluating its convergent, discriminant, and nomological validity; the third phase includes initial scale refinement. The fourth phase is validation, gathering data from a new sample to reexamine the scale's properties and cross-validate the scale. The final phase is norm development, establishing the typical statistical properties of the scale for a population of interest.

Conceptualization

The first phase of scale development is to define the construct's conceptual domain (MacKenzie et al., 2011). This phase includes several objectives: defining the construct in unambiguous terms, explaining differences between the present definition and prior definitions, identifying the structural properties of the construct (such as whether it is a behavior or

perception), specifying the applicable entity or entities, and explaining how the construct differs from other constructs. I will address each of these items below for certifier trustworthiness.

First, certifier trustworthiness is a perception, a "socio-cognitive evaluation or judgment of an entity's appropriateness that occurs between the individual and the collective" (Alexiou & Wiggins, 2019, p. 470). The construct considers information receivers' judgments about another entity, specifically certifiers. As the label makes clear, an organization is the focal entity of interest for certifier trustworthiness. The audience perceiving the organization can consist of a single stakeholder or a mix of stakeholders from one or more groups, including employees, investors, consumers, donors, government regulators, and other intermediaries and interested parties.

Following prior definitions of trustworthiness with the certifier context in mind, I define certifier trustworthiness as *the truster's degree of confidence that the organization can be counted upon to pursue and achieve its stated mission and goals*. I theorize that the CTS has three dimensions, mirroring Mayer and colleagues' model (1995): ability, integrity, and moral authenticity (benevolence).

The ability dimension of interpersonal trustworthiness describes the characteristics of a trustee, such as their skills and competencies that enable them to carry out their commitment to the truster. The ability dimension is critical to overall interpersonal trustworthiness because having the best intentions does not matter if the trustee cannot deliver upon them. Ability is also a critical feature of a trusted certifier. Intentions are meaningless without the resources and skills necessary to deliver upon them. For the CTS, *perceived ability* is the truster's belief that the certifier has the resources and expertise needed to achieve the organization's self-proclaimed mission and goals.

In Mayer and colleagues' (1995) conceptualization of interpersonal trustworthiness, the integrity dimension captures whether the trustee holds and consistently follows principles accepted by the truster. A trustee with integrity can be relied upon; their actions are predictable. A trustworthy certifier must also be predictable and not veer from the agreed-upon terms of exchange. Additionally, certifiers must be perceived as fair and treating producers consistently, without bias. Thus, *perceived integrity* is the truster's belief that the certifier strives to make choices consistent with its mission and goals and the principles accepted by its stakeholders.

In the CTS, the third dimension, perceived authenticity, replaces the benevolence dimension of interpersonal trustworthiness. *Perceived authenticity* is the truster's belief that the certifier sincerely values achieving its mission. Mayer and colleagues (1995) define benevolence as “the extent to which a trustee is believed to want to do good to the truster” (p. 718). Benevolence typically arises due to a pre-existing relationship between the truster and the trustee, which results in the trustee wanting to support the truster. In this way, their incentives are aligned. Exhibiting benevolence is similar to exhibiting altruism, helping others without expecting self-benefit (Ferrin et al., 2006; Frost et al., 1978; Mayer et al., 1998). Acting with benevolence involves acting out of interest for another, the truster. Similarly, acting with authenticity involves acting out of interest for an ideal. For certifiers, these ideals should align with producers' ideals; therefore, a certifier acting authentically is in an associated producer's best interest.

Other conceptualizations of trustworthiness follow this tradition across levels of analysis and for specific relationship types. At the organizational level, in addition to Pirson and Malhorta (2011), Peifer and Newman (2020) adapt Mayer and colleagues' model of trustworthiness to test

how corporate social responsibility initiatives affect workers' perceptions of firm trustworthiness; they maintain the three original dimensions of ability, integrity, and benevolence.

Within management, another common conceptualization of trust and trustworthiness is founded in the sociological literature. Cook and colleagues, for example, take an "encapsulated interest" view of trust (Cook, 2015). In this view, "[t]rust exists when one party to the relation believes the other party has the incentive to act in his or her interest or to take his or her interests to heart" (Cook et al., 2005, p. 2). Greater emphasis is placed on the relationship between the truster and the trustee and the context in which a trustworthiness judgment is made. Specifically, Cook (2015) argues for consideration of incentives of the trustee, past interactions, and reputational knowledge from other sources; she states:

Too many arguments about trust are decontextualized and too generic to represent the real complexities of trust relationships...trust relations are specific, often deeply embedded in social contexts (social and cultural), and have high information requirements. (p. 128)

By specifying the context of certifiers, I can address some of Cook's concerns with psychologically-based measurements of trustworthiness. In particular, knowing that the trustee is a certifier allows the measurement items to be more specific and relevant to the particular context.

For certifier trustworthiness, it does not matter whether the truster agrees with the organization's mission and goals. Instead, the truster must evaluate how well the organization can achieve its self-generated mission and goals. This differs from Pirson and Malhorta's (2011) inclusion of identification as a dimension of trustworthiness. Identification-based trust results from alignment between a truster and a trustee's shared values (Lewicki & Bunker, 1996).

Notably, Pirson and Malhorta (2011) hypothesize that identification matters only for deep relationships with firms (e.g., for employees and investors). It seems plausible that identification with an certifier's mission and goals is relevant for stakeholders with deep or shallow relationships with the organization.

Following in the tradition of Deephouse and Suchman (2008), I argue that trustworthiness is highly related to, but distinct from, the constructs of legitimacy and reputation. Both legitimacy and reputation are extensively researched constructs grounded in diverse theoretical frameworks and with many perspectives on how they should be defined and measured. Here, I will focus on some of the most common perspectives. From an institutional theory perspective, Suchman (1995) defines *legitimacy* as a widely agreed upon perception of organizations as "desirable, proper or appropriate within some socially constructed system of norms, values, beliefs and definitions" (p. 574). Similarly, corporate legitimacy has been defined as "[t]he degree to which broader publics view a company's activities as socially acceptable and desirable because its practices comply with industry norms and broader societal expectations" (Rindova et al., 2006, p. 55). Common across institutional definitions of legitimacy is that audiences view an entity to fit within the norms of society. Further, granting legitimacy is a mechanism of social control, promoting the structure and practices that judges view as beneficial to themselves and society (Bitektine, 2011). Legitimacy can be an antecedent to trust and perceptions of trustworthiness (Aldrich & Fiol, 1994).

Alexiou and Wiggins (2019) developed a scale of legitimacy perceptions intended to encompass a wide range of conceptualizations of legitimacy. Their scale included three legitimacy dimensions: pragmatic, moral, and cognitive. An organization has *pragmatic legitimacy* when immediate stakeholders perceive that they will gain value from the organization

(Alexiou & Wiggins, 2019; Dart, 2004). Pragmatic legitimacy resembles the ability dimension of trustworthiness. However, pragmatic legitimacy is less specific about why stakeholders perceive they can attain value from an organization than the ability dimension proposed here. The three pragmatic legitimacy items include “this organization creates value for its stakeholders, “[t]he policies of this organization cater to the interests of its stakeholders, and I “believe the activities of this organization benefit their immediate stakeholders.” Hence, pragmatic legitimacy does not capture *how* organizations are expected to meet stakeholder needs.

“*Moral legitimacy* is an active assessment of the degree to which an organization adheres to social norms and shared values in a manner that promotes societal welfare” (Alexiou & Wiggins, 2019, p. 472). This definition of moral legitimacy resembles the definition of integrity. However, the items of moral legitimacy fail to capture relevant aspects of integrity across certifiers, including expectations of consistency and legality. *Cognitive legitimacy* captures whether the organization is understandable as an entity and taken for granted. According to Alexiou and Wiggins (2019), cognitive legitimacy is a more passive and less effortful evaluation of an organization than pragmatic and moral legitimacy. The items of the cognitive legitimacy dimension include “I believe this organization is necessary,” “this organization provides an essential function, and “it is difficult to imagine a world in which this organization did not exist.” During the scale evaluation and validation phases, I will use the moral legitimacy dimension of Alexiou and Wiggins’ (2019) scale to assess the discriminant validity of the CTS.

A definition of reputation proposed by Fombrun and colleagues (2000) is that reputation is “a collective assessment of a company's ability to provide valued outcomes to a representative group of stakeholders” (p. 23). Outcomes of particular importance in reputation judgments are product quality and financial performance (Fombrun & Shanley, 1990; Rindova et al., 2006;

Shapiro, 1983). The final version of the reputation quotient adapted to assess discriminant validity in the present research identifies six dimensions of company reputation: emotional appeal, products and services, vision and leadership, workplace environment, social and environmental responsibility, and financial performance (Fombrun et al., 2000; see Appendix E for the complete scale). The dimensions more likely to correlate to the CTS are emotional appeal, products and services, and workplace environment. An example item of emotional appeal is "I trust this company," and an example item from the products and services dimension is this company "[o]ffers high quality products and services" (Fombrun et al., 2000, p. 253). An example workplace environment item is this company "is well-managed" (Fombrun et al., 2000, p. 253). Trustworthiness differs from reputation in that trustworthiness focuses more narrowly on the primary activities of the certifiers. Furthermore, reputation is often construed as a collective rather than individual judgment.

Scale Development

The second phase of the scale development process is to generate and validate items for each dimension. MacKenzie and colleagues (2011) state:

The ultimate goal of the item generation process is to produce a set of items that fully captures all of the essential aspects of the domain of the focal construct, while minimizing the extent to which the items tap concepts outside of the domain of the focal construct. (p. 304)

Item Generation

To generate items for an initial version of the CTS, I consulted a variety of existing measures of trust and trustworthiness, and the individual dimensions of ability, integrity, and moral authenticity (Mayer & Davis, 1999; Kervyn, Fiske & Malone, 2012; McAllister, 1995;

McKnight et al., 2002; Roy & Shekhar, 2010; Ennew & Sekhon, 2007; Buttner & Goritz, 2008). I ensured items could be applied to the organizational level of analysis and were relevant to certifiers. I also generated several items not captured from the review of existing measures. Then, I reviewed the item list independently and eliminated items that were essentially duplicates of other items. I collected feedback from 13 organizational scholars on an initial set of items, and based on their feedback decided on 43 items to assess for content validity.

Content Validity

Content validity relates to how well items represent the construct's content domain (Kerlinger, 1973; MacKenzie et al., 2011; Straub et al., 2004). The content validation process involves making two judgments. One judgment is whether the individual item represents the construct's content domain. The second judgment is whether the items collectively represent the construct's entire content domain. I conducted two content validity tests, each using panels of online participants. Each content validity test followed a similar procedure.

I adapted the content validity process introduced by Anderson and Gerbing (1991). I chose this approach instead of the one proposed by Hinkin and Tracey (2009) and preferred by MacKenzie and colleagues (2011) due to concerns about retaining participants' attention. Anderson and Gerbing's (1991) process involves participants sorting items into the appropriate construct definition. My approach differed in that the definitions provided for the sorting task were that of the hypothesized CTS dimensions (ability, integrity, and moral authenticity), the construct of *awareness*, and an "unsure" option.

Participants for each test completed an online Qualtrics survey. Upon consent, participants viewed definitions of sustainability certifiers, certifier trustworthiness, ability (tests 1 and 2 only), integrity, authenticity, and awareness (tests 1 and 2 only). Then, participants chose

which dimension they thought best fits the conceptual domain of each item by dragging each item listed on the left-hand side of the screen and dropping it within one of the boxes on the right. Each test included an "Unsure" box on the right-hand side. After completing the sorting task, participants could explain why they placed items in the "Unsure" box, provide any comments on the study, and provide relevant demographic information. In the first test, I asked participants whether they were current faculty members or students. In the second and third tests, I asked participants to provide their age bracket, their highest level of educational attainment, employment status, and whether they currently resided in the U.S. Participants were then thanked for their participation and prompted to exit the survey.

Content Validity Test 1

The item pool for the first test contained 48 total items, 43 items developed to measure certifier trustworthiness, and five items adapted from a brand awareness scale (Yoo & Donthu, 2001). I recruited 120 participants from Prolific's online research platform (mean age group = 2.83, SD = 1.30). All participants were English-speaking and aged 18 years or older. All participants passed the attention check of sorting the statement "Put this item in the "integrity" box" into the integrity box.

I calculated two indices to analyze the data: the proportion of substantive agreement (p_{sa}) and the substantive-validity coefficient (c_{sv} ; Anderson & Gerbing, 1991). The calculations for these indices are shown below, where N is the total number of participants, n_c is the number of participants who sorted the items correctly, and n_o is the maximum number of times an item was sorted into any other construct. The p_{sa} value can range from 0 to 1, where one means that all participants sorted the item correctly. The value of the c_{sv} ranges from -1 to 1, where -1 means

that none of the participants classified the item correctly, and 1 means all of the participants classified the item correctly.

$$p_{sa} = n_c / N$$

$$c_{sv} = (n_c - n_o) / N$$

The p_{sa} and c_{sv} would be the same if participants had two box options to sort items. In this situation, a p_{sa} and c_{sv} value greater than .5 would indicate that participants sorted the item into the hypothesized dimension at a rate greater than chance. In this first content validity test, participants had five concepts to choose from. Therefore, a p_{sa} of over .2 indicates that participants sorted the item into the hypothesized dimension at a rate greater than chance. The critical value for c_{sv} is calculated as follows, where m represents the critical number of assignments.

$$\bar{c}_{sv} = \frac{m - (N - m)}{N} = \frac{2m}{N} - 1$$

The critical number of assignments is found using the probability formula below, assuming a significance level of .05.

$$P(n_c \geq m) < .05$$

I assumed a normal distribution, excluded outlying values of n_c (values in the first and fourth quartiles), and calculated the critical value for $\bar{c}_{sv}$ as .670. Notably, Anderson and Gerbing (1991) state that “in practice, a researcher would most likely employ c_{sv} in a comparative manner, retaining the subset of items with the largest values for each construct, even though values for some items may not attain statistical significance” (p. 735). All ability items had a c_{sv} over .670, but only two integrity items surpassed the threshold and no authenticity items. Excluding the ability items from the estimate of the critical value provides a value of .465 (see Appendix A, Table 2). At this critical value, two authenticity items surpassed the $\bar{c}_{sv}$ threshold.

Thus, participants did not appear to find much distinction between the items intended for authenticity and integrity items. Part of the reason for this lack of distinction in the items between the two hypothesized dimensions may have resulted from survey fatigue due to the high number of items participants sorted. Therefore, I conducted a second content validity test focusing on the authenticity and integrity dimensions and using select items. In the second content validity test, I also tested three new items that were succinct and with strong action verbs to enhance their readability and meaningfulness. These new items included “...prioritizes environmental sustainability.”, “...holds all potential certification holders to the same standards.”, and “...enforces standards consistently”.

 Insert Table 2 about here

Content Validity Test 2

The second test included sorting boxes labeled “integrity,” “authenticity,” and “Unsure.” The item pool contained 16 items, including three new items and one item for a global indicator of trustworthiness. A global indicator item is helpful to assess whether the other items converge with the overall construct and to test for the incremental validity of the scale beyond a single-item measure (MacKenzie et al., 2011). I recruited 120 participants from Prolific's online research platform (mean age group = 2.92, SD = 1.07). All participants were English-speaking and aged 18 years or older. All participants passed the attention check of sorting the statement “Put this item in the “integrity” box” into the integrity box.

I assumed a normal distribution, excluded outlying values of n_c (values in the first and fourth quartiles), and calculated the critical value for $\bar{c}_{sv}$ as .697. One authenticity item surpassed

the critical value (“The certifier is authentic”), and three integrity items surpassed the critical value (“The certifier enforces standards consistently,” “The certifier shows high integrity,” and “The certifier would not knowingly break the law”). Because there were three categories (including the “Unsure” category), a p_{sv} value greater than $.33\bar{3}$ indicates that the item was sorted into the intended category at a rate greater than chance. All of the items surpassed the p_{sv} criteria. The results of the third content validity test supported the hypothesis that authenticity and integrity represent distinct dimensions of certifier trustworthiness.

Insert Table 2 about here

Scale Evaluation and Refinement

Pretest

The scale evaluation and refinement phase includes collecting data to conduct a pretest and, based on the results, removing or modifying items that do not meet the recommended statistical criteria.

Participants

I recruited 370 participants from Prolific's online research platform (mean age group = 3.80, SD = 1.33). One participant was dropped from the survey because the time they took to complete the survey was identified as an outlier, falling significantly below the first quartile of response durations. I removed 21 responses from the data analyses due to evidence of careless responding. I removed 17 responses due to manipulation check failure. I removed two responses due to the failure of at least one attention check. One additional response was dropped because Qualtrics identified it as a potential bot. This resulted in a final sample of 318 participants (Mean

age group = 3.82, SD = 1.33). There were no significant differences in age group, employment status, or educational attainment between the uncleaned sample of 370 participants and the cleaned sample of 318 participants.

I collected demographic information at the end of the online survey. To assess the age group, I asked participants, “How old are you?” Participants selected from one of the following options: 1 = Under 18; 2 = 18-24 years old; 3 = 25-34 years old; 4 = 35-44 years old; 5 = 45-54 years old; 6 = 55-64 years old; 7 = 65+ years old. To assess educational attainment, I asked participants, “What is the highest level of school you have completed or the highest degree you have received?” Participants selected from one of the following options: 1 = Less than high school degree; 2 = High school degree or equivalent (e.g., GED); 3 = Some college but no degree; 4 = Associate degree; 5 = Bachelor degree; 6 = Graduate degree. To assess employment status, I asked participants, “Which of the following categories best describes your employment status?” Participants selected from one of the following options: 1 = Employed, working 1-39 hours per week; 2 = Employed, working 40 or more hours per week; 3 = Not employed, looking for work; 4 = Not employed, NOT looking for work; 5 = Retired; 6 = Disabled, not able to work. See Table 3 in Appendix A for a summary of demographic information for the final sample.

Insert Table 3 about here

Procedure

After participants elected to join the study within the Prolific platform, they were automatically directed to the online Qualtrics survey. Each participant was randomly assigned to one of three conditions: (1) low ability, (2) low authenticity, or (3) low integrity. After reading

about the survey and providing their consent and Prolific identifiers, participants were asked to briefly read about how I defined a “certifying organization” and “environmental sustainability.” Participants were asked one question each about both of these definitions and if they failed, they were given the correct answer. Participants across all conditions were then asked to read information about “Green Tech,” a fictional certifier that certifies personal electronic devices. The next part of the prompts varied depending on the participants’ assigned condition; please find the prompt text in Figures 1a-c below.

Figure 1a – Low Ability

- Green Tech was founded in California in 2010.
- Green Tech started with the certification of personal computers and later expanded to the certification of all personal electronic devices.
- Green Tech has never been charged with any illegal activity.
- Experts accept the process that Green Tech uses to review applications.

- Last year, Green Tech lost most of its money in the stock market.

Green Tech had to lay off over half of its employees, including green technology specialists. It may close its Nashville office to reduce costs further.

Figure 1b – Low Integrity

- Green Tech was founded in California in 2010.
- Green Tech started with the certification of personal computers and later expanded to the certification of all personal electronic devices.
- The value of Green Tech’s profits are similar to other organizations of the same size.
- Experts accept the process that Green Tech uses to review applications.

- Last year, the IRS charged Green Tech with tax fraud.

The IRS claims Green Tech knowingly underreported its income on its previous federal income tax return. The case is planned to go to trial later this year.

Figure 1c – Low Authenticity

- Green Tech was founded in California in 2010.
- Green Tech started with the certification of personal computers and later expanded to the certification of all personal electronic devices.
- Green Tech's profits are in line with similar organizations of the same size.
- Green Tech has never been charged with any illegal activity

- Last year, Green Tech received a lot of bad press.

One writer claimed that Green Tech does not actually make sure that certification applicants follow its standards. Instead, Green Tech approves all applications if its fees are paid.

Measures

Perceived trustworthiness was measured using 23 total items based on the results of the content validity tests, including one global reflective indicator (“...is trustworthy”). Participants rated the extent to which they agreed or disagreed with each item on a scale from 1 (strongly disagree) to 7 (strongly agree). Responses were coded such that higher scores indicated higher levels of perceived trustworthiness. See Appendix A for a list of items and Appendix F for a correlation matrix.

To test the discriminant validity of the CTS, I included measures for the related constructs of cognitive legitimacy and reputation. I used a modified version of the Reputation Quotient (Franco et al., 2009; see Appendix E) to measure reputation and the Cognitive Legitimacy subscale of the legitimacy scale developed by Alexiou and Wiggins (2019).

Analyses and Results

Factor analysis. The factorability of the certifier trustworthiness items was demonstrated by the significant correlation of each item with at least one other item (see Appendix A, Table 4). To determine the factor structure of the CTS, I conducted parallel analysis and principal axis factoring in R. In a parallel analysis, R generated a random dataset with the same size and

structure as the pretest dataset and compared the eigenvalues from the generated dataset to the pretest dataset. The number of factors estimated from the parallel analysis is based on the point at which the eigenvalues from the pretest dataset are greater than the eigenvalues from the generated dataset. In this case, the parallel analysis suggested that there are three factors.

I used principal axis factoring with an oblique rotation (Promax) to extract the three potential factors; an oblique rotation is preferable when the factors are expected to be correlated (see 2). The eigenvalues for the three factors were 6.29, 5.44, and 4.78, suggesting that each factor represents a meaningful pattern in the data and should be retained. I then removed the global indicator item (“...is trustworthy”), two items due to cross-loading over .32 (“... is authentic.”, “...is sincere.”) and two integrity items that loaded with authenticity (“...holds all potential certification holders to the same standards.”, “... enforces standards consistently.”), resulting in 18 final items, including the global indicator “...is trustworthy.”

A second principal axis factoring resulted in acceptable eigenvalues of 5.41, 4.54, and 2.87. All items had extracted communalities above the typically recommended cutoff of .50 ($M = .75$, $SD = .10$; Garson, 2022; Kline, 2013). The three factors explained 75% of the variance in the items. The first factor represented *ability* and explained the most variance (.32, nine items), followed by *authenticity* (.27, five items) and *integrity* (.17, three items; see Table 4). Each identified factor demonstrated high reliability (Ability $\alpha = .92$, Authenticity $\alpha = .96$, Integrity $\alpha = .93$). There were significant inter-subscale associations with correlations ranging between .51 to .76, with Authenticity and Integrity correlating the highest at .76. Parallel analysis again indicated three factors.

 Insert Table 4 about here

Predictive and discriminant validity. I tested the predictive value of the working CTS, including its convergent and discriminant validity. I hypothesized that the composite CTS and each theorized dimension (ability, authenticity, and integrity) would positively correlate with the tested outcome variable of *purchase intention*. I indeed found that each dimension positively predicted purchase intention (β s = .22 to .39, $p < .001$), as did the composite CTS measure (see Table 5 in Appendix A; $\beta = .932$, $p < .001$, $R^2 = .538$).

Perceived trustworthiness should be positively correlated yet distinct from the organization's reputation and perceived legitimacy. The composite CTS measure positively predicted reputation ($\beta = .808$, $p < .001$; see Table 6 in Appendix A), as did each CTS dimension independently (β s = .148 to .457, $p < .001$; see Table 6). Additionally, the composite CTS positively predicted cognitive legitimacy ($\beta = .552$, $p < .001$), as did authenticity ($\beta = .341$, $p < .001$) and ability ($\beta = .241$, $p < .001$); integrity did not predict cognitive legitimacy ($\beta = -.047$; See Table 7 in Appendix A).

