

Designing measurement for calibration: an instrument-design framework for fsQCA applications in tourism research

Noppadol Manosuthi

Chulalongkorn University, Bangkok, Thailand, and

Prarawan Senachai

*Faculty of Business Administration and Accountancy, Khon Kaen University,
Khon Kaen, Thailand*

Received 14 July 2026
Revised 1 September 2026
Accepted 9 September 2026

Abstract

Purpose – This article introduces “Measurement Before Calibration,” a design-oriented framework for developing survey instruments capable of yielding interpretable, theory-driven fuzzy-set memberships in fuzzy-set qualitative comparative analysis. It addresses the upstream methodological gap of whether measurement scales are built to support substantive calibration.

Design/methodology/approach – The framework specifies a five-step sequence: set definition, membership meaning definition, trade-off item design, pre-calibration validation and theory-driven calibration. It is demonstrated through a pilot study of 129 respondents choosing between destinations varying in price, convenience and sustainability credentials.

Findings – Using trade-off ladders rather than symbolic Likert agreement items successfully prevented ceiling effects in socially desirable constructs. Pre-calibration screening effectively identified exact-crossover ambiguity. Configurational analysis revealed a single conjunctural path requiring all four conditions jointly for sustainable destination choice.

Research limitations/implications – The pilot relies on a convenience sample of 129 respondents and evaluates a single sufficient configuration. Future research should implement between-subjects experimental designs to test configurational patterns in larger, representative samples.

Originality/value – While prior guidelines focus on calibration mechanics and reliability, this article provides an upstream instrument-design methodology ensuring measurement precedes and supports valid set-theoretic calibration.

Keywords fsQCA, Fuzzy sets, Calibration, Measurement, Tourism, Instrument design, Configurational analysis

Paper type Research article

1. Introduction

Fuzzy-set Qualitative Comparative Analysis (fsQCA) has become an increasingly common analytic choice in tourism, hospitality and marketing research. Between 2005 and 2024, some 352 documents in hospitality and tourism indexed in Web of Science and Scopus employed the method, with 124 articles appearing in 2023 alone and an exponential acceleration in output since 2015 (Eluwole, Lasisi, Parvez, & Cobanoglu, 2024). The Journal of Business Research alone published 464 fsQCA studies since 2007, 185 of them in 2016 (Geremew, Huang, & Hung, 2024) and critical reviews of QCA practice in business research have accompanied this

© Noppadol Manosuthi and Prarawan Senachai. Published in *ABAC Journal*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/>

Funding: This Research is funded by Thailand Science Research and Innovation Fund Chulalongkorn University (SOC_FF_68_208_3900_002).



ABAC Journal
Emerald Publishing Limited
e-ISSN: 2730-3543
p-ISSN: 0858-0855
DOI 10.1108/ABAC-07-2026-0293

growth (Wagemann, Buche, & Siewert, 2016). Within tourism and hospitality specifically, systematic reviews have identified six knowledge clusters, including business competitiveness, tourist behavior and psychoanalysis, entrepreneurship, tourism business ecosystems, sustainable tourism and service quality (Kumar, Sahoo, Ali, & Cobanoglu, 2024). The same trajectory is documented in entrepreneurship and innovation research, where fsQCA has become a standard analytic tool (Kraus, Ribeiro-Soriano, & Schüssler, 2018).

The method's reach extends across the tourism domain and into adjacent consumer contexts. Applications include ecotourism stakeholder management (Salman, Jaafar, Mohamad, & Khoshkam, 2023), hotel innovation performance (Palacios-Marqués, Roig-Dobón, & Comeig, 2017), asymmetric modeling of tourism insurance purchase and loyalty (Olya & Altinay, 2016), pro-environmental travel intentions (Olya & Akhshik, 2019), destination and community-based tourism marketing (Wattanawaraporn & Manosuthi, 2025), customer engagement with tourism-related services (Gligor & Bozkurt, 2020), knowledge sharing (Olan, Liu, Neaga, Chen, & Nakpodia, 2019), entrepreneurial opportunity evaluation (Andrade-Valbuena & Dos Santos, 2026) and the full range of tourist behavior and service-quality topics mapped by Kumar *et al.* (2024). The appeal of the method is evident. Many tourism outcomes, including destination choice, loyalty, sustainable behavior and satisfaction, plausibly arise not from single variables acting in isolation but from combinations of conditions that jointly produce an outcome in more than one way. A tourist may choose a certified sustainable destination because she trusts the certification, accepts a higher price and values environmental protection, while another arrives at the same choice largely through social pressure from travel companions. Configurational thinking gives researchers a vocabulary for conjunctural causation, equifinality and causal asymmetry, none of which is well captured by net-effects regression (cf. Fiss, 2011; Ragin, 2008).