The positive relationship between the composite CTS and purchase intention held when reputation and cognitive legitimacy were added into the regression model as predictors ($\beta = .284$, $p < .001$, $R^2 = .655$; see Table 8 in Appendix A). When each CTS dimension was separately entered into a regression model with cognitive legitimacy, each dimension was predictive of purchase intention (β s = .22 to .30, $p < .001$, $R^2 = .59$). However, when each CTS dimension was separately entered into a regression model with reputation alone and reputation with cognitive legitimacy, only the authenticity dimension remained predictive (see Appendix A, Table 8).

Goodness of fit. For the goodness of fit measures, MacKenzie, Podsakoff, and Podsakoff (2011) recommend the following cutoffs: .95 for CFI, .08 for SRMR, and .06 for RMSEA. I conducted SEM in R using the lavaan package to assess the goodness of fit of the working model for the CTS. The working model has an RMSEA of .00, a Comparative Fit Index (CFI) of 1.00, and a Standardized Root Mean Square Residual (SRMR) of .00, all indicating a good fit of the working model.

Final Test

The final scale validation phase involves evaluating whether the CTS possesses discriminant and nomological validity and whether the items accurately represent certifier trustworthiness and the dimensions of perceived ability, authenticity, and integrity. For the final test of the CTS, I conducted a vignette experiment manipulating the degree of the certifier's product diversification; this is the same experiment discussed in paper two. However, I focus on the most relevant results to the present research here. A vignette experiment is appropriate for these purposes because understanding a specific certifier is necessary to assess the perceived trustworthiness of the certifier. With a simple survey, there is no guarantee that the participants have information on a given certifier, let alone that the participants have the same information on a given certifier. A vignette experiment provides the necessary context for participants to make informed judgments while keeping the information that participants have access to consistent. I hypothesized that certifiers with greater product diversification would receive lower ratings for each trustworthiness dimension.

Participants

Five hundred twenty participants were recruited from the online research platform Prolific. I required participants to be aged 18 years or older, with at least a high school degree or

its equivalent, and to speak English fluently; participants were located in the United States at the time of the experiment, and the sample was gender-balanced. I removed 84 observations from the sample due to participants failing the manipulation check at the end of the study, and I removed 19 observations that had at least one attention check failure. I removed one observation where the participant was identified as potential bot by the survey platform Qualtrics. Finally, I removed 14 observations where a response to at least one item of the CTS was missing.

The final sample included 402 participants, about seventy-seven percent of the initial sample. Slightly over half of the participants were female (51.74%), and the mean age was 39.96 years (S.D. 14.28). The observation removal rate is consistent with empirical evidence on data quality using a Prolific sample (Douglas et al., 2023). Welch's two-sided t-tests indicated that there were no significant differences in age or gender between the original sample and the final sample.

Procedure

The experiment used a between-subjects design; participants were randomly assigned to one of three conditions where the only difference between conditions were the degree of product specialization (diversification) by the certifier. The three conditions were (1) specialization, (2) related diversification, and (3) unrelated diversification. The experiment employed a vignette scenario, where participants were asked to imagine that they were the owner of a chocolate store making a decision about a supplier certified by the fictional certifier "Earth Friendly Goods." The experiment took an average of about nine minutes to complete. Before beginning the experiment, participants were assured of anonymity and asked to consent. For each condition, participants received the same instructions upon providing consent. The instructions read:

The goal of this survey is to better understand how the public reviews and evaluates certifying organizations.

You will first read important background information for this study and then be asked two questions about what you read.

Then, you will be asked to imagine yourself in a scenario related to the fictional organization "Earth Friendly Goods" and be asked to answer several questions about the organization.

Then, participants were provided background information on what was meant by a “certifying organization” and by “environmental sustainability” for the study. They were asked to respond to questions about the meaning of each term, and if they answered incorrectly, they were shown the correct response.

The following prompt explained the vignette scenario and was consistent across assigned conditions. The prompt read:

Imagine you're the proud owner of a cozy local chocolate shop. You have a wide range of chocolates on display, from tasty truffles to lush chocolate bars and mouth-watering fudge. Your customers have high expectations when it comes to both taste and presentation, and they also care about the environment. That's why you're choosing a chocolate supplier that is certified sustainable.

The following and last vignette prompt did vary by condition and described the certifier “Earth Friendly Goods.” The specialization condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies only chocolate and the raw foods for chocolate**, such as cocoa beans.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

The related diversification condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies many types of food products**, such as chocolate, coffee, vegetables, fruits, and meats.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

The unrelated diversification condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies all types of products**, such as personal electronic devices, paper products, toys, and foods.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

Next, participants were asked to respond to items to respond to items for the CTS, purchase intention, warmth and competence, political ideology, environmental concern, familiarity with certifiers, demographic variables, and a manipulation check to assess their understanding of the condition to which they were randomly assigned. Finally, participants were thanked, provided with contact information to raise any questions or concerns, and prompted to exit the survey.

Measures

Certifier trustworthiness (CTS). The survey prompted participants to “Please rate your agreement with the following statements about Earth Friendly Goods.” Participants responded on a 7-point Likert scale ranging “Strongly disagree” to “Strongly agree,” with an additional “N/A / I don’t know” option. Items for the working CTS were based on the results of the pretest, and are listed in Appendix I.

Purchase intention. Participants rated four items adapted from Moon and colleagues' (2008) purchase intention measure to assess the likelihood of purchasing a product certified by Earth Friendly Goods. The three items were: (1) I would be willing to pay more for a product certified by Earth Friendly Goods; (2) I would recommend products certified by Earth Friendly Goods to a friend; and (3) I would consider Early Friends Goods certification as part of my purchase decision. Participants rated all items on a 7-point Likert scale ranging "Strongly disagree" to "Strongly agree," with an additional "N/A / I don't know" option.

Control variables. Participants responded to nine measures of variables potentially driving their purchase intention.

Warmth and competence of the organization. Previous research indicates that audiences engage with an entity based, in part, on perceptions of the entity's warmth and competence (e.g., Kervyn, Fiske, & Malone, 2012). Perceived trustworthiness and perceived warmth and competence are considered to be related in various ways. For example, Fiske and colleagues (2007) include trustworthiness as part of the warmth dimension. Thus, the warmth and competence measure was used to test the discriminant validity of the CTS. I measured the perceived warmth and competence of the certifier following the procedure used by Aaker, Garbinsky, and Vohs (2012). Participants used a 5-point Likert scale ranging from "Not at all" to "Extremely" with a "N/A / I don't know option" to respond to the question, "To what extent do the following words describe Earth Friendly Goods?" for each of the words "competent," "capable," "warm," and "friendly."

Environmental concern. Participants with greater concern for the environment, in general, should be more likely to perceive a sustainability certifier such as Earth Friendly Goods as trustworthy due to greater identification with the organization (e.g., Woolley & Fishbach,

2017). Therefore, they should also be more likely to favor products verified by a sustainability certifier than participants with lower environmental concern. I measured environmental concern using four items, and participants responded on a 7-point agreement scale (see Appendix H).

Political ideology. A liberal political ideology is positively associated with support for environmental causes (e.g., Dunlap & McCright, 2011). Therefore, a liberal political ideology is likely to positively predict the dependent variable of purchase intention. To measure participants' political ideology, I asked, "When it comes to politics, would you describe yourself as liberal, conservative, or neither liberal nor conservative?" with the seven response options ranging from 1 = Extremely Liberal to 7 = Extremely Conservative.

Familiarity with certifiers. Familiarity with a specific entity has been shown to increase trust in some contexts (e.g., Ha & Perks, 2005). Thus, familiarity with a particular organizational form, such as a certifier, may similarly lead to increased perceptions of trustworthiness. I evaluated participants' familiarity with certifiers using a 3-item measure adapted from Kent and Allen (1994). The items included "I am familiar with certifying organizations.," "I am experienced with certifying organizations.," and "I am knowledgeable about certifying organizations." Participants responded on an agreement scale ranging from 1 = Strongly Disagree to 7 = Strongly Agree.

Demographics. Additional control variables include participants' age, gender, highest level of education attained, and employment status. Age was measured as a continuous variable, while the other control variables were measured based on responses to multiple-choice questions. See Table 9 in Appendix A for participants' demographic characteristics by condition. No significant differences existed for any of the demographic characteristics measured between conditions.

Insert Table 9 about here

Analyses and Results

The final experimental test of the CTS provided an opportunity to refine the scale items further and assess the scale's predictive, discriminant, and convergent validity.

Scale refinement and dimensionality. The tested version of the CTS contained 19 items, including eight for the ability dimension, five for the authenticity and integrity dimension, and one for the global indicator item. I conducted EFA on the full scale (19 items) to identify potentially redundant items. I used principal axis factoring with an oblique rotation (Promax) to extract the three potential factors. The eigenvalues for the three factors were 6.98, 4.16, and 3.62, suggesting that each factor represents a meaningful pattern in the data and should be retained.

I then conducted EFA without the global indicator item "...is trustworthy." because its inclusion was for validation purposes, and two integrity items due to cross-loading with the authenticity dimension ("...tries hard to be fair in dealings with others." and "...would do its best to follow regulations."). Again, the eigenvalues for the three factors were 6.29, 3.85, and 2.83, suggesting that each factor represents a meaningful pattern in the data and should be retained. Due to the high number of ability items remaining, I removed the two lowest loading ability items ("...has the information it needs to achieve its stated mission and goals." and "...has the resources it needs to set proper certification standards."). The EFA of the final CTS with 16 items (5 authenticity, four integrity, and seven ability) had eigenvalues of 5.08, 3.80, and 2.76.

I conducted ANOVA for each dimension based on the participants' assigned condition to further test the distinction between the theorized dimensions. For the authenticity dimension,

there was a significant main effect of condition ($F = 3.562, p < .05$). Tukey's honestly significant difference (HSD) test shows that there was a significant difference in authenticity dimension scores between the specialization condition and the unrelated diversification condition ($p < .05$) and a marginally significant difference between the specialization condition and the related diversification condition ($p < .1$). For the integrity dimension, there was a significant main effect of condition ($F = 5.866, p < .05$). Tukey's HSD test shows that there was a significant difference of integrity dimension scores between the specialization condition and the unrelated diversification condition ($p < .05$) and between the specialization condition and the related diversification condition ($p < .01$). There were no significant differences for the ANOVA test for the ability dimension.

Predictive and discriminant validity. For predictive validity to hold, the composite CTS and each theorized dimension (ability, authenticity, and integrity) should positively correlate with the outcome variable of *purchase intention*. Each dimension positively correlated with purchase intention ($r_s = .512$ to $.561, p < .001$), as did the composite CTS ($r = .902, p < .001$). The CTS also positively predicted purchase intention when including all of the control variables (see Table 10 in Appendix A, $B = .812, p < .001, R^2 = .400$). Pearson's correlation coefficient between the CTS and the warmth and competence measure was significant but less than .71, indicating good discriminant validity ($r = .356, p < .001$; Anderson & Gerbing, 1998). The correlation coefficient between the CTS also indicated good discriminant validity ($r = .132, p < .01$).

Insert Table 10 about here

Norm development. The final phase of the scale development process is norm development, establishing expectations for the distribution of results (MacKenzie et al., 2011). Establishing norms for a scale eases interpretation by researchers, allowing them to classify a score as average or, relatively high or low (Spector, 1992). A test should be conducted using a representative sample to establish the norms of the scale (Leong & Austin, 2006). Based on the collected demographic information for the CTS final test, the sample of participants was generally aligned with the demographic characteristics of adults currently living in the United States. For example, 51.7 percent of participants identified as female, while according to U.S. Census Bureau estimates, approximately 51.1 percent of people aged 18 to 64 in the United States identified as female as of 2022 (2022). Notably, the U.S. Census Bureau estimates that 32.9 percent of people in the United States aged 18 years or older have attained at least a bachelor's degree, while in the CTS final test sample, about 52 percent of participants reported to have at least a bachelor's degree (2022). This discrepancy is partially explained by the experimental requirement for participants to have at least a high school degree.

Scale norms can then be established by calculating the mean and standard deviation of results (MacKenzie et al., 2011). The expected distribution of the scale results is variable based on specific contexts and sample populations. Therefore, the scale norms should be adjusted upon further application of the CTS. In the CTS final test, the mean of the overall CTS is 6.33, with a standard deviation of 0.79. The distribution of the CTS is left-skewed and, per a Shapiro-Wilk normality test, is non-normal ($p < .001$). Please see Table 11 in Appendix A for the descriptive statistics of the CTS and each dimension.

Discussion

Certifier trustworthiness is the truster's confidence that the certifier can be counted upon to pursue and achieve its stated mission and goals. Developing a psychometrically valid certifier trustworthiness construct and corresponding measure provides a valuable tool for understanding these unique organizational forms and assessing their effectiveness. The present research conceptualizes certifier trustworthiness as a tri-dimensional construct following Mayer and colleagues' (1995) seminal model of interpersonal trustworthiness. The first dimension of ability closely follows how ability represents an audience member's belief that the organization has the resources and expertise to achieve its mission and goals. The second dimension of integrity encompasses the perception that the organization strives to make choices consistent with its mission and goals. Finally, authenticity relates to the perception that an organization sincerely values achieving its mission. Across five studies, I demonstrate that the certifier trustworthiness possesses construct validity, and the developed CTS with formative indicators has a three-dimensional structure with discriminant, nomological, and predictive validity.

The development and validation of the CTS makes a significant contribution to research on organizational trust and trustworthiness. It extends Mayer and colleagues' (1995) model of interpersonal trustworthiness to the organizational level. In-depth knowledge of the drivers of certifier trustworthiness are particularly critical as trustworthiness is especially relevant to certifiers. The CTS will be a valuable tool for better understanding the antecedents and consequences of certifier trustworthiness. For example, future research can use the scale to explore the effect of governance and financial structures on perceptions of an organization's ability, integrity, and authenticity. Trustworthiness predicts performance in many settings, and future research should also explore the relationship between trustworthiness and performance for certifiers.

Limitations and Future Directions

While the CTS is a promising measure, capturing the multidimensional nature of certifier trustworthiness, future research should address the limitations of the testing conducted thus far. First, all samples used to validate the CTS came from Prolific's online platform. The sample of participants was relatively representative of the U.S. adult population along the demographics measured. However, participants were significantly more likely to have completed a bachelor's degree than the final sample. The samples used may also substantially differ from the general public across variables not measured that influence the outcomes of the studies. Second, using vignette experiments, while valuable for experimental control, may only partially capture real-world decision-making processes. In the future, researchers should use participants with experience evaluating certifications to test the CTS. Ideally, the CTS should be used during a decision-making process by a producer interested in attaining certification.

Additionally, a limitation of the CTS validation process is that the scenarios primarily focused on environmental certifiers, limiting its applicability to other certification contexts. Future research could explore the applicability of the CTS across different certifying organizations to assess its robustness and generalizability. Furthermore, establishing scale norms was based on a specific sample and may not fully capture the variability in certifier trustworthiness perceptions across different populations. Future studies could replicate this process with more extensive and diverse samples to provide more robust norms for interpretation. Finally, the relationship between the CTS and theoretically related variables warrants continued evaluation.

Paper 1 Appendices

Appendix A: Content Validity Test 1

Table 1

Content Validity Test 1 Results

Item	P _{SA}	C _{SV}	Included in test 2
Ability dimension			
...is competent	.923	.846	Yes
...is very knowledgeable about product sustainability.	.923	.846	
...has the resources it needs to achieve its stated mission and goals.	.923	.846	Yes, modified
...is an expert in its area.	.923	.846	Yes, modified
...has access to the people needed to achieve its stated mission and goals.	.923	.846	No
...is efficient.	.769	.692	No
...has the information it needs to achieve its stated mission and goals.	.923	.846	Yes
Authenticity dimension			
...believes the statements it makes.	.615	.615	No
...is genuinely interested in improving the sustainability of pet products.	.692	.538	Yes, modified
...is honest about its goals.	.385	.077	Yes, modified
...seems to want to make a positive impact on the sustainability of the pet supply industry.	.462	.231	No
...is authentic.	.615	.462	No
...has a good grasp of its purpose as an organization.	.385	.154	No
...will recognize when its choices deviate from its goals.	.308	.000	No
...strives to be transparent about its choices.	.462	.154	No
...will make sacrifices to maintain its mission and goals.	.692	.538	No
Integrity dimension			
...ensures its certified products are sustainability made even though it may take longer.	.615	.385	No
...operates scrupulously.	.462	.308	No
...shows high integrity.	.923	.846	Yes
...acts in the best interests of its stakeholders.	.846	.769	Yes, modified
...is consistent in what it does.	.231	.154	No
...would keep its commitments.	.462	.385	No
...is reliable.	.308	-.154	No

Table 2

Content Validity Test 2 Results			
Item	P _{SA}	C _{SV}	Included in pretest
Authenticity dimension			
I trust the certifier to have environmental sustainability's best interests at heart.	.692	.442	Yes, modified
The certifier aims to set standards that support environmental sustainability.	.700	.425	Yes
The certifier genuinely supports environmental sustainability.	.808	.642	No
The certifier is authentic.	.925	.883	Yes
The certifier is sincere.	.667	.425	Yes
The certifier prioritizes environmental sustainability.	.725	.483	Yes
The certifier truly cares about environmental sustainability.	.792	.617	Yes
The certifier would not knowingly set standards that undermine what is best for environmental sustainability.	.575	.208	No
Integrity dimension			
I trust the certifier to do what it says it will do.	.592	.242	No
The certifier enforces standards consistently.	.825	.750	Yes
The certifier holds all potential certification holders to the same standards.	.700	.408	Yes
The certifier shows high integrity.	.850	.708	No
The certifier tries hard to be fair in dealings with others.	.708	.458	No
The certifier would do its best to follow regulations.	.808	.642	Yes
The certifier would not knowingly break the law.	.900	.825	Yes

Table 3*Pretest Demographic Characteristics of Participants*

	Ability		Authenticity		Integrity	
	N	%	N	%	N	%
Age group						
18-24	15	14.0	15	15.2	15	13.4
25-34	41	38.3	30	30.3	44	39.3
35-44	21	19.6	18	18.2	24	21.4
45-54	19	17.8	22	22.2	14	12.5
55-64	8	7.5	8	8.1	12	10.7
65+	3	2.8	6	6.1	3	2.7
Educational attainment						
High school degree or equivalent (e.g., GED)	14	13.1	15	15.2	11	9.8
Some college but no degree	26	24.3	25	25.3	21	18.8
Associate degree	9	8.4	7	7.1	12	10.7
Bachelor degree	42	39.3	41	41.4	51	45.5
Graduate degree	16	15.0	11	11.1	17	15.2
Employment status						
Employed, working 1-39 hours per week	31	29.0	34	43.3	36	32.1
Employed, working 40 or more hours per week	40	37.4	36	36.4	46	41.1
Not employed, looking for work	14	13.1	11	11.1	13	11.6
Not employed, NOT looking for work	11	10.3	8	8.1	10	8.9
Retired	3	2.8	7	7.1	5	4.5
Disabled, not able to work	8	7.5	3	3.0	2	1.8

Table 4*Pretest Factor Analysis*

	Factor Loading			
	Ability	Authenticity	Integrity	Communality
c. ...genuinely supports environmental sustainability.	-.028	.980	-.022	1.00
d. ...can be trusted to have environmental sustainability's best interests at heart.	-.001	.760	.192	1.13
e. ...truly cares about environmental sustainability.	-.005	.950	.007	1.00
f. ...aims to set standards that support environmental sustainability.	.066	.900	-.089	1.03
g. ...prioritizes environmental sustainability.	-.052	.920	.043	1.01
h. ...shows high integrity.	-.010	.020	.933	1.00
j. ...would not knowingly break the law.	-.109	-.070	.960	1.04
k. ...would do its best to follow regulations.	-.057	.150	.820	1.08
n. ...is able to achieve its goals.	.813	-.040	-.016	1.01
o. ...has the expertise needed to achieve its goals.	.863	.020	-.065	1.01
p. ...has the expertise needed to set appropriate standards.	.792	.160	-.056	1.09
q. ...has the resources it needs to check if products meet its standards.	.958	-.130	-.109	1.06
r. ...has the information it needs to achieve its stated mission and goals.	.743	.060	.028	1.01
s. ...has the resources it needs to set proper certification standards.	.950	-.180	-.108	1.09
t. ...is competent.	.536	.110	.285	1.62
u. ...knows how to enforce standards.	.588	.160	.172	1.33
v. ...shows high ability.	.583	.150	.209	1.40
Eigenvalue	4.536	5.406	2.867	
% of Total Variance	.267	.318	.169	
Total Variance	.754			

Table 5*Regression of CTS on Purchase Intention*

	Purchase Intention	
	(1)	(2)
Independent variables		
CTS (composite)		0.932*** (0.050)
Ability	0.277*** (0.063)	
Authenticity	0.395*** (0.064)	
Integrity	0.213*** (0.051)	
Control variables		
Age	0.001	0.001
Employment status	-0.004	-0.001
Educational Attainment	-0.002	-0.003
Constant	-.055	-0.104*
Observations	318	318
R ²	0.559	0.538
F-statistic	65.69	91.28

Note. Standard errors in parentheses.

** $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$; two-tailed tests.*

Table 6

<i>Regression of CTS on Reputation</i>		
	Reputation	
	(1)	(2)
Independent variables		
CTS (composite)	-	0.808*** (0.030)
Ability	0.457*** (.039)	-
Authenticity	0.148*** (0.039)	-
Integrity	0.213*** (0.031)	-
Control variables		
Age	-0.006 (0.004)	-0.007 (0.004)
Employment status	0.006 (0.004)	0.005 (0.004)
Educational Attainment	-0.002 (0.004)	-0.003 (0.003)
Constant	-0.001 (0.029)	0.007 (0.028)
Observations	318	318
R ²	0.711	0.705
F-statistic	127.500	187.100

Note. Standard errors in parentheses.

*†p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests*

Table 7*Regression of CTS on Cognitive Legitimacy*

	Cognitive Legitimacy	
	(1)	(2)
Independent variables		
CTS (composite)	-	0.552*** (0.061)
Ability	0.241*** (0.078)	-
Authenticity	0.341*** (0.079)	-
Integrity	-0.047 (0.063)	-
Control variables		
Age	.025** (.008)	.027** (.008)
Employment status	-.017* (.008)	-.015 (.008)
Educational Attainment	-.006 (.007)	-.004 (.007)
Constant	.244*** (.059)	.229*** (.061)
Observations	318	318
R ²	.266	.251
F-statistic	18.82	26.27

Note. Standard errors in parentheses.

p < .1; *p < .05; **p < .01; *p < .001; two-tailed tests*

Table 8

Regression of CTS on Purchase Intention

		Purchase Intention		
	(1)	(2)	(3)	(4)
<i>Independent variables</i>				
CTS (composite)	.284*** (.079)	-	-	-
Ability	-	.085 (.065)	-.082 (.067)	.218*** (.061)
Authenticity	-	.237*** (.057)	.279*** (.057)	.313*** (.062)
Integrity	-	.070 (.047)	.046 (.048)	.225*** (.049)
<i>Control variables</i>				
Reputation	.687*** (.083)	.709*** (.082)	.787*** (.081)	-
Cognitive legitimacy	.168*** (.041)	.158*** (.040)	-	.242*** (.044)
Age	.002** (.006)	.001 (.006)	.006 (.006)	-.005 (.006)
Employment status	-.005 (.006)	-.008 (.006)	-.011 (.006)	-.002 (.006)
Educational Attainment	.000 (.005)	.000 (.005)	-.001 (.005)	-.001 (.006)
Constant	-.147* (.041)	-.093* (.042)	-.054 (.042)	-.114* (.047)
Observations	318	318	318	318
R ²	.655	.677	.661	.599
F-statistic	98.22	80.95	86.39	66.11

Note. Standard errors in parentheses.

p < .1; **p < .05; *p < .01; ****p < .001; two-tailed tests*

Table 9*Final Test Demographic Characteristics of Participants*

	Specialist		Related diversification		Unrelated diversification	
	N	%	N	%	N	%
Gender						
Female	73	.53	59	.44	76	.58
Male	61	.44	72	.54	52	.40
Non-binary / third gender	4	.03	2	.02	3	.02
Educational attainment						
High school degree or equivalent (e.g., GED)	16	.12	22	.17	17	.13
Some college but no degree	30	.22	35	.26	31	.24
Associate degree	14	.10	13	.10	15	.11
Bachelor degree	62	.45	49	.37	52	.40
Graduate degree	16	.12	14	.11	16	.12
Employment status						
Employed, working 1-39 hours per week	39	.28	47	.35	33	.25
Employed, working 40 or more hours per week	52	.38	49	.37	49	.37
Not employed, looking for work	10	.07	14	.11	14	.11
Not employed, NOT looking for work	3	.02	2	.02	6	.05
Retired	12	.09	9	.07	12	.09
Homemaker	9	.07	5	.04	7	.05
Student	10	.07	5	.04	7	.05
Disabled, not able to work	3	.02	2	.02	3	.02
	Mean	SD	Mean	SD	Mean	SD
Age	39.15	14.51	40.5	13.57	40.29	14.79

Table 10

<i>Regression of CTS on Purchase Intention</i>			
	Purchase intention		
	(1)	(2)	(3)
Independent variables			
CTS (composite)	0.812*** (0.064)	-	-
Ability	-	0.420*** (0.098)	0.422*** (0.097)
Authenticity	-	0.312** (0.095)	0.342*** (0.092)
Integrity	-	0.108 (0.067)	0.111 (0.067)
Control variables			
Environmental concern	0.186* (0.084)	0.188* (0.085)	0.204* (0.084)
Political ideology	-0.068** (0.025)	-0.070** (0.026)	-0.074** (0.025)
Familiarity with certifiers	0.101** (0.031)	0.102** (0.032)	0.099** (0.030)
Warmth and competence	0.099 (0.062)	0.097 (0.063)	-
Age	-0.003 (0.003)	-0.003 (0.004)	-
Employment status	0.060* (0.025)	0.061* (0.025)	0.057* (0.025)
Educational Attainment	0.015 0.033	0.014 (0.033)	-
Constant	-1.541** (.574)	-1.619** (.583)	-1.425** (.500)
Observations	402	402	402
R ²	0.401	0.402	0.397
F-statistic	29.17	23.85	37.08

Note. Standard errors in parentheses.

p < .1; **p < .05; *p < .01; ****p < .001; two-tailed tests*

Table 11. Respondent Characteristics by Condition (Experimental Data)

	Specialization		Related Diversification		Unrelated Diversification	
	Count	%	Count	%	Count	%
Gender						
Female	73	52.9%	59	44.4%	76	58.0%
Male	61	44.2%	72	54.1%	52	39.7%
Free response	4	2.9%	2	1.5%	3	2.3%
Educational attainment						
High school degree or equivalent	16	11.7%	22	16.5%	17	13.0%
Some college but no degree	30	13.0%	35	26.3%	31	23.7%
Associate degree	14	23.7%	13	9.8%	15	11.5%
Bachelor degree	62	39.7%	49	36.8%	52	39.7%
Graduate degree	16	12.2%	14	10.5%	16	12.2%
Employment status						
Employed, working 1-39 hours per week	39	28.3%	47	35.3%	33	25.2%
Employed, working 40 or more hours per week	52	37.7%	49	36.8%	49	37.4%
Not employed, looking for work	10	7.3%	14	10.5%	14	10.7%
Not employed, not looking for work	3	2.2%	2	1.6%	6	4.6%
Retired	12	8.7%	9	6.8%	12	9.2%
Homemaker	9	6.5%	5	3.8%	7	5.3%
Student	10	7.3%	5	3.8%	7	5.3%
Disabled, not able to work	3	2.2%	2	1.5%	3	2.3%
	Mean		Mean		Mean	
Age	39.1		40.5		40.3	

Table 12: CTS Descriptive Statistics for Norm Development

Dimension	Mean	Standard deviation	Skewness	Kurtosis
CTS (composite)	6.328	0.792	-0.674	3.244
Ability	5.902	0.805	-0.726	3.132
Authenticity	6.052	0.770	-0.902	4.123
Integrity	7.781	1.204	-0.692	3.082

Note. Standard errors in parentheses.

*[†] $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$; two-tailed tests*

Appendix B: Content Validity Test 1 Items

- i. This certifier is competent.
- ii. The certifier is able to achieve its goals.
- iii. The certifier is an expert in environmental sustainability.
- iv. The certifier has the information it needs to achieve its stated mission and goals.
- v. The certifier has the expertise needed to achieve its goals.
- vi. The certifier knows how to enforce standards.
- vii. The certifier has the resources it needs to check if companies meet its standards.
- viii. The certifier is competent at setting standards.
- ix. The certifier knows how to set standards.
- x. The certifier is able to set rigorous standards.
- xi. The certifier has the resources it needs to set proper certification standards.
- xii. The certifier has the expertise needed to set appropriate standards.
- xiii. The certifier is authentic.
- xiv. The certifier is sincere.
- xv. The certifier can be relied upon to give honest recommendations.
- xvi. The certifier is honest about its mission.
- xvii. The certifier genuinely supports environmental sustainability.
- xviii. The certifier really looks out for what is important to environmental sustainability.
- xix. I trust the certifier to have environmental sustainability's best interests at heart.
- xx. The certifier truly cares about environmental sustainability.
- xxi. The certifier would not falsely certify a product.
- xxii. The certifier would not certify bad products just to make money.
- xxiii. The certifier aims to set standards that support environmental sustainability.
- xxiv. The certifier would not knowingly set standards that undermine what is best for environmental sustainability.
- xxv. The certifier shows high integrity.
- xxvi. The certifier acts in the best interests of its customers.
- xxvii. The certifier tries hard to be fair in dealings with others.
- xxviii. You can rely on this certifier.
- xxix. The certifier would not knowingly break the law.
- xxx. The certifier would do its best to follow regulations.
- xxxi. The certifier acts in the best interests of its stakeholders.
- xxxii. I trust the certifier to do what it says it will do.
- xxxiii. The certifier spends the required time and resources to make sure that products meet its standards.
- xxxiv. The certifier carefully checks that products meet its standards.
- xxxv. I believe that the certifier follows best practices when setting standards.
- xxxvi. The certifier's standards are consistent with its mission and goals.

Appendix C: Content Validity Test 2 Items

- a. The certifier is authentic.
- b. The certifier is sincere.
- c. The certifier genuinely supports environmental sustainability.
- d. I trust the certifier to have environmental sustainability's best interests at heart.
- e. The certifier truly cares about environmental sustainability.
- f. The certifier aims to set standards that support environmental sustainability
- g. The certifier would not knowingly set standards that undermine what is best for environmental sustainability.
- h. The certifier shows high integrity.
- i. The certifier tries hard to be fair in dealings with others.
- j. The certifier would not knowingly break the law.
- k. The certifier would do its best to follow regulations.
- l. I trust the certifier to do what it says it will do.
- m. The certifier prioritizes environmental sustainability.
- n. The certifier holds all potential certification holders to the same standards.
- o. The certifier enforces standards consistently.

Appendix D: Revised Certifier Trustworthiness Scale for Pretest

Prompt: “Please rate your agreement with the following statements about Green Tech. Use a scale from 1 – 7 where 1 = Highly Disagree and 7 = Highly Agree.

Green Tech...”

Perceived Ability

- ...has the expertise needed to achieve its goals.
- ...has the expertise needed to set appropriate standards.
- ...is competent.
- ...has the information it needs to achieve its stated mission and goals.
- ...has the resources it needs to set proper certification standards.
- ...has the resources it needs to check if products meet its standards.
- ...knows how to enforce standards.
- ...is able to achieve its goals.
- ...shows high ability.

Perceived Integrity

- ...enforces standards consistently.
- ...holds all potential certification holders to the same standards.
- ...tries hard to be fair in dealings with others.
- ...would not knowingly break the law.
- ...would do its best to follow regulations.
- ...shows high integrity.

Perceived Authenticity

- ...can be trusted to have environmental sustainability's best interests at heart.
- ...truly cares about environmental sustainability.
- ...aims to set standards that support environmental sustainability.
- ...prioritizes environmental sustainability.
- ...genuinely supports environmental sustainability.
- ...is sincere.
- ...is authentic.

Construct Global Indicator

- ...is trustworthy.

Appendix E: Adapted Reputation Quotient Scale

Source: Franco, A., Sarkar, M., Agarwal R., & Echambadi, R. (2009). Swift and smart: The moderating effects of technological capabilities on the market pioneering-firm survival relationship. *Management Science*, 55(11), 1842–186.

Emotional Appeal

1. I have a good feeling about this certifier.
2. I admire and respect this certifier.
3. I trust this certifier.

Products and Services

4. Stands behind its offerings.
5. Develops innovative certifications.
6. Offers high quality certifications.

Vision and Leadership

7. Has excellent leadership.
8. Has a clear vision for its future.
9. Recognizes and takes advantage of market opportunities.

Workplace Environment

10. Is well-managed.
11. Looks like a good organization to work for.
12. Looks like an organization that would have good employees.

Social and Environmental Responsibility

13. Supports good causes.
14. Is an environmentally responsible company.
15. Maintains high standards in the way it treats people.

Financial Performance

1. Has a strong record of profitability.
2. Tends to outperform its competitors.
3. Looks like an organization with strong prospects for future growth.

Appendix F: Pretest Correlation Matrix

Pretest Correlation Matrix		...is trustworthy.	a.	b.	c.	d.	e.	f.	g.	h.	i.	j.	k.
a.	...is authentic.	.822											
b.	...is sincere.	.809	.876										
c.	...genuinely supports environmental sustainability.	.713	.776	.775									
d.	...can be trusted to have environmental sustainability's best interests at heart.	.785	.796	.781	.880								
e.	...truly cares about environmental sustainability.	.710	.779	.750	.910	.880							
f.	...aims to set standards that support environmental sustainability.	.629	.693	.676	.818	.762	.825						
g.	...prioritizes environmental sustainability.	.702	.752	.750	.856	.835	.860	.838					
h.	...shows high integrity.	.883	.813	.800	.682	.746	.684	.599	.691				
i.	...tries hard to be fair in dealings with others.	.786	.786	.764	.729	.748	.735	.682	.724	.778			
j.	...would not knowingly break the law.	.760	.651	.669	.559	.613	.575	.479	.573	.801	.661		
k.	...would do its best to follow regulations.	.840	.775	.764	.688	.740	.714	.637	.685	.865	.786	.778	
l.	...holds all potential certification holders to the same standards.	.687	.720	.689	.715	.740	.721	.715	.753	.640	.743	.514	.670
m.	...enforces standards consistently.	.713	.734	.742	.734	.755	.743	.750	.755	.697	.747	.587	.705
n.	...is able to achieve its goals.	.342	.441	.401	.423	.405	.421	.415	.401	.336	.392	.230	.322
o.	...has the expertise needed to achieve its goals.	.376	.470	.433	.464	.456	.459	.480	.449	.378	.434	.237	.341
p.	...has the expertise needed to set appropriate standards.	.446	.555	.532	.554	.545	.575	.577	.538	.439	.524	.324	.425
q.	...has the resources it needs to check if products meet its standards.	.275	.359	.334	.355	.359	.366	.374	.315	.290	.337	.145	.250
r.	...has the information it needs to achieve its stated mission and goals.	.421	.527	.467	.520	.498	.521	.478	.450	.419	.468	.304	.425
s.	...has the resources it needs to set proper certification standards.	.217	.305	.263	.292	.318	.335	.340	.263	.208	.304	.146	.214
t.	...is competent.	.595	.667	.642	.612	.619	.619	.581	.609	.602	.589	.494	.550
u.	...knows how to enforce standards.	.522	.629	.629	.589	.600	.614	.593	.613	.542	.604	.450	.551
v.	...shows high ability.	.562	.678	.620	.601	.628	.624	.628	.599	.586	.611	.443	.539

<i>Pretest Correlation Matrix</i>										
	l.	m.	n.	o.	p.	q.	r.	s.	t.	u.
m. ...enforces standards consistently.	.802									
n. ...is able to achieve its goals.	.374	.406								
o. ...has the expertise needed to achieve its goals.	.405	.438	.660							
p. ...has the expertise needed to set appropriate standards.	.533	.538	.616	.767						
q. ...has the resources it needs to check if products meet its standards.	.318	.341	.666	.687	.699					
r. ...has the information it needs to achieve its stated mission and goals.	.427	.436	.615	.688	.722	.605				
s. ...has the resources it needs to set proper certification standards.	.262	.303	.630	.623	.636	.775	.612			
t. ...is competent.	.588	.632	.604	.624	.651	.522	.564	.518		
u. ...knows how to enforce standards.	.590	.623	.566	.651	.726	.563	.670	.545	.642	
v. ...shows high ability.	.597	.622	.615	.655	.688	.542	.623	.527	.794	.689

Appendix G: Final Test of the Certifier Trustworthiness Scale (CTS) Items

Prompt: “Please rate your agreement with the following statements **about Earth Friendly Goods.**”

Scale: 1 – 7 where 1 = Strongly Disagree and 7 = Strongly Agree

Perceived Ability

- ...is able to achieve its goals.
- ...has the expertise needed to achieve its goals.
- ...has the expertise needed to set appropriate standards.
- ...has the resources it needs to check if products meet its standards.
- ...is competent.
- ...knows how to enforce standards.
- ...shows high ability.

Perceived Integrity

- ...would not knowingly break the law.
- ...holds all potential certification holders to the same standards.
- ...enforces standards consistently.

Perceived Authenticity

- ...can be trusted to have environmental sustainability's best interests at heart.
- ...truly cares about environmental sustainability.
- ...aims to set standards that support environmental sustainability.
- ...prioritizes environmental sustainability.
- ...genuinely supports environmental sustainability.

Construct Global Indicator

- ...is trustworthy.

Appendix H: Environmental Concern Scale

Prompt: “Please rate your agreement with the following statements **about Earth Friendly Goods.**”

Scale: 1 – 7 where 1 = Strongly Disagree and 7 = Strongly Agree

Adapted from: Steel, B. S. (1996). Thinking globally and acting locally?: environmental attitudes, behaviour and activism. *Journal of environmental management*, 47(1), 27-36.

1. The balance of nature is very delicate and easily upset by human activities.
2. The earth is like a spaceship with only limited room and resources.
3. Plants and animals do not exist primarily for human use.
4. Modifying the environment for human use seldom causes serious problems.

Appendix I: Preliminary Certifier Trustworthiness Scale (CTS)

Prompt: “Please rate your agreement with the following statements **about [certifier name, e.g. Earth Friendly Goods].**”

Scale: 1 – 7 where 1 = Strongly Disagree and 7 = Strongly Agree

Perceived Ability

1. ...is able to achieve its goals.
2. ...has the expertise needed to achieve its goals.
3. ...has the expertise needed to set appropriate standards.
4. ...has the resources it needs to check if products meet its standards.
5. ...is competent.
6. ...knows how to enforce standards.
7. ...shows high ability.

Perceived Integrity

8. ...would not knowingly break the law.
9. ...holds all potential certification holders to the same standards.
10. ...enforces standards consistently.

Perceived Authenticity

11. ...can be trusted to have [cause(s) here, e.g., environmental sustainability’s] best interests at heart.
12. ...truly cares about [cause(s) here].
13. ...aims to set standards that support [cause(s) here].
14. ...prioritizes [cause(s) here].
15. ...genuinely supports [cause(s) here].

Appendix J: Study 1 – Demographic Questions**Age:**

Question: How old are you in years? (free response)

Gender:

Question: How do you describe yourself?

Response options:

- Female
- Male
- Non-binary / third gender
- Prefer to self-describe (free response)

Country of residence:

Question: Do you currently reside in the United States?

Response options:

- Yes
- No
- Other, please specify. (free response)

Note: Note reported in Study 1 results as all participants reported that they resided in the United State.

Educational attainment:

Question: What is the highest level of school you have completed or the highest degree you have received?

Response options:

- Less than high school degree
- High school degree or equivalent (e.g., GED)
- Some college but no degree
- Associate degree
- Bachelor degree
- Graduate degree

Employment status:

Question: Which of the following categories best describes your employment status?

Response options:

- Employed, working 1-39 hours per week
- Employed, working 40 or more hours per week
- Not employed, looking for work
- Not employed, NOT looking for work
- Retired
- Homemaker
- Student
- Disabled, not able to work

Paper 2: Sustainability certifier diversification: Effects on perceived trustworthiness and performance

As proposed and tested in Paper 1, perceived authenticity, ability, and integrity all contribute to the overall trustworthiness of a certifier. Furthermore, a certifier's perceived trustworthiness can significantly affect whether interested parties will use them. A logical next question is: What factors increase or decrease a certifier's perceived authenticity, ability, and integrity? In this paper, drawing on findings from organizational ecology, I theorize and find support for the idea that certifiers pursuing unrelated diversification are perceived to be less trustworthy than their specialist counterparts by their target audience – producer organizations whose offerings they seek to certify. Further, positive trustworthiness perceptions of certifiers drive demand and ultimately increase certifiers' revenue generated from certification-related activities.

In this paper, I explicitly focus on sustainability certifiers for two reasons. First, sustainability issues are a prevalent concern across industries, and certification is a predominant tool that firms use to communicate their work on environmental, social, and governance issues (McKinsey & NielsenIQ, 2023; Nielsen, 2023). Second, focusing on a specific certification domain increases the internal validity of my observational data. In the following sections, I provide more details on the proposed model related to sustainability certifiers and a critical audience group – the organizations they seek to certify- and review the empirical approach in progress to test the hypotheses stemming from the model.

Theoretical Background

Certifier Diversification and Performance

A diversified certifier certifies offerings and entities, such as organizations, operating in a wide array of market categories, whereas a specialist certifier focuses on relatively few categories (Hsu et al., 2009; Negro et al., 2010). Market categories delineate the space on a given dimension or dimensions to group like with like. They “provide shared frameworks for audiences to compare and evaluate competing organizations and their product offerings” (Porac et al., 1995; Zuckerman, 2000, as cited in Jensen & Kim, 2013). For example, one relevant dimension for certifiers is the offerings they certify. Along this dimension are the conceptual categories of products, services, organizations, and people; a certifier that certifies only products is more specialized than one that certifies both products and organizations. The relevant dimensions depend on the context. As such, I will describe the niche dimensions within the context of each study in the empirical section below.

Certifier diversification is likely to have direct effects on certifier performance. Specialist certifiers may underperform compared to diversified certifiers because they are less visible and less recognizable to audiences. Such logic follows the recognition heuristic. However, empirical findings related to the recognition heuristic indicate that decision-makers tend to rely on something other than name recognition when they can access more information (e.g., Pohl, 2006).

Many factors predict the extent of information search, including the perceived benefits, costs, and ability of the decision-maker to search (Schmidt & Spreng, 1996). Because achieving certification often requires substantial investment from the certification-seeking organization, an information search is likely to occur before beginning the certification process. Furthermore, information search is partially involuntary when pursuing certification. At a minimum, organizations must learn if they qualify for the certification and how to obtain it. Thus, a baseline

hypothesis that generalist certifiers are more likely to survive and grow because they are more visible to decision-makers is unlikely.

However, specialist certifiers can focus their resources and gain expertise in their chosen domain (Hsu et al., 2009). Greater expertise can translate into greater efficiency, quality, and speed (Negro et al., 2010). Additionally, specialist certifiers are better positioned than their generalist counterparts to tailor their offerings to appeal to their focal audiences (Hannan et al., 2003; Hsu, 2006; Hsu et al., 2009). Indeed, specialization has been found to increase performance in a wide range of industries and market sectors (e.g., Skinner, 1979; Leung & Sharkey, 2014; Vera et al., 2018; Zuckerman et al., 2003). For example, Hsu (2006) found that feature films spanning multiple genres attracted larger audiences of professional critics, but the critical appeal of such films was comparatively lower.

Not all forms of diversification are likely to be detrimental for a certifier. Specifically, it is critical to differentiate between *related* and *unrelated* diversification. Prior research shows that firms engaged in related diversification tend to outperform firms engaged in unrelated diversification (Bettis, 1981; Palich et al., 2000). Researchers following the resource-based view of the firm contend that economies of scope benefit firms pursuing related diversification but not unrelated diversification (Wan et al., 2011). Further, a firm operates in many distinct businesses, it becomes more difficult to manage (Grant et al., 1988; Jones & Hill, 1988). It is also critical to be methodical in distinguishing between related and unrelated diversification because two business areas classified into different categories may practically be highly related; for example, they may provide complementary offerings or vertically integrated (Villalonga, 2004). Based upon prior research on specialist versus generalist firms and the relative benefits of related versus unrelated diversification, the first hypothesis is as follows:

Hypothesis 1: *Certifiers pursuing a specialization or related diversification strategy will outperform certifiers pursuing an unrelated diversification strategy.*

Certifier Diversification and Perceived Trustworthiness

Certifier diversification is likely to predict perceived certifier trustworthiness negatively, whether evaluators make trustworthiness judgments quickly or after more thoughtful consideration. *Certifier trustworthiness* is “the trustor’s degree of confidence that the certifier can be counted upon to develop and administer quality standards related to its stated mission and goals (see Paper 1). Both effortful and automatic processing by evaluators can compromise diversified certifiers’ perceived trustworthiness. A diversified certifier operates across multiple business categories or cause areas and has a more difficult-to-interpret business model than a focused certifier. Because of this, a diversified can be said to have high ambiguity (Hannan et al., 2019). Prior research shows that an object’s ambiguity negatively affects the fluency of perception and the ease with which perceivers interpret the object. In turn, lower fluency decreases the appeal of the object (Hannan et al., 2019; Reber et al., 2004; Winkielman, 2006).

These relationships are explained by the “processing fluency account,” which holds that non-ambiguous objects are preferred to their more ambiguous counterparts because they are processed more quickly or easily (Jakesch et al., 2013; Lee & Aaker, 2004; Wänke et al., 1997). Thus, unrelatedly diversified certifiers should be perceived as less generally appealing than their specialized counterparts. Such a negative valence surrounding an unrelatedly diversified certifier should negatively influence audiences’ perceptions of the certifier’s overall trustworthiness and the trustworthiness dimensions of ability, integrity, and authenticity. This reasoning is supported by research on brands, which shows that low brand liking is negatively associated with brand trust (Almahdi et al., 2022). Additionally, research in experimental economics found that people who are more ambiguity-averse are less likely to trust (Li et al., 2019).

Mayer and colleagues (1995) conceptualized trustworthiness at the interpersonal level as a tri-dimensional construct with the dimensions of ability, integrity, and benevolence. Certifier trustworthiness can be similarly conceptualized as a tri-dimensional construct with the dimensions of ability, integrity, and authenticity. Perceived *ability* is the trustor's perception that the certifier has the resources and expertise needed to achieve the certifier's self-proclaimed mission and goals. Perceived *integrity* is the trustor's belief that the certifier is fair and just. Perceived *authenticity* is the trustor's belief that the certifier is sincere in its support for the purpose behind the certification (see Paper 1). Unrelated diversification is theorized to negatively predict each trustworthiness dimension through reduced audience member appeal directly and due to increased organizational ambiguity.