The popularity of fsQCA has, however, outpaced methodological discipline in a specific way. A representative published workflow now runs as follows: Theory, Likert questionnaire construction, data collection, data import into fsQCA software, calibration threshold selection and interpretation of recipes. We argue that this sequence is procedurally complete, yet it omits the elements that make set-theoretic analysis set-theoretic rather than a relabeled correlational exercise. What does membership in the set of “high willingness to choose a sustainable destination” actually mean, substantively? Why is the crossover point set at 4 on a 7-point scale rather than 3 or 5? On which logical remainders does the parsimonious solution depend, and are the counterfactual assumptions behind them tenable? Because the software runs whether or not these questions have been asked, their absence tends to go unnoticed.

Existing methodological guidelines (Nikou, Mezei, Liguori, & El Tarabishy, 2024; Oana, Schneider, & Thomann, 2021; Rasoolimanesh, Valaei, & Rezaei, 2023; Thomann & Maggetti, 2020) provide detailed calibration procedures, including direct versus indirect calibration, percentile-based thresholds and transformation functions, and recommend conventional reliability and validity checks before fsQCA is undertaken. These guidelines, however, address calibration mechanics and measurement reliability, not the upstream question of whether the instrument itself was designed to produce scores to which a theory-driven anchor can meaningfully be applied. A scale with a Cronbach's alpha of 0.90 can still suffer from ceiling effects that make calibration effectively impossible.

Also, we argue that the recurring symptom of this underlying issue appears in measurement itself. Items such as “I support sustainable tourism” reliably draw responses clustered at 6–7 on a 7-point scale, the ceiling-effect pattern documented for Likert-type measures generally (Wang, Zhang, McArdle, & Salthouse, 2008), and when researchers find that a crossover point of 4 leaves almost no cases below it, the common response is to adjust the threshold until the distribution looks more usable. This is the wrong repair, since the problem is not that the threshold was chosen incorrectly but that the item was never capable of discriminating meaningful degrees of membership in the first place. No choice of anchor can rescue an instrument that measures symbolic agreement rather than a substantive, graded construct. Moving the crossover from 4 to 5 to 6 merely relabels the same distribution, because the

instrument itself never asked respondents to distinguish between endorsing sustainability in the abstract and actually paying more, tolerating inconvenience or changing a destination choice because of it.

Tourism research appears particularly vulnerable to this failure mode. Systematic reviews identify sustainable tourism, ecotourism, green hotels and residents' support for tourism as frequent fsQCA domains (Geremew *et al.*, 2024; Kumar *et al.*, 2024), constructs in which symbolic agreement can substitute fairly easily for gradated commitment, that is, for revealing the point at which one would accept a real cost, inconvenience or trade-off. Salman *et al.* (2023) illustrate the risk directly, since their ecotourism stakeholder-management instrument measured stakeholder interest and influence using 5-point agreement items, the ceiling-prone format flagged here.

Ragin (2008) already addressed the calibration half of this problem. His central methodological contribution, namely that fuzzy-set membership must be calibrated against external and theory-driven qualitative anchors rather than the sample's own distribution, is well established and this article does not seek to revise it. What Ragin's framework presupposes but does not itself provide is an instrument capable of producing scores to which an external anchor can meaningfully be applied. That is the gap this article addresses, namely not how to calibrate but how to design a measurement instrument such that calibration, once undertaken properly, is possible at all.

Recent empirical work lends this gap urgency. Krogslund, Choi, and Poertner (2015), replicating three published fsQCA studies and running Monte Carlo simulations, found the method "extraordinarily sensitive" to the calibration anchors chosen, to the point of exhibiting a "built-in confirmation bias" (the full finding is detailed in Section 3.1). If calibration decisions alone can produce this degree of sensitivity, the upstream instrument feeding those decisions, the items whose distributions determine which anchors are even plausible, matters rather more than has generally been acknowledged.