First, unrelated diversification should negatively predict the perceived ability of a certifier. When entities take a generalist approach by spanning market categories, they signal to audience members that they lack expertise, even if that is inaccurate (Hsu et al., 2009; Negro et al., 2010). For example, Zuckerman and colleagues (2003) found support for their hypothesis that actors who were typecast to a given category are more likely to find work in their typecast category than non-typecast actors. Thus, typecasting is considered a heuristic for casting decisions, limiting consideration to actors with proven ability to perform the role. Research in multiple other disciplines has found positive relationships between specialization and performance expectations. For example, survey respondents expected high school athletes more specialized in sporting activities to perform better (Hill & Simmons, 1989).

Unrelated certifier diversification should also decrease perceived integrity. Perceived integrity is the trustor's belief that the certifier follows principles that the trustor finds acceptable and acts consistently, holding all potential certification holders to the same standards and

uniformly enforcing standards (Mayer et al., 1995). Certifiers offering many or less specific standards have less focused attention and less to lose if they are found liable for not abiding by best practices for a given certification area. It may also be assumed that diversified certifiers are less easily monitored and evaluated for maintaining ethical practices due to relatively high organizational complexity.

Finally, unrelated diversification should decrease perceived authenticity. Perceived authenticity is closely linked to engagement – one-way and two-way communication with the target audience. Engagement encompasses a broad set of actions, including learning about the nuances of specific audiences, customizing offerings to appeal to those audiences, and creating a relevant identity (Hannan et al., 2003). Prior research has found that engagement within a category tends to correlate with category membership (Hsu et al., 2009). Regular engagement with a category audience should enhance perceived authenticity, particularly moral authenticity (Hsu et al., 2009; Verhaal et al., 2017). Moral authenticity is where decisions and actions match their sincere intentions (Carroll & Wheaton, 2009). Therefore, certifier diversification should negatively influence perceived certifier trustworthiness and its sub-dimensions of perceived ability, integrity, and authenticity. This reasoning leads to the following hypotheses:

Hypothesis 2a: *Unrelatedly diversified certifiers receive lower evaluations of trustworthiness, specifically:*

Hypothesis 2b: *Unrelatedly diversified certifiers receive lower evaluations on the trustworthiness dimension of perceived **ability** than specialized or relatedly diversified certifiers.*

Hypothesis 2c: *Unrelatedly diversified certifiers receive lower evaluations on the trustworthiness dimension of perceived **integrity** than specialized or relatedly diversified certifiers.*

Hypothesis 2d: *Unrelatedly diversified certifiers receive lower evaluations on the trustworthiness dimension of perceived **authenticity** than specialized or relatedly diversified certifiers.*

Perceived Trustworthiness as a Mediator

Theoretical and empirical evidence points to perceived trustworthiness as an integral mediator in the positive relationship between the degree of certifier diversification and their performance. Certifier diversification should directly affect the perceived trustworthiness dimensions of ability, integrity, and authenticity as described above, and in turn, attributions of trustworthiness should directly relate to certifier performance.

Prior management research supports the hypothesized direct relationship between trustworthiness and performance. For instance, buyers' trust in their suppliers increases suppliers' performance (Poppo et al., 2016), there is a direct link between interorganizational trust and performance (Gulati & Nickerson, 2008; Zaheer et al., 1998), and employees' trust in management increases business unit performance (Davis et al., 2000). Further, trust increases cooperation within organizations (Gulati & Westphal, 1999; Williams, 2001) and between different stakeholder groups (Jensen, 2003; Uzzi, 1997). The marketing literature provides evidence for trust enhancing consumer satisfaction (Doney & Cannon, 1997), brand loyalty and market share (Chaudhuri & Holbrook, 2001), and consumer commitment to a brand (Delgado-Ballester, & Munuera-Aleman, 2001). Additionally, consumers assign higher ratings to restaurants they perceive as authentic, one of the three dimensions of certifier trustworthiness (Paper 1; Kovacs et al., 2014).

Therefore, perceived trustworthiness should also be a positive predictor of certifier performance. This has practical implications as producers considering certification are more likely to choose certifiers they perceive as trustworthy and, importantly, believe consumers will also perceive as trustworthy. Recent evidence also suggests that interpersonal trust mediates the relationship between entrepreneur specialization and the funds they raise at their ventures' initial

public offering (IPO; Souitaris et al., 2023). These findings lead to the formulation of the final four hypotheses:

***Hypothesis 3a:** Certifier trustworthiness mediates the negative relationship between unrelated certifier diversification and performance.*

***Hypothesis 3b:** The certifier trustworthiness dimension of perceived **ability** mediates the negative relationship between unrelated certifier diversification and performance.*

***Hypothesis 3c:** The certifier trustworthiness dimension of perceived **integrity** mediates the negative relationship between unrelated certifier diversification and performance.*

***Hypothesis 3d:** The certifier trustworthiness dimension of perceived **authenticity** mediates the negative relationship between unrelated certifier diversification and performance.*

Context of Sustainability Certifiers

The present research focuses on the context of sustainability certifiers, organizations that sponsor “ecolabels.” The Environmental Protection Agency (EPA) defines ecolabels as “marks placed on product packaging or in e-catalogs that can help consumers and institutional purchasers quickly and easily identify those products that meet specific environmental performance criteria and are therefore deemed “environmentally preferable” (United States, 2021). Therefore, sustainability labels are marks on product packaging or in e-catalogs that can help consumers and institutional purchasers quickly and easily identify those products that meet specific environmental, social, and economic performance criteria and are therefore deemed “socially preferable.”

The market of sustainability certifications spans product industries, supply chain stages, and geographic boundaries. Despite differences across these dimensions, sustainability certifications share similar purposes: to promote environmental and social well-being and fill regulatory gaps. Some sustainability certifiers conduct operations outside the scope of certification. For example, the Rainforest Alliance also engages in lobbying (Rainforest Alliance,

2022; OpenSecrets.org; 2022). Sustainability certifiers also span sectors; some operate as nonprofit organizations, others are for-profit, and some are government-operated.

Data and Methods

I conducted two primary studies to test the hypotheses within the context of sustainability certifiers and an additional supplementary study to classify the certifiers' diversification approaches. The first study (Study 1) is an online vignette experiment where I manipulated the degree of certifier specialization. The supplementary study (Study 2a) for the final observational study is an online survey where participants identified the market categories that they perceived as related to each other, the results of which are used to identify unrelated diversification efforts in the observational study (Study 2b). Study 2b uses a novel data set of nonprofit sustainability certifiers with headquarters in the United States. Below is a description of each study.

Study 1 – Experiment, Mechanism of Perceived Trustworthiness

Study 1 tests hypotheses 2a thru 3d. Hypotheses 2a-d hold that diversified certifiers will be perceived as less trustworthy overall and less able, integrous, and authentic. Hypotheses 3a-d hold that perceived trustworthiness and its compositional dimensions of perceived ability, integrity, and authenticity mediate the negative relationship between certifier diversification and performance. The experiment uses a vignette design with three conditions varying the extent of certifier diversification.

Sample

The sample included an online panel of respondents sourced from Prolific. Prolific has been found to have high external validity for questions about perceptions and beliefs (Tang et al., 2022). The sample initially included 520 participants aged 18 years or older, with at least a high school degree or its equivalent, and spoke English fluently; participants were located in the

United States at the time of the experiment, and the sample was gender-balanced. I removed 84 observations from the sample due to participants failing the manipulation check at the end of the study. I removed 19 observations that had at least one attention check failure. I removed one observation where the participant was identified as a potential bot by the survey platform Qualtrics. Finally, I removed 14 observations that had missing responses to items within the measure for the mediator variable of perceived certifier trustworthiness. The final sample included 402 participants, about seventy-seven percent of the initial sample. Most participants were female (51.74%), and the mean age was 39.96 years (S.D. 14.28). The observation removal rate is consistent with empirical evidence on data quality using a Prolific sample (Douglas et al., 2023). Welch's two-sided t-tests indicated that there were no significant differences in age or gender between the original sample and the final sample.

Experimental Design

The experiment used a between-subjects design; participants were randomly assigned to one of three conditions. The only difference between conditions was the certifier's degree of product specialization (diversification). The three conditions were (1) specialization, (2) related diversification, and (3) unrelated diversification. The experiment employed a vignette scenario, where participants were asked to imagine that they were the owner of a chocolate store making a decision about a supplier certified by the fictional certifier "Earth Friendly Goods." The experiment took an average of about nine minutes to complete. Before beginning the experiment, participants were assured of anonymity and asked to consent. For each condition, participants received the same instructions upon providing consent. The instructions read:

The goal of this survey is to better understand how the public reviews and evaluates certifying organizations.

You will first read important background information for this study and then be asked two questions about what you read.

Then, you will be asked to imagine yourself in a scenario related to the fictional organization "Earth Friendly Goods" and be asked to answer several questions about the organization.

Then, participants were provided background information on what was meant by a “certifying organization” and by “environmental sustainability” for the study. They were asked to respond to questions about the meaning of each term, and if they answered incorrectly, they were shown the correct response.

The following prompt explained the vignette scenario and was consistent across assigned conditions. The prompt read:

Imagine you're the proud owner of a cozy local chocolate shop. You have a wide range of chocolates on display, from tasty truffles to lush chocolate bars and mouth-watering fudge. Your customers have high expectations when it comes to both taste and presentation, and they also care about the environment. That's why you're choosing a chocolate supplier that is certified sustainable.

The following and last vignette prompt did vary by condition and described the certifier “Earth Friendly Goods.” The specialization condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies only chocolate and the raw foods for chocolate**, such as cocoa beans.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

The related diversification condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies many types of food products**, such as chocolate, coffee, vegetables, fruits, and meats.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

The unrelated diversification condition read:

You're thinking about choosing a chocolate supplier certified by **Earth Friendly Goods**. Earth Friendly Goods is a nonprofit organization founded in 2013 and based in Seattle, Washington. Here are some more facts about Earth Friendly Goods:

- **Certifies all types of products**, such as personal electronic devices, paper products, toys, and foods.
- Sets and maintains environmental standards for the products it certifies.
- Works with local partners, such as on-site production auditors, to verify that products meet their standards.
- The value of Earth Friendly Good's profits is similar to other organizations of the same size.

Next, participants were asked to respond to items to measure the mediating variable (perceived certifier trustworthiness), dependent (purchase intention), control (warmth and competence, political ideology, environmental concern, familiarity with certifiers), demographic variables (gender, age, highest level of education attained, employment status, and geographic location), and a manipulation check to assess their understanding of the condition to which they were randomly assigned. Finally, participants were thanked, provided contact information to raise questions or concerns, and prompted to exit the survey.

Measures

Mediating variable - perceived certifier trustworthiness. The Certifier Trustworthiness Scale (CTS) developed and validated in Paper 1 assessed respondents'

perceptions of the trustworthiness of the certifier. The CTS contains three dimensions: perceived ability, perceived integrity, and perceived authenticity; it has 15 items (see Appendix J).

Dependent variable – purchase intention. Participants rated four items adapted from Moon and colleagues' (2008) purchase intention measure to assess the likelihood of purchasing a product certified by Earth Friendly Goods. The three items were: (1) I would be willing to pay more for a product certified by Earth Friendly Goods; (2) I would recommend products certified by Earth Friendly Goods to a friend; and (3) I would consider Early Friends Goods certification as part of my purchase decision. Participants rated all items on a 7-point Likert scale ranging from 1 (*very unlikely*) to 7 (*very likely*).

Control variables. Participants responded to nine measures of variables potentially driving their purchase intention.

Warmth and competence of certifier. Previous research indicates that audiences determine their engagement with an entity based on part on perceptions of the entity's warmth and competence (e.g. Kervyn et al., 2012). I measured the perceived warmth and competence of the certifier following the procedure used by Aaker, Garbinsky, and Vohs (2012). Participants used a 5-point scale (1= *Not at all*, 5 = *Extremely*) to respond to the question of, "To what extent do the following words describe Earth Friendly Goods?" for each of the words "competent," "capable," "warm," and "friendly."

Environmental concern. Participants with greater concern for the environment, in general, should be more likely to perceive a sustainability certifier such as Earth Friendly Goods as trustworthy due to greater identification with the organization (e.g., Woolley & Fishbach, 2017). Therefore, they should also be more likely to favor products verified by a sustainability

certifier than participants with lower environmental concern. I measured environmental concern using four items, and participants responded on a 7-point agreement scale (see Appendix K).

Political ideology. A liberal political ideology is positively associated with support for environmental causes (e.g., Dunlap & McCright, 2011). Therefore, a liberal political ideology is likely to predict the dependent variable of purchase intention positively. To measure participants' political ideology, I asked, "When it comes to politics, would you describe yourself as liberal, conservative, or neither liberal nor conservative?" with the seven response options ranging from 1 = Extremely Liberal to 7 = Extremely Conservative.

Familiarity with certifiers. Familiarity with a specific entity has been shown to increase trust in some contexts (e.g., Ha & Perks, 2005). Thus, familiarity with a particular organizational form, such as a certifier, may similarly lead to increased perceptions of trustworthiness. I evaluated participants' familiarity with certifiers using a 3-item measure adapted from Kent & Allen (1994). The items included "I am familiar with certifying organizations.," "I am experienced with certifying organizations.," and "I am knowledgeable about certifying organizations." Participants responded on an agreement scale ranging from 1 = Strongly Disagree to 7 = Strongly Agree.

Demographics. Additional control variables include participants' age, gender, highest level of education attained, and employment status. Age was measured as a continuous variable, while the other control variables were measured based on responses to multiple-choice questions. See Appendix A for participants' demographic characteristics by condition. No significant differences existed for any of the demographic characteristics measured between conditions.

Results

The results of Study 2 support hypotheses 2a-c. Participants in the specialization condition gave the highest perceived trustworthiness ratings of the three conditions for the certifier, Earth Friendly Goods, overall (2a; $p < .001$), and the highest rating for the individual dimensions of perceived ability (2b; $p < .001$), perceived integrity (2c; $p < .01$), and perceived authenticity (2d; $p < .01$; see Tables 12a-3b in Appendices L-O). Additionally, familiarity with certifiers, environmental concern, and the perceived warmth and competence of the certifier all positively moderated the effects of specialization on perceived trustworthiness (see Table 12b in Appendix M and 13b in Appendix O).

Overall, the results support Hypothesis 3a, which holds that certifier trustworthiness mediates the positive relationship between certifier specialization and performance. The indirect effect of specialization on performance via certifier trustworthiness was insignificant when comparing all three conditions independently. Upon further inspection, the pattern of results was similar for both the related and unrelated diversification conditions, indicating that the observations for the two conditions could be combined. The lack of differences comparing all three conditions is due to the similarity between the related and unrelated conditions. Welch's two-sided t-tests indicated that there were no significant differences in perceived trustworthiness or purchase intention between the two conditions.

Therefore, I tested the hypotheses while combining the two diversification conditions into one "diversification" group. There was a significant indirect effect of specialization on performance via certifier trustworthiness when comparing the specialization condition to the aggregated diversification group and the specialization condition to solely the unrelated diversification condition. Furthermore, the individual dimensions of perceived ability, perceived integrity, and perceived authenticity each fully mediate the positive relationship between certifier

specialization and purchase intention when comparing the specialization condition to the single diversification group, supporting hypotheses 3b-d (see Tables 16c-f in Appendix S).

Study 1 Discussion

This experimental study breaks new ground by providing initial evidence that certifier trustworthiness and its dimensions of perceived ability, integrity, and authenticity mediate the positive relationship between specialization and certifier performance. Importantly, the study also reveals a positive correlation between perceived trustworthiness and the behavioral intention measure of purchase intention, further validating the meaningfulness of the findings.

Study 2a – Survey, Identifying Unrelated Diversification

This study aims to differentiate between related and unrelated diversification across product market categories. The product market categories used are from Intengine.com, one of the primary data sources on certification schemes used in Study 2b. The findings from this study are used in Study 2b to test hypotheses 3a-d, predicting significant performance differences based on the degree of diversification. Participants completed a card-sort task by grouping business (product market) categories based on their perceived similarity. The results support the distinction of the business categories into four groups.

Sample

The sample included an online panel of respondents sourced from Prolific. The sample initially included 52 participants aged 18 years or older, with at least a high school degree or its equivalent, and spoke English fluently; participants were located in the United States at the time of the experiment, and the sample was gender-balanced. I removed four observations from the sample due to participants failing at least one attention check. The final sample included 48 participants, about ninety-two percent of the initial sample. Most (25) participants were female,

and the mean age was 38.76 years (S.D. 13.68). Welch's two-sided t-tests indicated no significant differences in age, and a chi-square test indicated no significant differences in gender between the original and final samples.

Design

The study was an online version of a card sort. The survey took an average of about seven minutes to complete. Before beginning the experiment, participants were assured of anonymity and asked to consent. After providing consent, participants who moved on to the next page were asked to complete an "item-sorting task." The instructions were as follows:

Instructions:

- Click, drag, and drop "Items" (business categories) on the left into "Group" boxes on the right.
- **Group similar items together using your best judgment.**
- Use the "Miscellaneous" box only for items that appear entirely unrelated to other items.
- There are no right or wrong answers.

On the left-side of participants' screen was a list of 31 items (see Appendix U). Twenty-nine items were product category names; some were simplified from the name used on Intengine.com to ease readability and interpretation. The remaining two items were for attention checks ("put in Group 4" and "move to Group 1"). Notably, I excluded two of the product categories used by Intengine.com at the time of data collection ("Industry" and "Specialty") because they are vague and would likely not be meaningfully sorted by participants. On the right side of the participants' screen were eleven boxes for participants to group the items. The first box was labeled "Miscellaneous," and the other boxes were labeled "Group 1" through "Group 10."

After completing the item-sorting task, participants were asked to answer three questions about their experience with the task. The first question was multiple choice, "How difficult did you find the item-sorting task?" with response options on a 5-point scale ranging from 1 (*very*

difficult) to 5 (*very easy*). The second and third feedback questions had open-ended responses. The second question was, “Do you have any recommendations on how to improve the item-sorting task?” and the third question was, “Please provide any other feedback or comments for the researchers about your experience.” Next, participants were asked to respond to several demographic questions optionally. The demographic questions asked about their gender, age, highest level of education, employment status, and geographic location. Finally, participants were thanked, provided contact information to raise questions or concerns, and prompted to exit the survey.

Analysis and Results

I used a combination of hierarchical and k-means clustering, two of the most common unsupervised machine-learning clustering algorithms, to assess the best-fitting categorization from the survey results. I used a co-occurrence matrix as the basis for each clustering analysis. A co-occurrence matrix identifies shows how often pairs of data elements appear together. I used hierarchical agglomerative clustering in Python via the *scipy* library for an initial assessment and visualization of emerging clusters. In agglomerative clustering, each data point is first considered an individual cluster. Then, as the algorithm iterates, similar clusters merge based on the Euclidean distance between the two clusters (Patlolla, 2018). The hierarchical clustering results are visually interpretable using a dendrogram (see Appendix AD).

The hierarchical clustering results show a clear distinction between two main clusters: one composed of agricultural and environmental categories and one seemingly a catch-all of everything else. Because of inherent limitations with hierarchical clustering with noisy data and path dependency in cluster identification, I used k-means clustering to identify possible sub-categories within each main category.

K-means clustering is another unsupervised machine-learning technique to partition data into a pre-specified number of groups (Rousseeuw & Kaufman, 2009). The k-means algorithm sorts each data point into a cluster using the distance between points as a similarity measure (Bholowalia & Kumar, 2014). This technique requires the pre-identification of the number of clusters. I used the Elbow and silhouette methods to identify the optimal number of clusters. The premise of the Elbow method is to minimize the number of clusters while maximizing the percentage of variance explained. The first cluster will explain the most variance, the second cluster will explain the second-most variance, and so forth. The number of clusters “k” should be chosen where the marginal gain from adding another cluster drops dramatically (Bholowalia & Kumar, 2014). This is typically assessed visually using a line graph.

For the first primary group identified using hierarchical clustering, the elbow method did not show any clear k-value (see Figure 1 in Appendix AF). Therefore, I also used the silhouette method initially proposed by Rousseeuw (1987). The silhouette method “defines for each object the measure of how this object is similar to other objects from the same cluster (cohesion, compactness) in comparison with objects of other clusters (separation)” (Dudek, 2020, p. 19). Silhouette values can range from -1 to 1, with higher values representing a greater fit to the assigned cluster. The results of the silhouette analysis for the first primary group indicated that (up to) two clusters best fit the data. Paired with the initial hierarchical clustering results and my knowledge of the product market categories, I kept the first primary group as a single cluster.

For the second primary group identified using hierarchical clustering, the elbow method indicated two to three clusters best fit the data (see Figure 3 of Appendix AF). I used the silhouette method initially proposed by Rousseeuw (1987). The results of the silhouette analysis

for the second primary group indicated three or four clusters best fit the data (see Figure 4 of Appendix AF). Therefore, I used $k = 3$ for the k-means clustering analysis of the second primary group. The results of the k-means clustering are shown in Appendix AE.

Study 2a Discussion

The results of this study lay the groundwork for evaluating the advantages of related versus unrelated diversification in Study 2b. Including lay participants enhances this assessment's external validity, aligning with decision-maker profiles in producing organizations. The broad delineation of business categories offers a conservative testing ground for hypotheses concerning the benefits of related diversification, as explored in Study 2b.

Study 2b – Observational, Ecolabels

Study 2b tests Hypothesis 1, which predicts that certifier specialization positively influences certifier performance. The study uses a unique observational data set of sustainability certifiers (i.e., ecolabel-providing organizations) based in the United States (U.S.), providing strong ecological validity.

Sample

The data set originated from two sources. The first source was the EcoLabel Index, a global directory of ecolabels. The EcoLabel Index is operated by the for-profit, certified B Corporation, Big Room Inc., based in Vancouver, Canada (Ecolabel Index, 2022), and is used in peer-reviewed research uses the Ecolabel Index (e.g., Couckuyt & Amy, 2021; Darnall et al., 2018). The Ecolabel Index was first published online in 2007. It is unclear when the Ecolabel Index was last updated, but it did not appear to be up-to-date as of the commencement of data collection in 2021. Therefore, the EcoLabel Index data was supplemented by data on certifying organizations found in the online directory *Intengine.com* to create the most complete database

of sustainability certifiers possible. *Intengine.com* is promoted as having “the world’s most comprehensive and relevant database of responsible businesses” and allows users to search for sustainability-related certifications (Intengine.com, 2022). Intengine.com has entries for individual sustainability certifications linked to the organizations that provide them.

The research focuses on certifiers based in the United States for ease of comparability of the financial dependent variables. Likewise, the final sample is limited to nonprofit certifiers because their annual financial information is publicly available in most cases from their IRS filings (limited exceptions include religious organizations not required to disclose their financial information). The final data set consists of 1289 observations representing 87 organizations from 1989 to 2021. Please see Table 18 in Appendix W for descriptive statistics of the final sample and Figure 2 in Appendix Z for a longitudinal count of the sustainability certifiers included in the data set.

Dependent Variable – Certifier Financial Performance

IRS financial data from the organizations’ Form 990s was obtained from the National Center for Charitable Statistics (NCCS) Data Archive, an online repository maintained by the Urban Institute, a nonprofit research organization (Urban, 2023). All financial data was transformed to a logarithmic scale to correct for skewness and inflation-adjusted using the annual CPI published by the U.S. Bureau of Labor Statistics.

Annual program services revenue. Annual program services revenue (PSR) is the primary dependent variable used in this study. PSR indicates how much of an organization's financial resources come from its mission-related activities. PSR includes, but is not limited to, fees for services or goods provided, registration fees, contractual income from government grants that require services, and royalties. PSR was chosen over total revenue due to its closer

representation of the performance of an organization's certification efforts, as total revenue often includes revenue from unrelated activities such as grants and investments.

Independent Variables – Certifier Specialization

Active product categories. The product categories are those specified by the data source *Intengine.com* (see Appendix U). Each certification offered was classified by the Large Language Model (LLM) Chat GPT 3.1 using the following prompt: "Which of the following product categories are covered by the [insert certification name] certification from [insert certifier name]? Please provide a simple list of the relevant product categories: [list of product categories]." Below are two example responses from Chat GPT 3.1.

Example 1: The EPEAT - Computers and Displays certification from the Green Electronics Council primarily covers the "Computers & Electronics" product category, specifically focusing on computers and displays that meet certain environmental criteria.

Example 2: The "Safe + Sustainable Site Certification" from Earth Advantage Inc primarily covers the "Construction" and "Real Estate" product categories. This certification is focused on recognizing and promoting safe and sustainable practices in the construction and real estate sectors.

The count of active product categories is at the certifier level. The regression analyses include three variations of the active product categories variable. The first is *active product categories*, a continuous measure ranging from 1-31. The second is *two or more active product categories*, a binary variable where zero indicates that the certifier had only one active product category, and a value of one indicates that the certifier had at least two active product categories during the year in question. The third version of this independent variable is *three or more active product categories*, a binary variable where a value of zero indicates that the certifier had two or fewer active product categories, and a value of one indicates that the certifier had three or more active product categories. A majority of the observations had two or less active product categories.

Five-hundred forty-two of the 1,289 total observations (about 42%) had one active product category, and 250 of the 1,289 total observations (about 19%) had two active product categories.

Unrelated product category diversification. Unrelated product category diversification. The survey results described in Study 2a provide the categorical variables for identifying unrelated product category diversification. A certifier is classified as following unrelated product category diversification if they span two or more product category groupings, as shown in Appendix AE. There are four unrelated diversification variables tested. The first, “two primary groups,” is based on the initial clusters identified through hierarchical clustering and includes one group related to agricultural and environmental business categories and another (“other”) that is an assortment of the remaining business categories. The second, “four groups,” holds the first grouping consistent as the “two primary groups” clustering but deconstructs the “other” grouping into three sub-groups, including “Lifestyle,” “Industry Solutions,” and “Travel and Hospitality.” The third unrelated diversification variable (“within first primary group”) considers only the first primary group (of the “two primary groups” clustering), specifically, the two identified sub-clusters within it using k-means clustering as described in Study 2a. Finally, the fourth unrelated diversification variable, “within second primary group,” isolates the second primary grouping (of the “two primary groups” clustering”, the “other” cluster) and the three identified sub-clusters within it.

Sustainability goal diversification. In addition to diversifying across product categories, certifiers may diversify across sustainability causes. I use the United Nations’ Sustainable Development Goals framework to assess the degrees of environmental and social sustainability diversification. The Sustainable Development Goals (SDGs) were adopted by the United Nations General Assembly in 2015 to act as a “blueprint for peace and prosperity for people and planet”

(United Nations, 2023). There are 17 SDGs (see Appendix Y). Each SDG has a set of targets. For example, a target for the “No poverty” goal is to “eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day” by 2030. I classified each of the 169 targets as either “environmental” or “social.” Then, based on the dominant classification of the targets, I classified the SDGs as either environmental or social. Two SDGs have a near-equal mixture of environmental and social targets; therefore, I assigned these targets to both categories. The final SDG, “Partnerships for the goals,” does not represent social or environmental causes and, therefore, was excluded from the analyses.

The social SDG diversification variable has a possible range of 0-11, while the environmental SDG diversification variable has a potential range of 0-7. However, each variable has an observed range of 0-5. Similarly to the active product categories variable, social and environmental sustainability goal diversification variables are at the certifier level. Each certification was assigned a count of SDGs addressed based on responses from ChatGPT 3.1 to the prompt: “Which Sustainable Development Goals from the United Nations are directly addressed by the [certification name] certification from [certifier name]? Clearly indicate which SDG is directly versus indirectly addressed.” I only included SDGs that Chat GPT 3.1 tagged as directly addressed in the counts.

Mediating Variables – Perceived and Objective Certifier Trustworthiness

I used multiple data sources to gain observational insight into the perceived trustworthiness of the sample of certifiers. No single source is ideal for measuring the perceived trustworthiness of the certifier over time. However, three data sources provide preliminary insight into stakeholders’ perceptions of these certifiers and a relatively objective measure of

their trustworthiness. These data sources, their corresponding variables, and their strengths and weaknesses are described below.

Glassdoor ratings. The online platform Glassdoor provides organization ratings for each review by current or former employees. The ratings range from one to five stars, with five stars representing the highest-performing organizations (Ratings from Glassdoor, 2004). The Glassdoor ratings provide insight into internal stakeholders' perceptions of the certifiers. These perceptions may include, but are not limited to, perceptions of the certifier's trustworthiness and the dimensions of ability, integrity, and authenticity.

I calculated three primary variables from the Glassdoor ratings. The first variable, "annual average rating," is the average employee rating for a given year; it is a continuous variable ranging from one to five. The second variable, "annual median rating," is the median employee rating for a given year; it is a factorial variable ranging from one to five. The third variable, "average rating, all years," is the overall average rating (ranging from 1-5) for the organization for all ratings provided, regardless of when the review was provided; as such, there is a single "average rating, all years" value per organization. I also included variables for the annual standard deviation of the employee ratings and the total reviews by year and overall.

In addition to rating their current or former employer, reviewers are asked to describe the pros and cons of their experience in open-response questions. I am reviewing this qualitative data using the software MAXQDA24 to assess the role of trustworthiness and its theorized dimensions in employees' assessments of their employers. Early findings suggest that for at least some employees, the certifier's trustworthiness, not only as an employer but as a sustainability certifier, plays a critical role in employees' ratings.

Charity Navigator ratings. Charity Navigator employees research and rate nonprofit organizations across four domains, called “beacons.” The four beacons are impact and measurement, accountability and finance, leadership and adaptability, and culture and community (How we rate charities, 2024; Rating methodology guide, 2024). As described in Charity Navigator’s methodology guide (2024), the impact and measurement beacon assesses the degree to which the nonprofit makes good use of its resources. The accountability and finance beacon assesses the nonprofit’s financial well-being, accountability, and transparency. The leadership and adaptability beacon assesses the nonprofit’s ability to respond to change and the effectiveness of its leadership. Finally, the culture and community beacon assess the nonprofit’s connection to its stakeholders and internal culture. An overall rating is publicly available for over 225,000 organizations on Charity Navigator’s website (About us, 2024). The rating is provided as a percentage score and a star score, with the highest star rating being four stars. For example, the U.S. Green Building Council is currently listed as a four-star charity with a rating of ninety-six percent. For the Study 2b analysis I used the overall percentage rating for the organization from November 2023. Notably, the Charity Navigator rating is an expert opinion of the certifier’s trustworthiness. The experts employed by Charity Navigator to rate organizations are likely to meaningfully differ in their knowledge of certifiers and the nonprofit sector compared to producers.

Great Nonprofits reviews. Great Nonprofits touts itself as the “leading platform for community-sourced stories about nonprofits” (Great Nonprofits, 2024). The information included for each organization on greatnonprofits.org includes up to its location, a list of causes supported, a descriptive overview of the organization, and “stories” (reviews) from stakeholders, including volunteers, clients served, donors, employees, and other supporters. The diversity of

review sources is beneficial for the present research due to my primary interest in the perspectives of clients served (members of producer organizations). However, the coverage of the certifiers within my sample is limited; only 18 of the 87 organizations in the observational sample had at least one review on [greatnonprofits.org](https://www.greatnonprofits.org) as of November 2023. Therefore, I am pursuing directed content analysis using MAXQDA of the review texts to ascertain the relative importance of trustworthiness, ability, integrity, and authenticity for reviewers. To-date, my analysis indicates that approximately forty-seven percent of the descriptive review content relates to one or more dimensions of certifier trustworthiness. For example, an employee of a client served by Sustainable Travel International wrote that the certifier has “been an invaluable partner to Travel Oregon on numerous fronts by providing us with quality information on the sustainable/green travel market... STI's staff is knowledgeable and professional.”

Control Variables

I control for the observation year and the count of active certifications. The count of active certifications is the sum of the count of individual active certifications for the certifier for each year in the observation period. For example, in 2021, the U.S. Green Building Council had seven active certifications (LEED ID+C, LEED Cities and Communities, LEED Zero, LEED O+M, LEED, LEED Homes, and LEED BD+C). I used various sources to identify the year certifications were first offered as accurately as possible. Whenever possible, information came from the certifier’s website or publications, such as their annual reports. If the certification start year was not available directly from the certifier, the next preferred source was the EcoLabel Index. If not available from either of those preferred sources, I used online news publications to identify the certification’s launch year.

Additionally, I control for the geographic reach. Specifically, I use a control for whether the certifier operates outside North America based on the aggregation of certification-level information obtained from Chat GPT 3.1. Furthermore, as a proxy for the year founded, I control for the year the certifier was first granted tax-exempt status by the United States Internal Revenue Service (IRS; “IRS ruling year”).

Analyses and Results

The final data set consists of 87 certifiers with observations from 1989 through 2021. The data set is an unbalanced panel. Data may be missing for one of two reasons. First, the certifier may have yet to exist during the given year (there was a steady increase in the number of sustainability certifiers for most of the observation period; see Figure 1 in Appendix Z). Second, the certifiers did not always report PSR on their IRS tax form.

Product category spanning. I first conducted fixed-effects panel linear regression analyses with the certifier as the unit of analysis (see Table 19 in Appendix AB). I controlled for observation year in each model due to the study's longitudinal design. Generally, there is a positive relationship between the number of certifications the certifier offers (certification count) and PSR. When including the independent variables of spanning three or more product categories and environmental SDG count together with the control variables, spanning three or more product categories has a significant negative relationship with PSR ($p < .01$).

SDG spanning. The environment and social SDG count variables tend to have a significant positive relationship with PSR (see Table 19 in Appendix AB). This pattern of results suggests that cause-related diversification benefits certifier performance, contradictory to Hypothesis 1 but in line with the alternative argument that diversification across sustainability causes facilitates the certifier's perceived authenticity and overall trustworthiness.

Unrelated diversification. Next, I conducted pooled ordinary least squares regression (OLS) to test for the effects of unrelated diversification (see Tables 21 and 22 in Appendix AB). Pooled OLS assumes that there are no unobservable entity-specific effects. This assumption is not ideal for my data set because there are likely varied certifier-specific characteristics that I do not explicitly capture with control variables. However, pooled OLS is suitable for testing the effects of unrelated diversification, geographic reach, and Glassdoor ratings because they have little to no time variance. The results of these analyses generally support the hypothesis that unrelated diversification negatively affects certifier performance. For example, in Model 16, the “two primary group” variable for unrelated diversification has a significant negative relationship with the outcome variable of PSR (numeric results) while controlling for the observation year, the count of environmental and social SDGs, the geographic reach of the certifier, and IRS ruling year ($p < .001$).

Mediation by certifier appeal. Hypotheses 3a-d predict that the trustworthiness, ability, integrity, and authenticity of a certifier as perceived by evaluators from producer organizations mediate the negative relationship between certifier diversification and certifier performance. I have not identified a measure of perceived certifier trustworthiness that allows me to test this prediction directly. However, the Charity Navigator rating allows for the testing of the perceived trustworthiness of the certifier by an expert audience. Additionally, the Glassdoor ratings allow me to test the appeal of the certifier to present and past employees.

Mediation by Charity Navigator ratings. The Charity Navigator ratings variable is available for 785 of the 1289 observations, with a mean value of 90.19 and a standard deviation of 10.76, with a left-skewed distribution. I conducted mediation analysis using the Lavaan package in R with a maximum likelihood (ML) estimator and NLMINB optimization, which is

well-suited for non-normally distributed variables. In the first mediation model, I used “unrelated diversification, two groups” as the independent variable and PSR as the dependent variable.

The findings indicate that perceived trustworthiness by experts fully mediates the relationship between unrelated diversification and certifier performance. First, there is a significant direct effect of two-group unrelated diversification on Charity Navigator ratings ($p < .001$), suggesting that unrelated diversification negatively influences the perceived trustworthiness of a certifier by experts. Second, there is a significant positive effect of Charity Navigator ratings on PSR ($p < .001$), indicating that perceived trustworthiness by experts positively influences certifier performance. The direct effect of two-group unrelated diversification on PSR was insignificant ($p = .427$); the indirect effect of unrelated diversification on PSR while controlling for the observation year was significant ($p < .001$). In the second Charity Navigator ratings mediation model, I used four-group unrelated diversification as the mediating variable instead of two-group unrelated diversification; the pattern of results was similar (see Tables 24 and 26 in Appendix AC).

Mediation by Glassdoor ratings. The Glassdoor average rating variable is available for only 192 of 1289 observations. Therefore, any findings with the Glassdoor average rating variable are preliminary. The Glassdoor average rating variable has a mean value of 3.39 and a standard deviation of 1.24, with a left-skewed distribution. I used the same analysis approach described above for the Charity Navigator ratings mediation tests. I used “unrelated diversification, two groups” as the independent variable in the first mediation model. In the second mediation model, I used “unrelated diversification, four groups” as the independent variable (see Tables 23 and 25 in Appendix AC). The results of the first mediation model indicate a significant positive relationship between Glassdoor average ratings and PSR ($p < .05$),

which aligns with the hypothesized relationship between perceived trustworthiness and performance. However, the second mediation model results indicate a significant negative relationship between Glassdoor average ratings and PSR ($p < .05$), which contradicts the hypothesized relationship between perceived trustworthiness and performance. There are no other statistically significant findings for any of the relationships of interest.

Control variables. The pooled OLS results consistently demonstrate a significant positive correlation between the international presence of certifiers and PSR ($p < .001$). Certifiers operating internationally, in addition to North America, generate higher PSR. A significant negative relationship exists between the IRS ruling year and PSR ($p < .001$). Firms recognized by the IRS later tend to exhibit lower PSR, aligning with the notion that certifiers typically experience growth over time.

Furthermore, specific models reveal a significant positive correlation between the count of active certifications a certifier offers and PSR ($p < .001$). This intuitive finding suggests that certifiers with a broader array of certifications tend to enjoy greater PSR. Finally, a consistent and significant positive relationship exists between the observation year and PSR ($p < .001$).

Study 2b Discussion

The results of the longitudinal observational analyses in Study 2b provide mixed support for the hypothesis that certifier specialization (low diversification) positively affects certifier performance. Specifically, the hypothesis is supported for unrelated business category diversification but not unrelated diversification across environmental or social sustainability causes. Indeed, diversification has a positive effect across environmental causes. A plausible explanation for this finding aligns with the theoretical model that certifier trustworthiness, particularly certifier authenticity, mediates the relationship between certifier diversification and

performance. Certifiers catering to multiple environmental causes may be viewed as more authentic – they care more about the environment in general. Additionally, Study 2b supports certifier trustworthiness as a mediator between unrelated certifier diversification and performance. Further qualitative analysis of reviews from diverse stakeholders will provide further insight into the relationship between the dimensions of trustworthiness and certifier appeal.

Discussion

The motivation for Paper 2 is to explain the performance variances of certifiers, in general, and sustainability certifiers, in particular. Certifiers are a unique information intermediary that has received little attention from organizational scholars despite being a pervasive component of global markets. Like other organizations, certifiers have responsibilities to donors, investors, employees, and other stakeholders to survive and expand their operations. There is significant variation in the organizational forms of certifiers on dimensions such as profit-seeking status, geographic reach, and, most notably for the present research, degree of market specialization. Little research has examined the strategies and performance predictors of these unique but influential organizational forms.

Specifically, the present research supports that the perceived trustworthiness of certifiers mediates the negative relationship between certifiers' diversification (specialization) and performance. The experimental evidence provided by Study 1 supports the idea that specialized certifiers are rated as having higher overall perceived certifier trustworthiness and higher ability, integrity, and authenticity. Additionally, Study 1 finds that overall perceived certifier trustworthiness and its dimensions (ability, integrity, and authenticity) positively predict participants' intention to use the certifier and that certifier trustworthiness fully mediates the

negative relationship between specialization and participants' intentions to use the certifier. The findings from the observational, longitudinal study (Study 2b) generally support these findings. First, certifiers engaged in unrelated diversification have reduced financial performance compared to certifiers engaged in related diversification or specialization. Second, expert opinions of the quality of certifiers as measured by Charity Navigator ratings are negatively predicted by certifiers' engagement in unrelated diversification and, in turn, positively predict certifiers' financial performance.

These findings align with and extend prior strategy research showing that firms engaged in unrelated diversification tend to underperform their peers (e.g., Bettis, 1981; Palich et al., 2000). Previously, research has focused on the resource disadvantages of unrelated diversification. Specifically, compared to firms that engage in related diversification, firms with unrelated diversification strategies benefit less from economies of scope and are likely to have more complex organizations to manage (Grant et al., 1988; Jones & Hill, 1988; Wan et al., 2011). In addition to the potential for these realized disadvantages, organizations that pursue an unrelated diversification strategy may also be penalized by the perception that they are worse off. Namely, diversified organizations are more likely to be perceived as having less ability, integrity, and authenticity and overall as less trustworthy.

The present research also contributes to the organizational ecology literature. Researchers drawing from an organizational ecology perspective have found that organizations that span categories and, as such, occupy a broad market niche tend to be penalized (e.g., Hsu, 2006; Leung & Sharkey, 2014; Vera et al., 2018; Zuckerman et al., 2003). A proposed rationale for this is that such entities are more ambiguous to observers, making them more challenging to understand and reducing their appeal (Hannan et al., 2019). The present research aligns with this

argument and finds additional evidence for this playing out at the field level. Prior research on niche width has left unknown and untested any underlying cognitive mechanisms that make occupying a narrow niche, specializing, a potentially helpful strategy. Though, according to researchers within the fields of organizational ecology, specializing is thought to signal enhanced resources, expertise, and engagement within the specialized domain (Hannan et al., 2003; Hsu et al., 2009; Negro, Koçak, & Hsu, 2010). The present research supports perceived trustworthiness and its dimensions, including perceived ability, as one (of potentially multiple) cognitive mechanisms underlying the relationship between specialization and performance.

Limitations and Future Directions

Despite the comprehensive nature of this research, certain limitations warrant consideration, and there are several opportunities for future research to uncover a deeper understanding of the performance drivers of certifiers. One limitation of the present research is that it does not delve into potential fluency effects in decision-making processes related to certifier trustworthiness. Future inquiries could explore how automatic processing versus deliberate decision-making shapes stakeholders' perceptions of certifiers and their subsequent performance evaluations.

While this research highlights the significant role of producers in determining certifier performance, the perspective of other stakeholders, such as end consumers and donors, is also critical. Future research could explore differences in the beliefs and attitudes of stakeholder groups and their influence on certifiers' performance outcomes. Additionally, future investigations could undertake a finer-grained analysis by mapping Charity Navigator beacons to specific dimensions of certifier trustworthiness, such as integrity, ability, authenticity, and

trustworthiness. This nuanced approach holds promise for further understanding the intricacies of stakeholder perceptions regarding certifiers.

Finally, as theorized, unrelated diversification emerges as a meaningful predictor of certifier trustworthiness. However, there are certainly other factors contributing to certifier trustworthiness that warrant investigation. Future studies could explore the impact of variables such as certifiers' profit-seeking status, partnerships, and organizational structure on perceived trustworthiness and performance metrics.

Conclusion

In summary, this dissertation offers a wide-reaching exploration of certifiers as a unique organizational form and their role as information intermediaries between producers and consumers in the market for sustainable offerings. Global markets are witnessing a surge in demand for environmentally and socially conscious products and services. Certifiers serve as crucial intermediaries in this market, bridging the gap between producers and consumers by assuring the quality of certified offerings.

The development and validation of the Certifier Trustworthiness Scale (CTS) represent a significant contribution to the literature, providing a robust framework for assessing the perceived trustworthiness of certifiers. Through a nuanced examination of perceived ability, integrity, and authenticity, the CTS illuminates the multifaceted nature of trust in certifiers.

Furthermore, the second paper delves into the implications of certifier diversification, particularly in the context of sustainability certifications. By drawing on organizational ecology principles, the study reveals how unrelated diversification can diminish certifiers' perceived trustworthiness and, consequently, their performance outcomes. This underscores the importance of specialization in signaling expertise and commitment within a specific domain, thus enhancing certifiers' credibility and effectiveness.

Overall, this dissertation underscores the pivotal role of trustworthiness in certification markets, emphasizing the need for continued research to inform certifiers' strategies and practices, meet stakeholders' evolving demands, and advance sustainability and quality standards globally.

Paper 2 Appendices

Appendix J: Study 1 – Preliminary Certifier Trustworthiness Scale (CTS)

Prompt: “Please rate your agreement with the following statements **about [certifier name, e.g. Earth Friendly Goods].**”

Scale: 1 – 7 where 1 = Strongly Disagree and 7 = Strongly Agree

Perceived Ability

8. ...is able to achieve its goals.
9. ...has the expertise needed to achieve its goals.
10. ...has the expertise needed to set appropriate standards.
11. ...has the resources it needs to check if products meet its standards.
12. ...is competent.
13. ...knows how to enforce standards.
14. ...shows high ability.

Perceived Integrity

11. ...would not knowingly break the law.
12. ...holds all potential certification holders to the same standards.
13. ...enforces standards consistently.

Perceived Authenticity

16. ...can be trusted to have [cause(s) here, e.g., environmental sustainability’s] best interests at heart.
17. ...truly cares about [cause(s) here].
18. ...aims to set standards that support [cause(s) here].
19. ...prioritizes [cause(s) here].
20. ...genuinely supports [cause(s) here].

Appendix K: Environmental Concern Scale

Prompt: “Please rate your agreement with the following statements **about Earth Friendly Goods.**”

Scale: 1 – 7 where 1 = Strongly Disagree and 7 = Strongly Agree

Adapted from: Steel, B. S. (1996). Thinking globally and acting locally?: environmental attitudes, behaviour and activism. *Journal of environmental management*, 47(1), 27-36.

1. The balance of nature is very delicate and easily upset by human activities.
2. The earth is like a spaceship with only limited room and resources.
3. Plants and animals do not exist primarily for human use.
4. Modifying the environment for human use seldom causes serious problems.

Appendix L: Study 1 – Respondent Characteristics by Condition**Table 11.** Respondent Characteristics by Condition (Study 1, Experimental Data)

	Specialization		Related Diversification		Unrelated Diversification	
	Count	%	Count	%	Count	%
Gender						
Female	73	52.9%	59	44.4%	76	58.0%
Male	61	44.2%	72	54.1%	52	39.7%
Free response	4	2.9%	2	1.5%	3	2.3%
Educational attainment						
High school degree or equivalent	16	11.7%	22	16.5%	17	13.0%
Some college but no degree	30	13.0%	35	26.3%	31	23.7%
Associate degree	14	23.7%	13	9.8%	15	11.5%
Bachelor degree	62	39.7%	49	36.8%	52	39.7%
Graduate degree	16	12.2%	14	10.5%	16	12.2%
Employment status						
Employed, working 1-39 hours per week	39	28.3%	47	35.3%	33	25.2%
Employed, working 40 or more hours per week	52	37.7%	49	36.8%	49	37.4%
Not employed, looking for work	10	7.3%	14	10.5%	14	10.7%
Not employed, not looking for work	3	2.2%	2	1.6%	6	4.6%
Retired	12	8.7%	9	6.8%	12	9.2%
Homemaker	9	6.5%	5	3.8%	7	5.3%
Student	10	7.3%	5	3.8%	7	5.3%
Disabled, not able to work	3	2.2%	2	1.5%	3	2.3%
	Mean		Mean		Mean	
Age	39.1		40.5		40.3	

Appendix M: Study 1 – Regression Tables**Table 12a.** Impact of Specialization-Diversification Condition on Perceived Certifier Trustworthiness (Study 1, Experimental Data)

	Certifier trustworthiness	Ability	Integrity	Authenticity
	(1)	(2)	(3)	(4)
Conditions				
Specialization	6.071*** (0.062)	6.017*** (0.068)	6.001*** (0.078)	6.187*** (0.065)
Related diversification	-0.225* (0.089)	-0.216* (0.097)	-0.298** (0.112)	-0.195* (0.093)
Unrelated diversification	-0.187* (0.089)	-0.132 (0.098)	-0.264* (0.112)	-0.218* 0.093
Observations	402	402	402	402
R-squared	0.018	0.012	0.021	0.016
F-statistic	3.709*	2.513 ⁺	4.297*	3.309*

Note. Standard errors in parentheses.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Table 12b. Impact of Specialization-Diversification Condition on Perceived Certifier Trustworthiness (Study 1, Experimental Data)

	Certifier			
	trustworthiness	Ability	Integrity	Authenticity
	(5)	(6)	(7)	(8)
Conditions				
Specialization	4.458*** (0.370)	4.733*** (0.416)	4.168*** (0.474)	4.249*** (0.384)
Related diversification	-0.170* (0.084)	-0.168+ (0.095)	-0.238* (0.108)	-0.131 (0.088)
Unrelated diversification	-0.119 (0.084)	-0.071 (0.095)	-0.192+ (0.108)	-0.143 -0.087
Moderating variables				
Political ideology	-0.027 (0.019)	-0.010 (0.021)	-0.058* (0.024)	-0.032 (0.019)
Familiarity with certifiers	0.063** (0.023)	0.068** (0.026)	0.073* (0.029)	0.05 (0.024)
Environmental concern	0.134* (0.061)	0.110 0.069	0.172* (0.079)	0.145 (0.064)
Warmth/Competence	0.280*** (0.043)	0.244*** (0.050)	0.279*** (0.055)	0.330 (0.045)
Control variables				
Age	-0.003 (0.003)	-0.005 (0.003)	-0.002 (0.003)	-0.002 (0.003)
Gender	-0.030 (0.066)	-0.034 (0.074)	-0.003 (0.084)	-0.041 (0.068)
Educational attainment	-0.036 (0.024)	-0.047 (0.027)	-0.039 (0.031)	-0.019 (0.025)
Employment status	0.020 (0.019)	0.013 (0.021)	0.036 (0.024)	0.020 (0.019)
Observations	402	402	402	402
R-squared	0.108	0.155	0.126	0.172
F-statistic	7.19***	4.734***	5.655***	8.112***

Note. Standard errors in parentheses.

+p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Appendix N: Study 1 – Regression Tables**Table 13a.** Impact of Specialization-Diversification Condition on Perceived Certifier Trustworthiness (Study 1, Experimental Data)

	Certifier trustworthiness	Ability	Integrity	Authenticity
	(1)	(2)	(3)	(4)
Conditions				
Related diversification	5.846*** (0.063)	5.800*** (0.070)	5.707*** (0.080)	5.992*** (0.066)
Specialization	0.225* (0.089)	0.216* (0.097)	0.298** (0.112)	0.194* (0.093)
Unrelated diversification	0.038 (0.090)	0.084 (0.099)	0.034 (0.113)	-0.023 (0.094)
Observations	402	402	402	402
R-squared	0.0183	0.012	0.021	0.016
F-statistic	3.709*	2.513 ⁺	4.297*	3.309*

Note. Standard errors in parentheses.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Appendix O: Study 1 – Regression Tables**Table 13b.** Impact of Specialization-Diversification Condition on Perceived Certifier Trustworthiness (Study 1, Experimental Data)

	Certifier			
	trustworthiness	Ability	Integrity	Authenticity
	(5)	(6)	(7)	(8)
Conditions				
Related diversification	4.288*** (0.366)	4.565*** (0.412)	3.930*** (0.469)	4.117*** (0.380)
Specialization	0.17* (0.084)	0.168. (0.095)	0.238* (0.108)	0.131 (0.088)
Unrelated diversification	0.051 (0.085)	0.098 (0.096)	0.046 (0.109)	-0.011 -0.088
Moderating variables				
Political ideology	-0.027 (0.019)	-0.010 (0.021)	-0.058* (0.024)	-0.032 -0.019
Familiarity with certifiers	0.063** (0.023)	0.068** (0.026)	0.073* (0.029)	0.050 (0.024)
Environmental concern	0.134* (0.061)	0.109 (0.069)	0.172* (0.079)	0.145 (0.064)
Warmth/Competence	0.280*** (0.043)	0.244*** (0.049)	0.279*** (0.055)	0.330 (0.045)
Control variables				
Age	-0.003 (0.003)	-0.005. (0.003)	-0.002 (0.003)	-0.002 (0.003)
Gender	-0.030 (0.066)	-0.034 (0.074)	-0.003 (0.084)	-0.041 (0.068)
Educational attainment	-0.036 (0.024)	-0.047+ (0.027)	-0.039 (0.031)	-0.019 (0.025)
Employment status	0.020 (0.019)	0.013 (0.021)	0.036 (0.024)	0.020 (0.019)
Observations	402	402	402	402
R-squared	0.155	0.108	0.126	0.172
F-statistic	7.190***	4.734***	5.655***	8.112***

Note. Standard errors in parentheses.

+p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Appendix P: Study 1 – Regression Tables**Table 14a.** Impact of Specialization-Diversification Condition on Purchase Intention (Study 1, Experimental Data)

	Purchase intention	
	(1)	(2)
Conditions		
Specialization	2.239*** (0.601)	5.213*** (0.102)
Related diversification	-0.004 (0.137)	-0.120 (0.145)
Unrelated diversification	-0.111 (0.137)	-0.225 (0.146)
Moderating variables		
Political ideology	-0.091** (0.030)	
Familiarity with certifiers	0.160*** (0.037)	
Environmental concern	0.302** (0.100)	
Warmth/Competence	0.346*** (0.070)	
Control variables		
Age	-.091 (0.030)	
Gender	0.012 -0.106	
Educational attainment	-0.015 -0.039	
Employment status	0.083** -0.03	
Observations	402	402
R-squared	0.156	0.006
F-statistic	7.201***	1.202

Note. Standard errors in parentheses.

[†]p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Appendix Q: Study 1 – Regression Tables**Table 15a.** Certifier Trustworthiness on Purchase Intention (Study 1, Experimental Data)

	Purchase intention			
	(1)	(2)	(3)	(4)
Independent variables				
Certifier trustworthiness	0.878*** (0.068)	-	-	-
Ability	-	0.731*** (0.063)	-	-
Integrity	-	-	0.565*** (0.057)	-
Authenticity	-	-	-	0.732*** (0.070)
Moderating variables				
Political ideology	-0.068** (0.025)	-0.085** (0.026)	-0.059* (0.027)	-0.068* (0.027)
Familiarity with certifiers	0.102** (0.031)	0.109*** (0.032)	0.117*** (0.033)	0.122*** (0.033)
Environmental concern	0.186* (0.084)	0.225** (0.086)	0.206* (0.090)	0.197* (0.089)
Warmth/Competence	0.099 (0.062)	0.169** (0.062)	0.187** (0.065)	0.104 (0.066)
Control variables				
Age	-0.003 (0.003)	-0.002 (0.004)	-0.004 (0.004)	-0.004 (0.004)
Gender	0.050 (0.089)	0.050 (0.091)	0.023 (0.095)	0.049 (0.094)
Educational attainment	0.013 (0.033)	0.016 (0.034)	0.004 (0.035)	-0.003 (0.034)
Employment status	0.061* (0.025)	0.069** (0.026)	0.058* (0.027)	0.065* (0.027)
Constant	-1.624**	-1.216*	-0.069	-0.840
Observations	402	402	402	402
R-squared	0.404	0.373	0.324	0.34
F-statistic	29.49***	25.85***	20.9***	22.44***

Note. Standard errors in parentheses.

* $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$; two-tailed tests.

Appendix R: Study 1 – Mediation Tables**Table 16a.** Mediation by Certifier Trustworthiness

	Estimate	Z-value	P-value
Condition → Certifier trustworthiness	0.020 (0.045)	0.436	0.663
Certifier trustworthiness → Purchase intention	0.974*** (0.065)	15.062	0.000
Condition → Purchase intention	-0.071 (0.059)	-1.219	0.223
Condition → Certifier trustworthiness → Purchase intention (Indirect)	0.019 0.044	0.436	0.663

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

Table 16b. Mediation by Certifier Trustworthiness

	Estimate	Z-value	P-value
Specialization ¹ → Certifier trustworthiness	0.206** (0.076)	2.699	0.007
Certifier trustworthiness → Purchase intention	0.975*** (0.065)	14.917	0.000
Specialization ¹ → Purchase intention	-0.029 (0.101)	-0.288	0.774
Specialization ¹ → Certifier trustworthiness → Purchase intention (Indirect)	0.201** (0.076)	2.656	0.008

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

¹Specialization compared to Related and Unrelated Diversification, combined

Appendix S: Study 1 – Mediation Tables**Table 16c.** Mediation by Perceived Ability

	Estimate	Z-value	P-value
Specialization ¹ → Ability	0.175* (0.084)	2.079	0.038
Ability → Purchase intention	0.818*** (0.062)	13.197	0.000
Specialization ¹ → Purchase intention	0.029 (0.105)	0.28	0.779
Specialization ¹ → Ability → Purchase intention (Indirect)	0.143* (0.070)	2.054	0.040

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

¹Specialization compared to Related and Unrelated Diversification, combined

Table 16d. Mediation by Perceived Integrity

	Estimate	Z-value	P-value
Specialization ¹ → Integrity	.281** (0.096)	2.927	0.003
Integrity → Purchase intention	0.667*** (0.056)	11.991	0.000
Specialization ¹ → Purchase intention	-0.015 (0.108)	-0.143	0.886
Specialization ¹ → Integrity → Purchase intention (Indirect)	0.188** (0.066)	2.844	0.004

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

¹Specialization compared to Related and Unrelated Diversification, combined

Appendix T: Study 1 – Mediation Tables**Table 16e.** Mediation by Perceived Authenticity

	Estimate	Z-value	P-value
Specialization ¹ → Authenticity	0.206** (0.080)	2.570	(0.010)
Authenticity → Purchase intention	0.83*** (0.066)	12.616	0.000
Specialization ¹ → Purchase intention	0.001 (0.107)	0.011	0.991
Specialization ¹ → Authenticity → Purchase intention (Indirect)	0.171* (0.068)	2.518	0.012

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

* $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$; two-tailed tests.

¹Specialization compared to Related and Unrelated Diversification, combined

Appendix U: Study 2a-b - Intengine product/business categories

Original label from Intengine	Label used in Study 2a
1. Agriculture, Horticulture & Viticulture	agriculture and horticulture
2. Apparel, Textiles & Accessories	apparel, textiles, and accessories
3. Automotive & Personal Transportation	automotive and personal transportation
4. Baby, Toddlers & Kids Care	baby, toddlers, and kids care
5. Business & Professional Services	business and professional services
6. Computers & Electronics	computers and electronics
7. Conferences, Events & Trade Shows	conferences and events
8. Construction	construction
9. Dining & Entertainment	dining and entertainment
10. Energy, Environment & Carbon	energy and environment
11. Finance	finance
12. Fisheries & Aquaculture	fisheries and aquaculture
13. Fitness, Sports & Recreation	fitness, sports, and recreation
14. Food & Beverage	food and beverage
15. Forestry & Silviculture	forestry
16. Healthcare, Nutrition & Healing	healthcare and nutrition
17. Hobbies, Crafts & Toys	hobbies, crafts, and toys
18. Home & Garden	home and garden
19. Industry	N/A
20. Media & Communications	media and communication
21. Office Equipment & Supplies	office supplies
22. Packaging	packaging
23. Personal Care, Beauty & Spas	personal and beauty care
24. Real Estate	real estate
25. Specialty	N/A
26. Spirituality & Consciousness	spirituality and consciousness
27. Sustainability Reporting	sustainability reporting
28. Transportation & Public Transit	public transportation
29. Travel, Tourism & Lodging	travel, tourism, and lodging
30. Waste Management & Recycling	waste management and recycling
31. Water Quality & Water Resource Management	water quality and water management

Appendix V: Glassdoor Star Ratings

Source: Ratings on Glassdoor, 2024

Star Rating	Interpretation
0.00 – 1.50	Employees are “Very Dissatisfied”
1.51 – 2.50	Employees are “Dissatisfied”
2.51 – 3.50	Employees say it’s “OK”
3.51 – 4.00	Employees are “Satisfied”
4.01 – 5.00	Employees are “Very Satisfied”

Appendix W: Study 2 – Certifier Descriptive Statistics**Table 18.** Certifier Descriptive Statistics

	1989	2000	2010	2021
Count ¹	2	23	55	72
IRS Subsection Type				
Public Charity	2	15	37	51
Business League	-	8	17	18
Civic League	-	-	1	2
Private Operating Foundation	-	-	-	1
Other	-	-	-	2
Average Active Certifications	1	1.478	2.382	3.806
Maximum Active Certifications	1	6	12	33
Average Active Product Categories	1	7.565	5.727	7.472
Maximum Active Product Categories	1	31	31	31

Notes. ¹Excludes certifiers identified as outliers based on Total Revenue for at least one year in the period from 2010 to 2021

Appendix Y: Classification of United Nations Sustainable Development Goals (SDGs)

Goal Number	Goal Name	Classification
1	No poverty	Social
2	Zero hunger	Social
3	Good health and well-being	Social
4	Quality education	Social
5	Gender equality	Social
6	Clean water and sanitation	Both environmental and social
7	Affordable and clean energy	Environmental
8	Decent work and economic growth	Social
9	Industry, innovation, and infrastructure	Social
10	Reduced inequalities	Social
11	Sustainable cities and communities	Both environmental and social
12	Responsible consumption and production	Environmental
13	Climate action	Environmental
14	Life below water	Environmental
15	Life on land	Environmental
16	Peace, justice, and strong institutions	Social
17	Partnerships for the goals	N/A

Appendix Z: Study 2 Figures (page 1 of 2)

Figure 1.

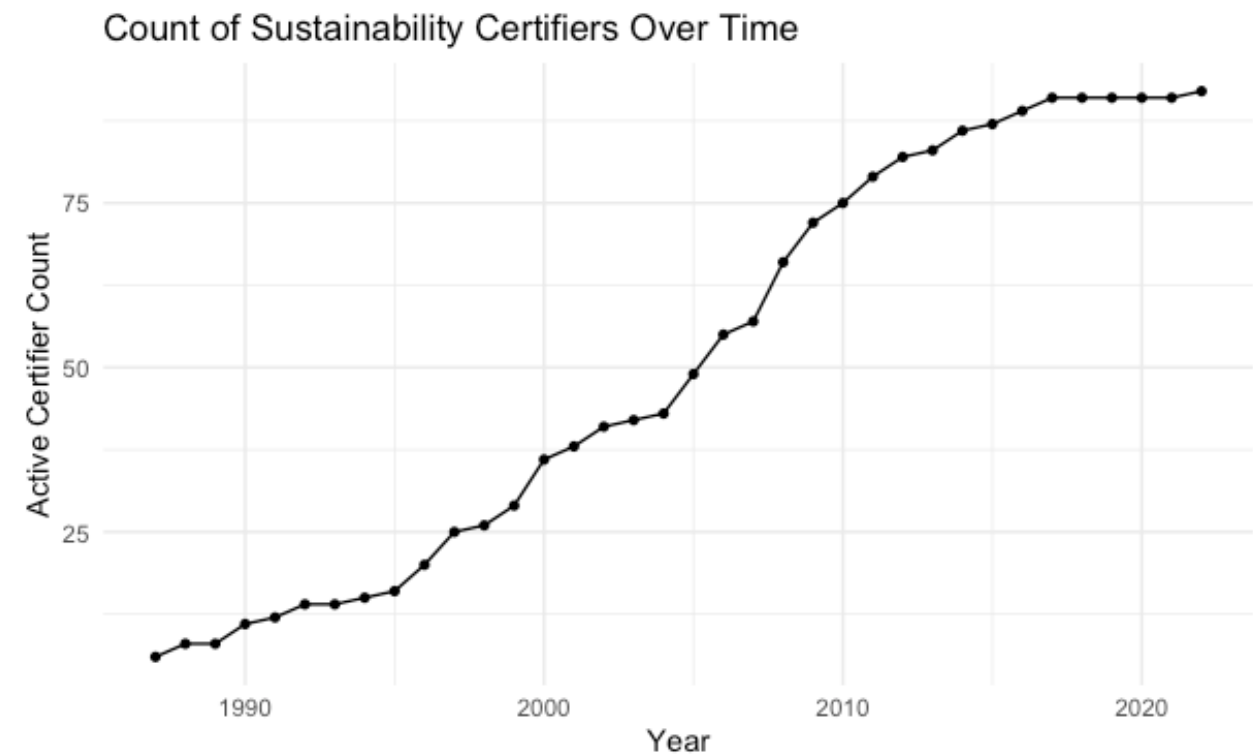


Figure 2.

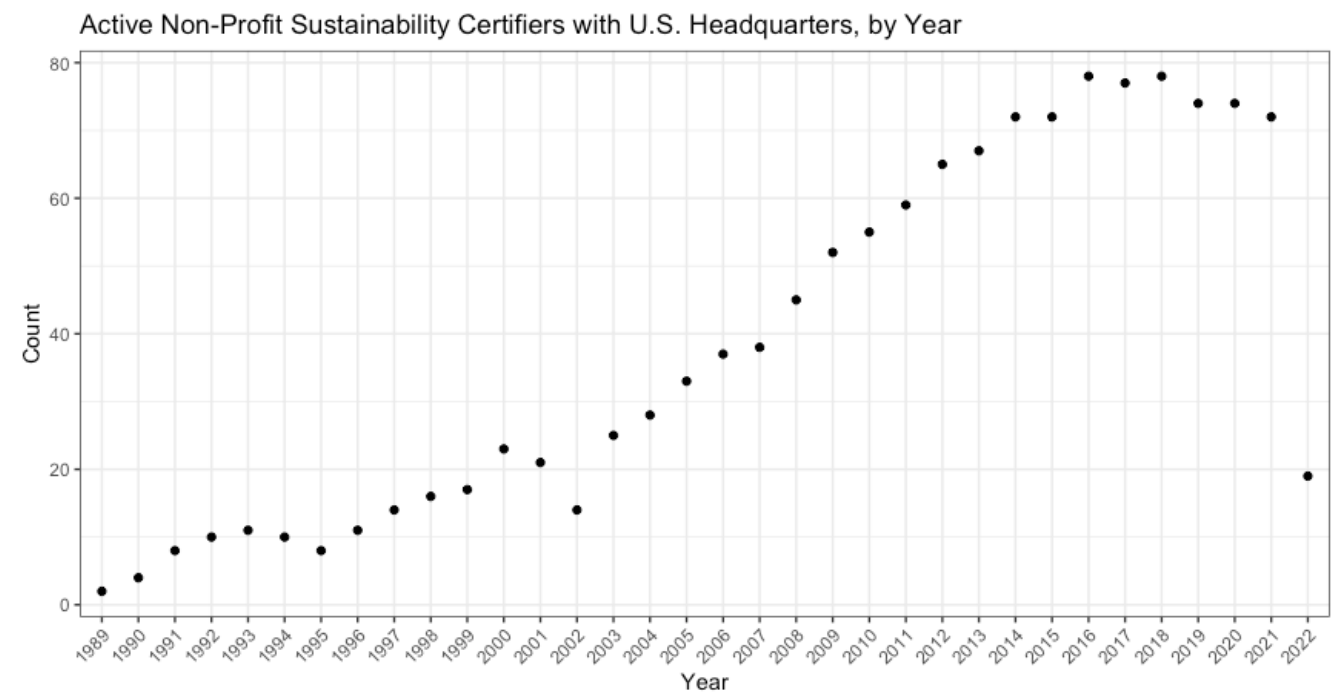
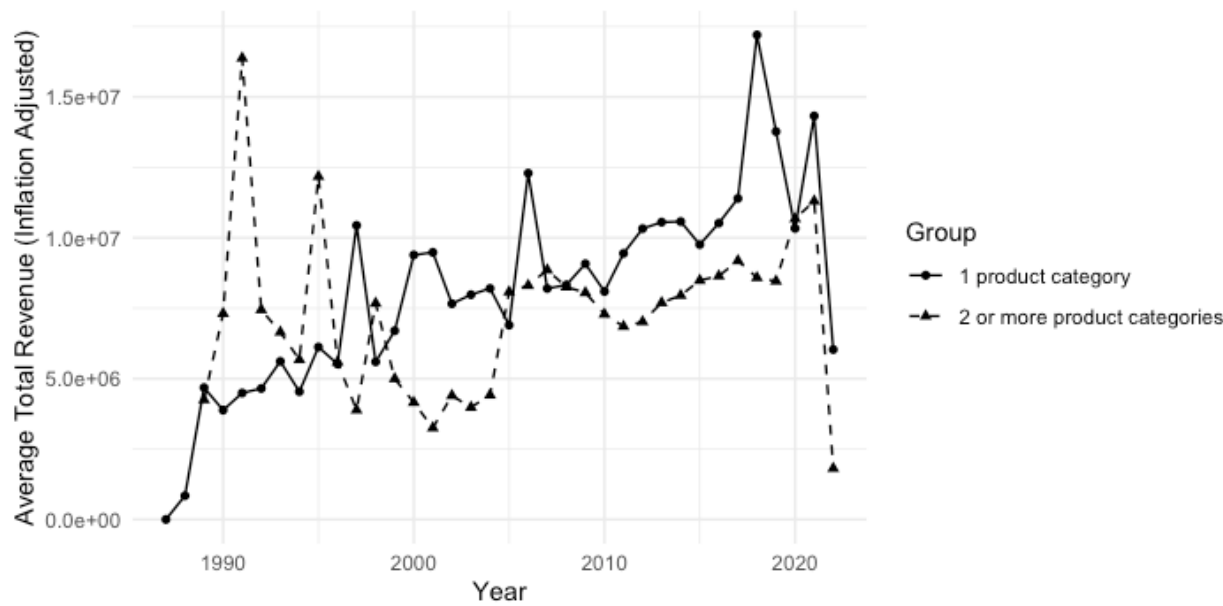
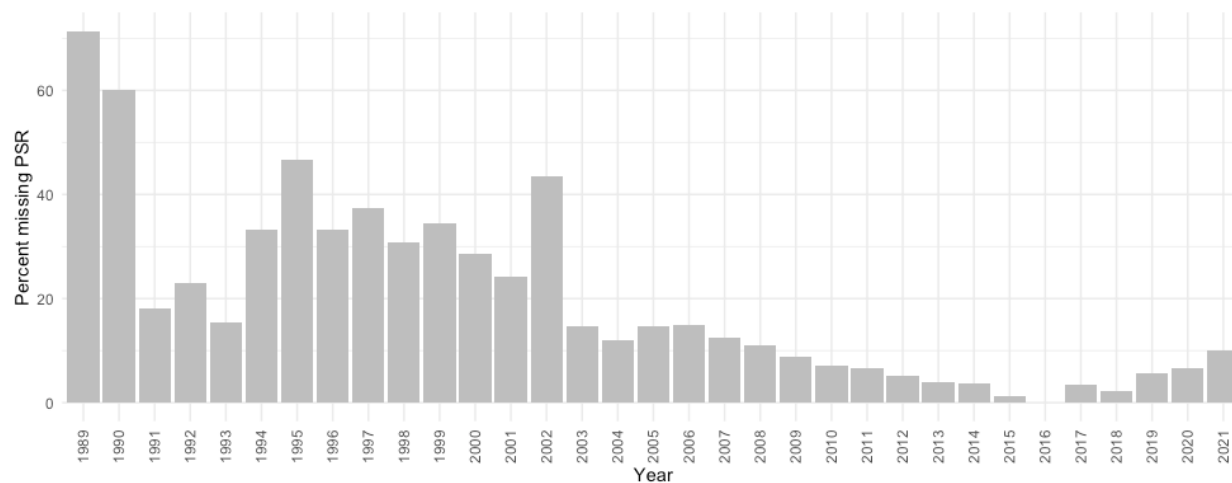


Figure 3.
Longitudinal Plot of Average Total Revenue, Inflation Adjusted



Appendix AA: Study 2 Figures (page 2 of 2)**Percentage of missing Program Services Revenue values by observation year**

Appendix AB: Study 2 - Panel Regression Results

Table 19. Fixed Effects Panel Model for predicting Program Services Revenue

	Model 1	2	3	4	5	6	7
Product categories (continuous)	0.028 (0.017)	-	-	-	-	-	-0.103 (0.054)
2+ product categories (binary)	-	-	-	-	-	-0.017 (0.131)	-
3+ product categories (binary)	-	-0.209 (0.138)	-	-0.473** (0.163)	-0.533** (0.166)	-	-
Square of product categories	-	-	-	-	-	-	0.004* (0.002)
Environmental SDG count	-	-	0.267*** (0.079)	0.327** (0.106)	0.320*** (0.106)	0.150 (0.094)	0.256* (0.103)
Social SDG count	-	-	-	-	0.157 (0.088)	0.100 (0.087)	-
Certification count	0.037* (0.015)	0.053*** (0.013)	-	0.031* (0.015)	0.028 (0.015)	0.033* (0.015)	0.038* (0.017)
Year	0.077*** (0.004)	0.078*** (0.005)	0.080*** (0.004)	0.078*** (0.004)	0.077*** (0.004)	0.076*** (0.005)	0.078*** (0.005)
n	86	87	87	87	87	87	87
N	1274	1289	1289	1289	1289	1289	1289
R ²	0.280	0.280	0.277	0.285	0.287	0.281	0.285
R ² adj.	0.226	0.226	0.224	0.232	0.233	0.227	0.231
F-statistic	155.464 ***	155.281 ***	230.128 ***	119.652 ***	96.528***	93.663***	95.452***

¹All financial data is inflation adjusted to 2022 USD²All financial data is transformed to a logarithmic scale due to skewness

+p < .1; *p < .05; **p < .01; ***p < .001

Table 20: Fixed Effects Panel Model for predicting Program Services Revenue

	Model 8	9	10	11	12	13
Unrelated diversification, two primary groups ³	0.057 (0.136)	-	-0.026 (0.221)	0.214 (0.200)	-0.222 (0.361)	-0.096 (0.214)
Unrelated diversification, four groups ⁴	-	-0.180 (0.130)	-	-	-	-
Unrelated diversification within first primary group ⁴	-	-	-	-0.653* (0.278)	-	0.952*** (0.286)
Unrelated diversification within second primary group ⁵	-	-	-0.075 (0.186)	-	-0.118 (0.187)	-
Environmental SDG count	-	-	-	-	0.113 (0.125)	0.527*** (0.143)
Social SDG count	-	-	-	-	0.174 (0.094)	0.316** (0.114)
Certification count	0.077*** (0.005)	0.048*** (0.013)	0.047*** (0.006)	0.066*** (0.018)	0.037* (0.016)	0.035 (0.019)
Year	0.077*** (0.005)	0.084*** (0.004)	0.072*** (0.006)	0.083*** (0.006)	0.070*** (0.006)	0.080*** (0.006)
n	87	87	74	57	74	57
N	1289	1289	1015	832	1015	832
R ²	0.279	0.278	0.239	0.284	0.242	0.304
R ² adj.	0.225	0.225	0.176	0.228	0.178	0.248
F-statistic	154.311 ***	154.239 ***	73.420 ***	76.442 ***	49.859***	56.013***

¹All financial data is inflation adjusted to 2022 USD

²All financial data is transformed to a logarithmic scale due to skewness

³See Appendix K, “Primary Grouping” for the two groupings used

⁴See Appendix K, “Sub-groups for First Primary Group” for the groupings used

⁵See Appendix K, “Sub-groups for Second Primary Group” for the groupings used

+p < .1; *p < .05; **p < .01; ***p < .001

Table 21: Pooled OLS Model for predicting Program Services Revenue

	Model 14	15	16	17	18	19
Unrelated diversification, two primary groups ³	-0.667*** (0.116)	-0.894*** (0.114)	-0.961*** (0.107)	-1.451*** (0.172)	-1.226*** (0.139)	-
Unrelated diversification, four groups ⁴	-	-	-	-	-	-0.842*** (0.107)
Unrelated diversification within first primary group ⁴	-	-	-	1.061*** (0.155)	-	-
Unrelated diversification within second primary group ⁵	-	-	-	-	0.552*** (0.136)	-
Environmental SDG count	-	0.213*** (0.057)	0.400*** (0.055)	0.335*** (0.067)	0.270*** (0.062)	0.406*** (0.055)
Social SDG count	-	0.266*** (0.051)	0.351*** (0.048)	0.381*** (0.057)	0.249*** (0.056)	0.327*** (0.048)
International certifier	-	0.932*** (0.107)	0.889*** (0.101)	1.401*** (0.137)	0.811*** (0.127)	0.851*** (0.101)
IRS Ruling Year	-	-	-0.038*** (0.003)	-	-	-0.040*** (0.003)
Certification count	0.092*** (0.017)	0.022 (0.021)	-0.021 (0.020)	0.006 (0.023)	0.011 (0.022)	-0.025 (0.020)
Year	0.050*** (0.008)	0.044*** (0.007)	0.072*** (0.007)	0.056*** (0.023)	0.024** (0.009)	0.075*** (0.007)
n	87	87	85	57	74	85
N	1289	1289	1286	832	1015	1286
R ²	0.073	0.149	0.254	0.297	0.143	0.244
R ² adj.	0.071	0.146	0.250	0.291	0.138	0.240
F-statistic	33.668***	37.551***	62.268***	49.680***	24.098***	58.839***

¹All financial data is inflation adjusted to 2022 USD²All financial data is transformed to a logarithmic scale due to skewness³See Appendix K, “Primary Grouping” for the two groupings used⁴See Appendix K, “Sub-groups for First Primary Group” for the groupings used⁵See Appendix K, “Sub-groups for Second Primary Group” for the groupings used

*p < .1; **p < .05; ***p < .01; ****p < .001

Table 22: Pooled OLS Model for predicting Program Services Revenue

	Model 20	21	22	23	24	25
Unrelated diversification, two primary groups ³	-0.894** (0.114)	-	0.119 (0.266)	0.268 (0.263)	0.118 (0.269)	-0.532*** (0.127)
Unrelated diversification, four groups ⁴	-	-0.605*** (0.113)	-	-	-	-
Environmental SDG count	0.213*** (0.057)	0.201*** (0.057)	0.184* (0.091)	0.162 (0.089)	0.185* (0.092)	0.255*** (0.058)
Social SDG count	0.266*** (0.051)	0.231*** (0.051)	0.097 (0.098)	0.027 (0.097)	0.102 (0.098)	0.210*** (0.006)
Annual average Glassdoor rating	-	-	-0.191 (0.099)	-	-	-
Annual median Glassdoor rating	-	-	-	-	-0.129 (0.095)	-
Average Glassdoor rating (all years)	-	-	-	-0.580*** (0.154)	-	-
Charity Navigator rating	-	-	-	-	-	0.036*** (0.006)
International certifier	0.932*** (0.107)	0.885*** (0.108)	1.509*** (0.294)	1.471*** (0.287)	1.516*** (0.296)	1.033*** (0.124)
IRS Ruling Year						-0.021*** (0.005)
Certification count	0.022 (0.021)	0.145 (0.021)	-0.003 (0.029)	0.007 (0.028)	-0.006 (0.029)	-0.010 (0.019)
Year	0.044*** (.007)	0.044*** (0.007)	-0.150*** (0.039)	-0.140*** (0.037)	-0.154*** (0.039)	0.084*** (0.008)
n	87	87	45	45	45	49
N	1289	1289	192	192	192	785
R ²	0.149	0.128	0.279	0.317	0.271	0.298
R ² adj.	0.146	0.124	0.251	0.291	0.244	0.291
F-statistic	37.551***	31.352***	10.152***	12.193***	9.785***	41.148***

¹All financial data is inflation adjusted to 2022 USD²All financial data is transformed to a logarithmic scale due to skewness³See Appendix K, “Primary Grouping” for the two groupings used⁴See Appendix K, “Sub-groups for First Primary Group” for the groupings used⁵See Appendix K, “Sub-groups for Second Primary Group” for the groupings used

*p < .1; **p < .05; ***p < .01; ****p < .001

Appendix AC: Study 2b – Mediation Results**Table 23.** Study 2b - Mediation by Charity Navigator ratings

	Estimate	Z-value	P-value
Unrelated diversification, two groups → Charity Navigator ratings	-2.181** (0.764)	-2.856	0.004
Charity Navigator ratings → Program services revenue	0.052*** (0.006)	8.895	0.000
Unrelated diversification, two groups → Program services revenue	-0.101 (0.127)	-0.794	0.427
Control variable: Year	0.088*** (0.008)	10.788	0.000
Unrelated diversification, two groups → Charity Navigator ratings → Program services revenue (Indirect)	-0.112** (0.041)	-2.719	0.007

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

+p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

N = 785

Table 24. Study 2b - Mediation by Charity Navigator ratings

	Estimate	Z-value	P-value
Unrelated diversification, four groups → Charity Navigator ratings	-5.552*** (0.748)	-7.427	0.000
Charity Navigator ratings → Program services revenue	0.053*** (0.006)	8.904	0.000
Unrelated diversification, four groups → Program services revenue	-0.121 (0.132)	0.913	0.361
Control variable: Year	0.085*** (0.008)	10.397	0.000
Unrelated diversification, four groups → Charity Navigator ratings → Program services revenue (Indirect)	-0.295*** (0.052)	-5.703	0.000

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

+p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

N = 785

Table 25. Study 2b - Mediation by 5 ratings

	Estimate	Z-value	P-value
Unrelated diversification, two groups → Glassdoor ratings	0.335 ⁺ (0.180)	1.861	0.063
Glassdoor ratings → Program services revenue	0.228* (0.102)	-2.236	0.025
Unrelated diversification, two groups → Program services revenue	0.660 (0.257)	2.568	0.010
Control variable: Year	-0.167*** (0.039)	-4.233	0.000
Unrelated diversification, two groups → Glassdoor ratings → Program services revenue (Indirect)	-0.076 (0.053)	-1.430	0.153

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

N = 192

Table 26. Study 2b - Mediation by Glassdoor ratings

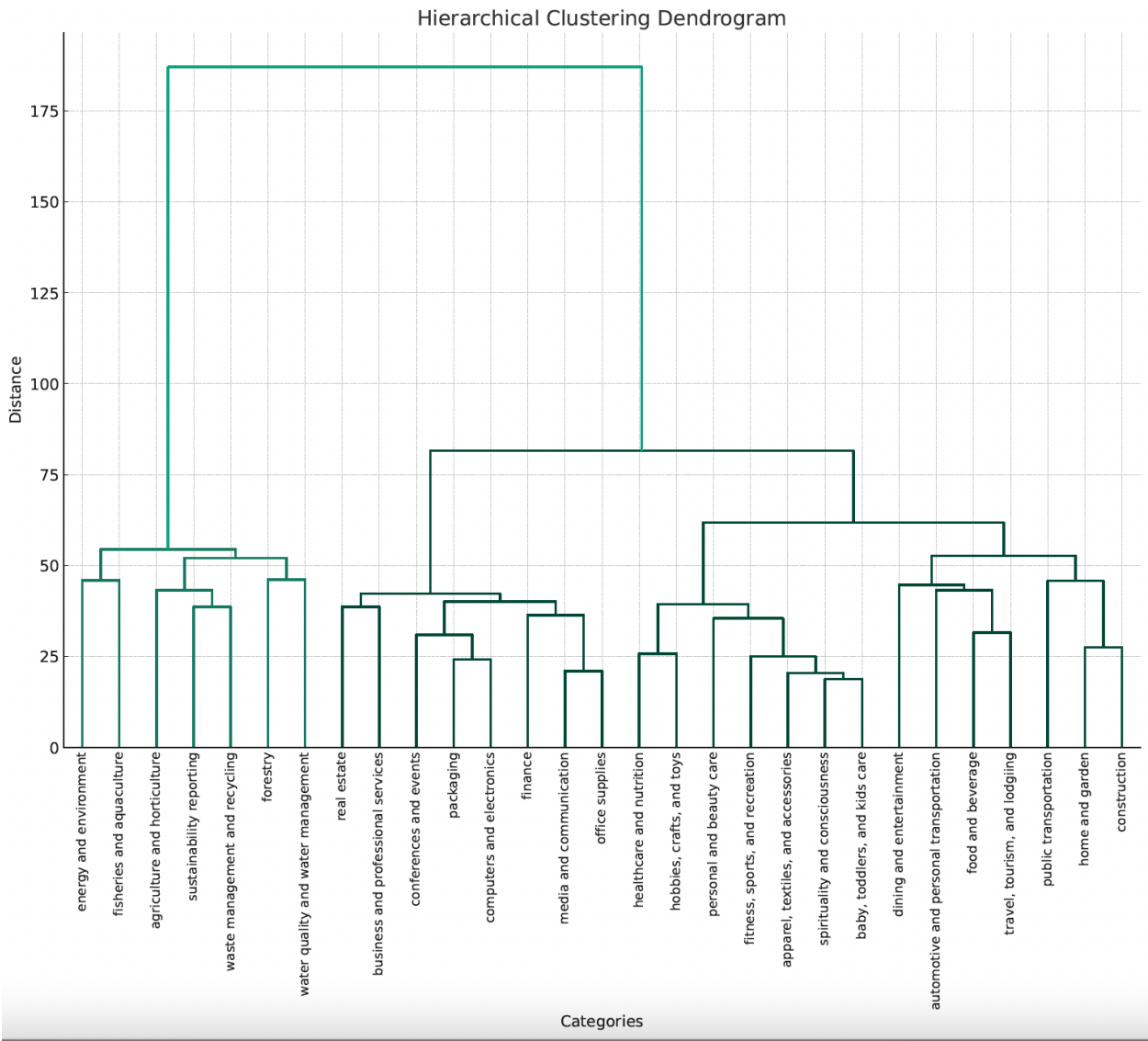
	Estimate	Z-value	P-value
Unrelated diversification, four groups → Glassdoor ratings	0.046 (0.189)	0.243	0.808
Glassdoor ratings → Program services revenue	-0.204* (0.101)	-2.012	0.044
Unrelated diversification, four groups → Program services revenue	0.630* (0.266)	2.369	0.018
Control variable: Year	-0.156*** (0.040)	-3.943	0.000
Unrelated diversification, four groups → Glassdoor ratings → Program services revenue (Indirect)	-0.009 (0.039)	-0.242	0.809

Note. Standard errors in parentheses; Lavaan package in R; ML estimator, NLMINB optimization.

⁺p < .1; *p < .05; **p < .01; ***p < .001; two-tailed tests.

N = 192

Appendix AD: Study 2a – Hierarchical Clustering



Appendix AE: Study 2a - Panel Regression Results

Business Category	Primary Grouping (Two Groups) from hierarchical clustering	Sub-groups for First Primary Group from k-means clustering	Sub-groups for Second Primary Group from k-means clustering	Secondary Grouping (Four Groups) from k-means clustering
energy and environment	Managing the Natural Environment	Energy and Fisheries	N/A	Managing the Natural Environment
fisheries and aquaculture				
agriculture and horticulture				
sustainability reporting				
waste management and recycling				
forestry				
water quality and water management				
apparel, textiles, and accessories	Other	N/A	Lifestyle	Lifestyle
baby, toddlers, and kids care				
fitness, sports, and recreation				
healthcare and nutrition				
hobbies, crafts, and toys				
home and garden				
personal and beauty care				
spirituality and consciousness				
business and professional services			Industry Solutions	Industry Solutions
computers and electronics				
conferences and events				
construction				
finance				
media and communication				
office supplies				
packaging				
real estate				
automotive and personal transportation			Travel and Hospitality	Travel and Hospitality
dining and entertainment				
food and beverage				
public transportation				
travel, tourism, and lodging				

Appendix AF: Study 2a – K-means Clustering Diagrams

Figure 1: First Primary Group, Elbow Method

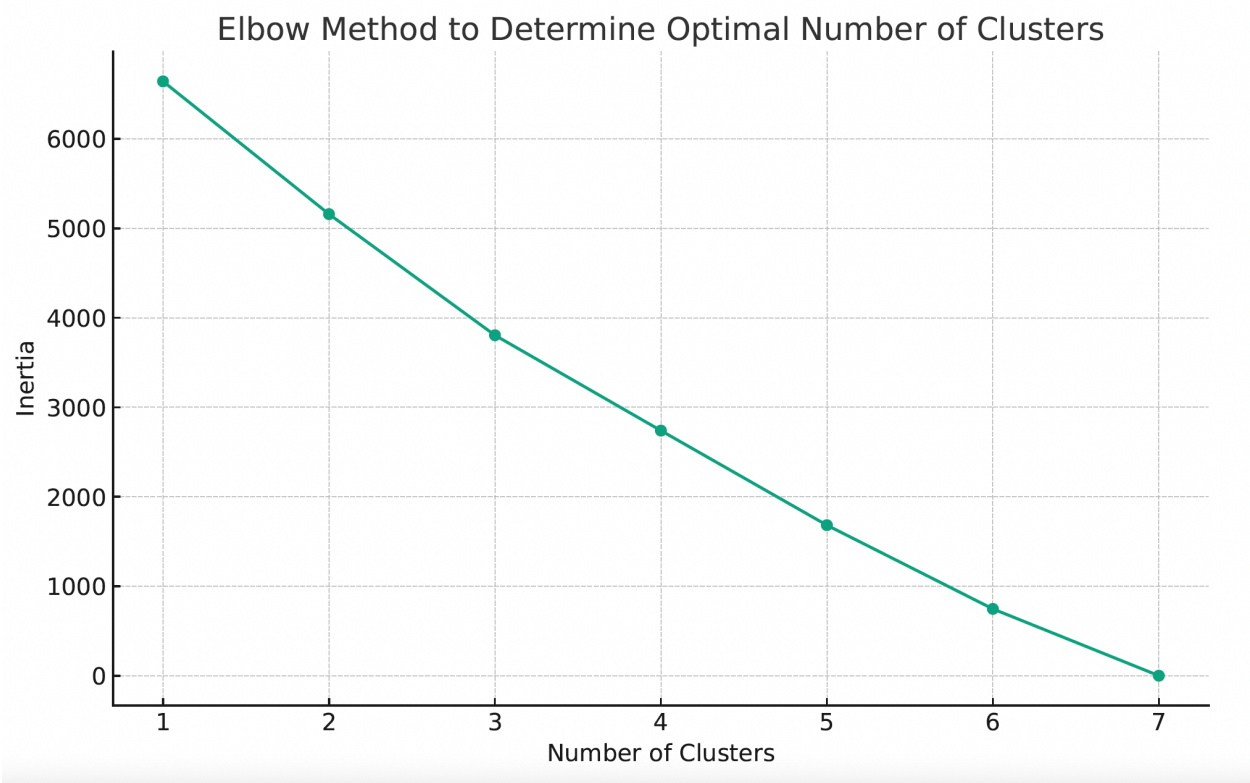


Figure 2: First Primary Group, Silhouette Scores

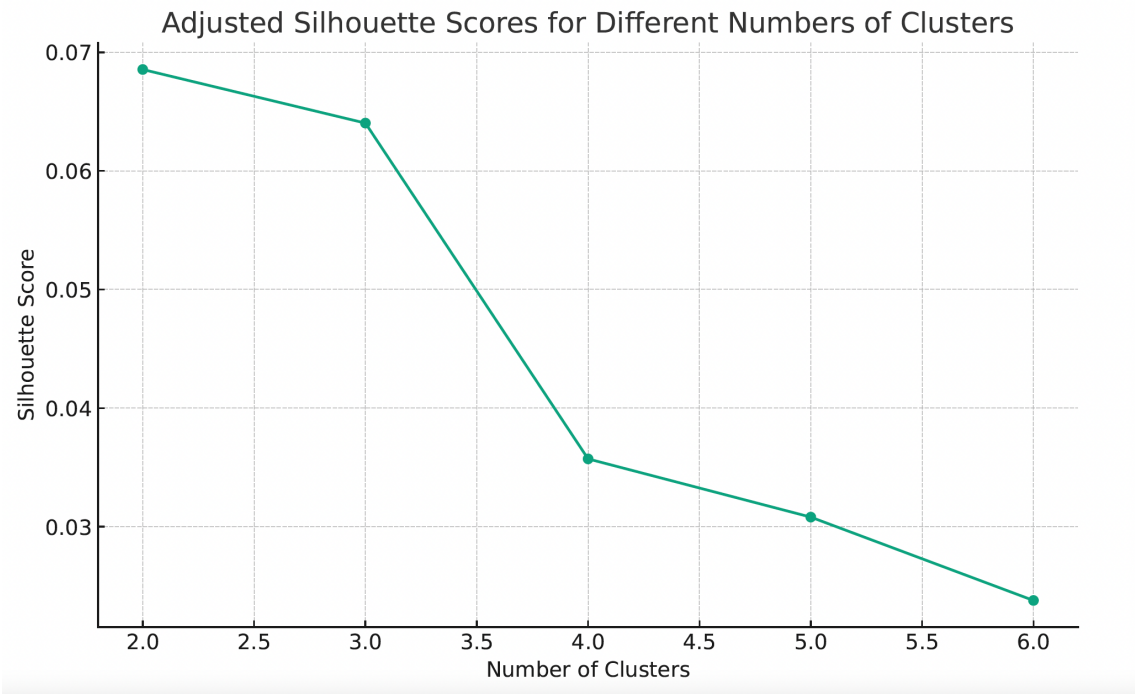


Figure 3: Second Primary Group, Elbow Method

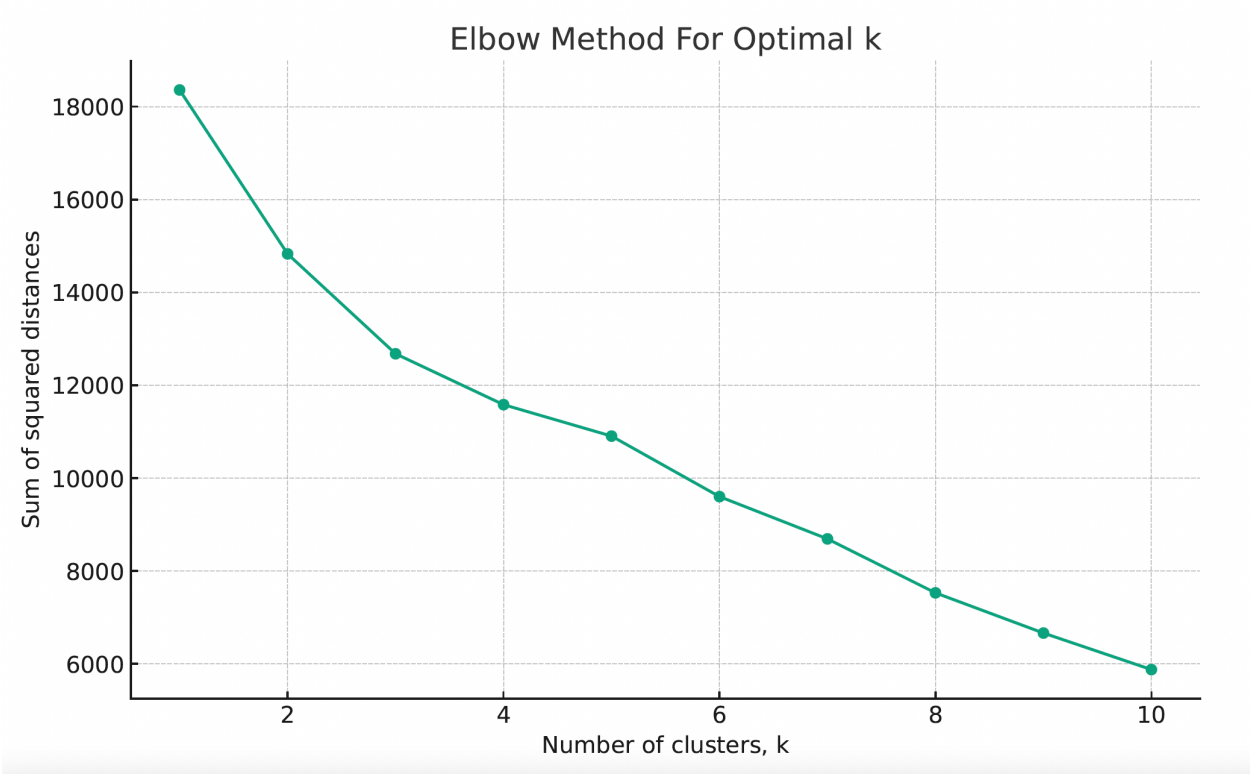
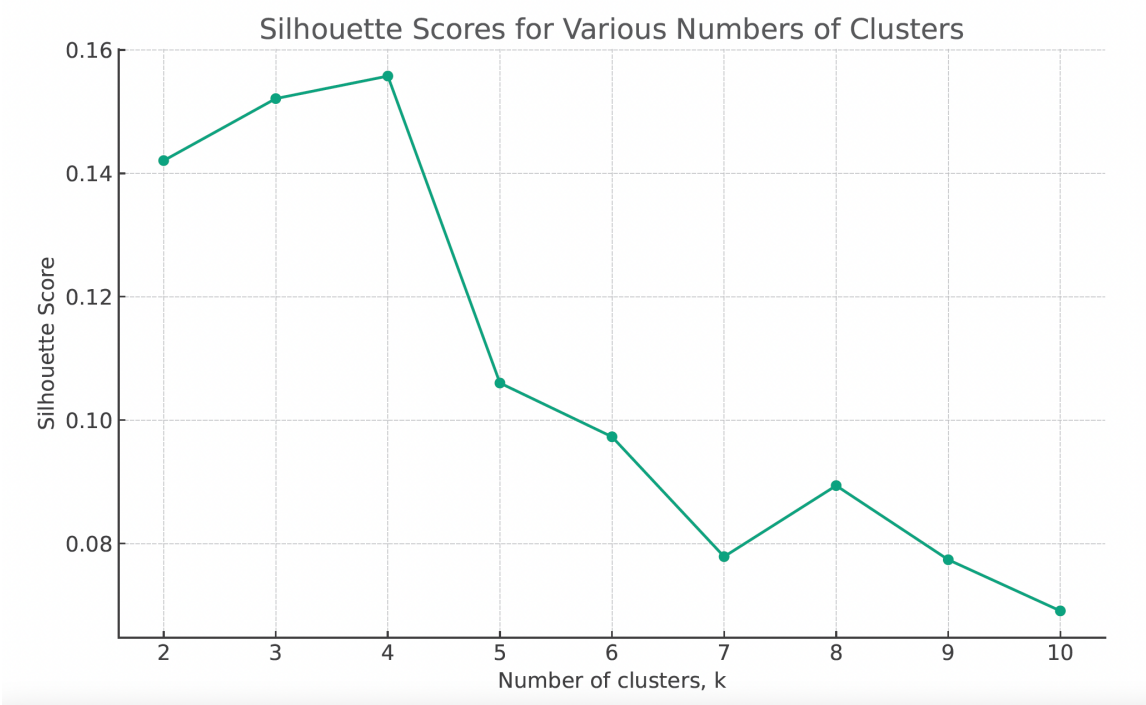


Figure 4: Second Primary Group, Silhouette Scores



References

- About us. (2024). Charity Navigator. Retrieved from <https://www.charitynavigator.org/about-us/>
- Agarwal, R., Sarkar, M., & Echambadi, R. (2002). The Conditioning Effect of Time on Firm Survival: An Industry Life Cycle Approach. *The Academy of Management Journal*, 45(5), 971–994.
- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Review*, 19(4), 645-670.
- Alexiou, K., & Wiggins, J. (2019). Measuring individual legitimacy perceptions: Scale development and validation. *Strategic Organization*, 17(4), 470-496.
- Almahdi, M. H., Alsayed, N., & Alabbas, A. (2022). In influencers we trust? A model of trust transfer in social media influencer marketing. In *Future of organizations and work after the 4th industrial revolution: The role of artificial intelligence, big data, automation, and robotics* (pp. 159-173). Cham: Springer International Publishing.
- Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, 76(5), 732.
- Ashforth, B. E., Schinoff, B. S., & Brickson, S. L. (2020). My company is friendly, “Mine’s a Rebel”: anthropomorphism and shifting organizational identity from “What” to “Who.” *Academy of Management Review*, 45(1), 29-57.
- Barnett, J. M. (2019). The Certification Paradox. In Contreras, J. L. (Ed.), *The Cambridge Handbook of Technical Standardization Law Further Intersections of Public and Private Law* (pp. 252). Cambridge University Press.

- Bettis, R. A. (1981). Performance differences in related and unrelated diversified firms. *Strategic management journal*, 2(4), 379-393.
- Bhaskarabhatla, A. (2016). The Moderating Role of Submarket Dynamics on the Product Customization–Firm Survival Relationship. *Organization Science*, 27(4), 1049-1064.
- Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International journal of computer applications*, 105(9).
- Bitektine, A. (2011). Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review*, 36(1), 151-179.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables* (Vol. 210). John Wiley & Sons.
- Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305.
- Carroll, G. R., & Wheaton, D. R. (2009). The organizational construction of authenticity: An examination of contemporary food and dining in the US. *Research in Organizational Behavior*, 29, 255-282.
- Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing*, 65(2), 81-93.
- Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation Without Trust?*. Russell Sage Foundation.
- Cook, K. (2015). Institutions, Trust, and Social Order. In E. Lawler, S. Thye, & J. Yoon (Eds.), *Order on the Edge of Chaos: Social Psychology and the Problem of Social Order* (pp. 125-144). Cambridge: Cambridge University Press.

- Coukyut, D. & Van Looy, A. (2021). An exploration of green business process maturity based on ecolabels. *Business Process Management Journal*, 27(7), 1999-2020.
- Council for Economic Education. (2020, February 5). *Survey of the States*.
<https://www.councilforeconed.org/survey-of-the-states-2020/>
- Coyle-Shapiro, J., & Kessler, I (2000). Consequences of the psychological contract for the employment relationship: A large scale survey. *Journal of Management Studies*, 37: 903–930.
- Darnall, N., Ji, H., & Vázquez-Brust, D.,A. (2018). Third-party certification, sponsorship, and consumers' ecolabel use: JBE. *Journal of Business Ethics*, 150(4), 953-969.
- Dart, R. (2004), The legitimacy of social enterprise. *Nonprofit Management and Leadership*, 14(4), 411–424.
- Davis, F. D., Schoorman, F. D., Mayer, R. C., & Tan, H. H. (2000). The trusted general manager and business unit performance: Empirical evidence of a competitive advantage. *Strategic Management Journal*, 21, 563-576.
- Deephouse, D. L., & Suchman, M. (2008). Legitimacy in organizational institutionalism. *The Sage Handbook of Organizational Institutionalism*, 49, 77.
- Dobrev, S. (2013). Organizational ecology. In E. Kessler (Ed.), *Encyclopedia of Management Theory* (Vol. 1, pp. 548-553). SAGE Publications, Ltd.
- Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer seller relationships. *Journal of Marketing*, 61, 35-51.
- Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one*, 18(3), e0279720.

- Dunlap, R. E., & McCright, A. M. (2011). Organized climate change denial. *The Oxford handbook of climate change and society*, 1, 144-160.
- Ecolabel Index. (2022). *All ecolabels*. Ecolabel index. <https://www.ecolabelindex.com/ecolabels/>
- Elbedweihy, A. M., Jayawardhena, C., Elsharnouby, M. H., & Elsharnouby, T. H. (2016). Customer relationship building: The role of brand attractiveness and consumer–brand identification. *Journal of Business Research*, 69(8), 2901-2910.
- Ennew, C., & Sekhon, H. (2007). Measuring trust in financial services: The trust index. *Consumer Policy Review*, 17(2), 62.
- Ferrin, D. L., Dirks, K. T., & Shah, P. P. (2006). Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of applied psychology*, 91(4), 870.
- Fiol, C. M., & Romanelli, E. (2012). Before identity: The emergence of new organizational forms. *Organization Science*, 23(3), 597-611.
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. In *Trends in Cognitive Sciences* (Vol. 11, Issue 2, pp. 77–83). <https://doi.org/10.1016/j.tics.2006.11.005>
- Franco, A., Sarkar, M., Agarwal R., & Echambadi, R. (2009). Swift and smart: The moderating effects of technological capabilities on the market pioneering-firm survival relationship. *Management Science*, 55(11), 1842–1860.
- Fombrun, C. J., Gardberg, N. A., & Sever, J. M. (2000). The Reputation QuotientSM: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management*, 7(4), 241-255.
- Fombrun, C., & Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, 33(2), 233-258.

- Frost, T., Stimpson, D. V., & Maughan, M. R. (1978). Some correlates of trust. *Journal of psychology*, 99(1), 103.
- Get certified. (2024). Vegan Action. <https://vegan.org/certification/>
- Gorton, M., Tocco, B., Yeh, C. H., & Hartmann, M. (2021). What determines consumers' use of eco-labels? Taking a close look at label trust. *Ecological Economics*, 189, 107173.
- Grant, R. M., Jammie, A. P., & Thomas, H. (1988). Diversity, Diversification, and Profitability among British Manufacturing Companies, 1972-84. *The academy of management journal*, 31(4), 771–801. <https://doi.org/10.2307/256338>
- Guidestar (2022). *Candid*. <https://www.guidestar.org/>
- Gulati, R., & Westphal, J. D. (1999). Cooperative or controlling? The effects of CEO-board relations and the content of interlocks on the formation of joint ventures. *Administrative Science Quarterly*, 44(3), 473-506.
- Ha, H. Y., & Perks, H. (2005). Effects of consumer perceptions of brand experience on the web: Brand familiarity, satisfaction and brand trust. *Journal of Consumer Behaviour: An International Research Review*, 4(6), 438-452.
- Hannan, M. T. (2005). Ecologies of organizations: Diversity and identity. *Journal of Economic Perspectives*, 19(1), 51-70.
- Hannan, M. T., Carroll, G. R., & Pólos, L. (2003). The organizational niche. *Sociological Theory*, 21(4), 309-340.
- Hill, G. M., & Simons, J. (1989). A study of the sport specialization on high school athletics. *Journal of Sport and Social Issues*, 13(1), 1-13.
- How we rate charities (2024). Charity Navigator. Retrieved from <https://www.charitynavigator.org/about-us/our-methodology/ratings/>

- Hsu, G. (2006). Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. *Administrative Science Quarterly*, 51(3), 420-450.
- Hsu, G., & Hannan, M. T. (2005). Identities, genres, and organizational forms. *Organization Science*, 16(5), 474-490.
- Hsu, G., Hannan, M. T., & Koçak, Ö. (2009). Multiple category memberships in markets: An integrative theory and two empirical tests. *American Sociological Review*, 74(1), 150-169.
- Intengine (2022). The marketplace & directory for eco-friendly, sustainable businesses. <https://intengine.com/>
- ISEAL Alliance. (2023). Our mission. <https://www.isealalliance.org/about-iseal/our-mission>
- Jakesch, M., Leder, H., & Forster, M. (2013). Image ambiguity and fluency. *PLoS one*, 8(9), e74084.
- Jensen, M. (2003). The role of network resources in market entry: Commercial banks' entry into investment banking, 1991-1997. *Administrative Science Quarterly*, 48, 466-497.
- Jensen, M., & Kim, B. K. (2014). Great, madama butterfly again! how robust market identity shapes opera repertoires. *Organization Science*, 25(1), 109-126.
- Jones, G. R., & Hill, C. W. (1988). Transaction cost analysis of strategy-structure choice. *Strategic management journal*, 9(2), 159-172.
- Kent, R. J., & Allen, C. T. (1994). Competitive interference effects in consumer memory for advertising: the role of brand familiarity. *Journal of Marketing*, 58(3), 97-105.
- Kerlinger, F. N. (1973). *Review of research in education*.

- Kervyn, N., Fiske, S. T., & Malone, C. (2012). Brands as intentional agents framework: How perceived intentions and ability can map brand perception. *Journal of Consumer Psychology*, 22(2), 166-176.
- Kovács, B., Carroll, G. R., & Lehman, D. W. (2014). Authenticity and consumer value ratings: Empirical tests from the restaurant domain. *Organization Science*, 25(2), 458-478.
- Lee, A. Y., & Aaker, J. L. (2004). Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion. *Journal of personality and social psychology*, 86(2), 205.
- Leong, F. T., & Austin, J. T. (2006). *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants*.
- Leung M.D., Sharkey A.J. (2014). Out of sight, out of mind? Evidence of perceptual factors in the multiple-category discount. *Organization science*, 25(1):171–184.
- Li, C., Turmunkh, U., & Wakker, P. P. (2019). Trust as a decision under ambiguity. *Experimental economics*, 22, 51-75.
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 293-334.
- Martin, F. (2021). Organizational Virtues and Organizational Anthropomorphism. *Journal of Business Ethics*, 1-17.
- Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), 123.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709-734.

- Market Research Future (2020). Green buildings market: Information by product (interiors, exteriors), application (residential, non-residential), and region - forecast till 2030.
- McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24-59.
- McKinsey & NielsenIQ. (2023). Consumers care about sustainability—and back it up with their wallets. Retrieved from <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/consumers-care-about-sustainability-and-back-it-up-with-their-wallets#/>
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The Journal of Strategic Information Systems*, 11(3-4), 297-323.
- McPherson, M. (1983). An ecology of affiliation. *American Sociological Review*, 519-532.
- Moon, J., Chadee, D., & Tikoo, S. (2008). Culture, product type, and price influences on consumer purchase intention to buy personalized products online. *Journal of Business Research*, 61(1), 31-39.
- Mordor Intelligence. (2023). Testing, Inspection, and Certification (TIC) Market Size & Share Analysis - Growth Trends & Forecasts (2023 - 2028).
- Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A., & Ferreira, M. E. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 3.
- NCCS Data Archive (2023). <https://nccs-data.urban.org/>
- Negro, G., Hannan, M. T., & Rao, H. (2010). Categorical contrast and audience appeal: Niche width and critical success in winemaking. *Industrial and Corporate Change*, 19(5), 1397-1425.

- Negro, G., Koçak, Ö., & Hsu, G. (2010). Research on categories in the sociology of organizations. In *Categories in Markets: Origins and Evolution*. Emerald Group Publishing Limited
- Nielsen. (2023). Nielsen study reveals: Sustainability-themed advertising is here to stay. Retrieved May 1, 2024, from <https://www.nielsen.com/news-center/2023/nielsen-study-reveals-sustainability-themed-advertising-is-here-to-stay/>
- Non-dairy pints. (2024). Ben & Jerry's. <https://www.benjerry.com/flavors/non-dairy>
- Office of Energy Efficiency & Renewable Energy. (2022). Energy Star®. <https://www.energy.gov/eere/buildings/energy-starr>
- Palich, L. E., Cardinal, L. B., & Miller, C. C. (2000). Curvilinearity in the diversification–performance linkage: an examination of over three decades of research. *Strategic management journal*, 21(2), 155-174.
- Pirson, M., & Malhotra, D. (2011). Foundations of organizational trust: What matters to different stakeholders?. *Organization Science*, 22(4), 1087-1104.
- Pirson, M., Martin, K., & Parmar, B. (2019). Public trust in business and its determinants. *Business & Society*, 58(1), 132-166.
- Pólos, L., Hannan, M. T., & Carroll, G. R. (2002). Foundations of a theory of social forms. *Industrial and Corporate Change*, 11(1), 85-115.
- Poppo, L., Zhou, K. Z., & Li, J. J. (2016). When can you trust "trust"? calculative trust, relational trust, and supplier performance. *Strategic Management Journal*, 37(4), 724-741.
- Porac, J. F., Thomas, H., Wilson, F., Paton, D., & Kanfer, A. (1995). Rivalry and the industry model of Scottish knitwear producers. *Administrative Science Quarterly*, 203-227.

- Patlolla, C. R. (2018). Understanding the concept of hierarchical clustering technique. *Towards data science*. Retrieved from <https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec>
- Prell, M., Zanini, M.T., Caldieraro, F. and Migueles, C. (2020), "Sustainability certifications and product preference", *Marketing Intelligence & Planning*, Vol. 38 No. 7, pp. 893-906.
<https://doi.org/10.1108/MIP-12-2019-0616>
- Rainforest Alliance. (2002). *Our approach*. <https://www.rainforest-alliance.org/approach/>
- Rao, H., Morrill, C., & Zald, M. N. (2000). Power plays: How social movements and collective action create new organizational forms. *Research in Organizational Behavior*, 22, 237-281.
- Ratings on Glassdoor. (2004). Glassdoor. Retrieved from
https://help.glassdoor.com/s/article/Ratings-on-Glassdoor?language=en_US
- Rating methodology guide. (2024). Charity Navigator. Retrieved from
https://docs.google.com/document/d/1zPMbhIIA-CaZrppzFus-TakBK5npSWpxFpBhUp_5OCY/edit
- Rindova, V. P., Pollock, T. G., & Hayward, M. L. (2006). Celebrity firms: The social construction of market popularity. *Academy of Management Review*, 31(1), 50-71.
- Rothenberg, E., Chanoine, H., Karr, R., & Martin, C. (2019). Companies face demand for third-party CSR certifications. *CFO.Com*.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393-404.
- Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational applied mathematics*, 20:53–65

- Rousseeuw, P. J., & Kaufman, L. (2009). *Finding groups in data: An introduction to cluster analysis*. Hoboken, New Jersey: John Wiley & Sons.
- Roy, S. K., & Shekhar, V. (2010). Dimensional hierarchy of trustworthiness of financial service providers. *International Journal of Bank Marketing*.
- Schmidt, J. B., & Spreng, R. A. (1996). A proposed model of external consumer information search. *Journal of the Academy of Marketing Science*, 24, 246-256.
- Semrush (2022). <https://www.semrush.com/>
- Sharkey, A., Kovacs, B., & Hsu, G. (2023). Expert critics, rankings, and review aggregators: The changing nature of intermediation and the rise of markets with multiple intermediaries. *Academy of Management Annals*, 17(1), 1-36.
- Skinner, W. (1974). The focused factory. *Harvard business review*, 114-121.
- Smith, A. (2002) *The Wealth of Nations*. Oxford, England: Bibliomania.com Ltd. [Web.] Retrieved from the Library of Congress, <https://lccn.loc.gov/2002564559>.
- Spector, P. E. (1992). *Summated rating scale construction: An introduction* (Vol. 82). Sage.
- Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, 13(1), 24.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571-610.
- Tang, J., Birrell, E., & Lerner, A. (2022). How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys. *arXiv preprint arXiv:2202.14036*.
- The Cornucopia Institute. (2019). *Certifier report and guide*. <https://www.cornucopia.org/research/certifier-report/>

United Nations (2023). *The 17 goals*. <https://sdgs.un.org/goals>

United States Environmental Protection Agency (2022). *Introduction to ecolabels and standards for greener products*. <https://www.epa.gov/greenerproducts/introduction-ecolabels-and-standards-greener-products>

Urban Institute. (2023). About the Urban Institute. <https://www.urban.org/about>

U.S. Census Bureau (2022), American Community Survey 5-year estimates subjects tables,
Retrieved from data.census.gov

U.S. Green Building Council (2022). USGBC country market brief: United States.
<https://www.usgbc.org/resources/country-market-brief>

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42, 35-67.

Vera, A., Petratos, P., & Torsten, O. S. (2018). A widening gap? Static and dynamic performance differences between specialist and general hospitals. *Health Care Management Science*, 21(1), 25-36.

Verhaal, J. C., Hoskins, J. D., & Lundmark, L. W. (2017). Little fish in a big pond: Legitimacy transfer, authenticity, and factors of peripheral firm entry and growth in the market center. *Strategic Management Journal*, 38(12), 2532-2552.

Villena, V. H., Choi, T. Y., & Revilla, E. (2019). Revisiting interorganizational trust: is more always better or could more be worse?. *Journal of Management*, 45(2), 752-785.

Wänke, M., Bohner, G., & Jurkowitsch, A. (1997). There are many reasons to drive a BMW: Does imagined ease of argument generation influence attitudes?. *Journal of consumer research*, 24(2), 170-177.

- Williams, M. (2001). In Whom We Trust: Group Membership as an Affective Context for Trust Development. *Academy of Management Review*, 26(3), 377–396.
- Woolley, K., & Fishbach, A. (2017). A recipe for friendship: Similar food consumption promotes trust and cooperation. *Journal of Consumer Psychology*, 27(1), 1-10.
- Yaniv, E. (2011). Construct clarity in theories of management and organization. *Academy of Management Review*, 36(3), 590-592.
- Yoo, B., & Donthu, N. (2001). Developing and validating a multidimensional consumer-based brand equity scale. *Journal of Business Research*, 52(1), 1-14.
- Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization Science*, 9(2), 141-159).
- Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American Journal of Sociology*, 104(5), 1398-1438.
- Zuckerman, E. W. (2000). Focusing the corporate product: Securities analysts and de-diversification. *Administrative Science Quarterly*, 45(3), 591-619.
- Zuckerman, E. W., Kim, T-Y., Ukanwa, K., & von Rittman, J. (2003). Robust identities or nonentities? Typecasting in the feature film labor market. *American Journal of Sociology*, 108, 1018–1074.