The article pursues three objectives, and one restructuring note follows from them. First, it explains the core logic of set-theoretic analysis, including membership, necessity, sufficiency, equifinality, asymmetry, consistency and coverage, using a fully worked and verifiable toy example computed in R. Because the purpose of that example is pedagogical, it is presented in full in Appendix F and summarized briefly in Section 2. Second, it catalogs the methodological problems that recur in tourism fsQCA applications, including mechanical calibration, ceiling effects, distribution-driven threshold-setting, missing counterfactual reasoning and unreported robustness checks, each grounded in published evidence (Section 3). Third, it proposes and illustrates a design-oriented framework, Measurement Before Calibration, that sequences instrument design ahead of calibration, demonstrated with a pilot study of 129 respondents choosing between destinations that differ in price, convenience and verified sustainability credentials (Sections 4–5). Section 6 discusses implications, and Section 7 concludes.

2. A primer on set-theoretic thinking

2.1 Set membership and calibration anchors

In fuzzy-set analysis, a case does not simply belong or fail to belong to a set. It belongs to some degree, expressed as a score between 0 (full nonmembership) and 1 (full membership), with 0.5 marking the crossover point of maximum ambiguity. Every condition and outcome in a QCA analysis is expressed this way from the outset. The foundations of fuzzy-set measurement are set out in Ragin (2000) and Smithson and Verkuilen (2006).

Consider a tourist deciding between two destinations. She is not simply "sustainable" or "not sustainable." She may lean strongly toward the certified option (near 1.0), be genuinely indifferent (0.50) or lean strongly toward the cheaper conventional option (near 0.0). A well-designed instrument captures this gradation, whereas a poorly designed one collapses most respondents into one or two scores, leaving little variation for calibration to work with.

Calibration is the process by which raw data are transformed into fuzzy-set membership scores. The standard procedure (Schneider & Wagemann, 2012, as cited in Rasoolimanesh *et al.*, 2023) specifies three qualitative anchors, namely full membership (near 1.0), the crossover point (0.50, maximum ambiguity) and full nonmembership (near 0.0). Two broad approaches exist. Direct calibration uses percentiles (commonly the 5th, 50th and 95th) or other statistical thresholds, whereas indirect calibration relies on expert judgment instead (Nikou *et al.*, 2024). The transformation may be linear or logistic, with logistic generally recommended for normally distributed data.

One clarification is owed to Ragin (2008) directly. The distinction between theory-driven and distribution-driven anchors, namely that anchors must come from substantive knowledge external to the sample and not from sample percentiles or standard deviations, is his contribution and not this article's, and Section 5.5 relies on it without modification. What this article contributes begins one step earlier, at the instrument-design decisions that determine whether a scale is even capable of supporting anchors of that kind.

2.2 Necessity and sufficiency

A condition X is *necessary* for outcome Y if Y cannot occur without X . In this sense, every instance of Y is accompanied by X , though X may occur without Y . In fuzzy-set terms, necessity corresponds to Y being consistently less than or equal to X , such that Y forms a subset of X . A condition X is *sufficient* for Y if the presence of X guarantees Y , even though Y may also arise from other conditions. In fuzzy-set terms, sufficiency corresponds to X being consistently less than or equal to Y , such that X forms a subset of Y . Most applied research deals not with single necessary or sufficient conditions but with combinations, namely conjunctions that are jointly sufficient without any one member being individually necessary, commonly called Insufficient but Non-redundant part of an Unnecessary but Sufficient (INUS) conditions. These are insufficient but necessary parts of a condition that is itself unnecessary but sufficient for the result (Mackie, 1965; Pappas & Woodside, 2021).

2.3 Consistency and coverage

Consistency and coverage are the fit statistics that make these subset claims gradable rather than all-or-nothing. Intuitively, consistency asks how often, or how consistently, membership in the proposed condition comes bundled with membership in the outcome, whereas coverage asks how much of the outcome's membership the condition actually accounts for.

Sufficiency consistency asks how much of the membership that cases have in X is matched by membership in Y .

$$\text{Cons}_{\text{suf}}(X \rightarrow Y) = \frac{\sum_i \min(X_i, Y_i)}{\sum_i X_i}$$

Sufficiency coverage asks the complementary question, namely of the membership that cases have in Y how much is explained by X .

$$\text{Cov}_{\text{suf}}(X \rightarrow Y) = \frac{\sum_i \min(X_i, Y_i)}{\sum_i Y_i}$$

Necessity consistency reverses which set is being explained, since it asks of the membership in Y how much is covered by X .

$$\text{Cons}_{\text{nec}}(Y \rightarrow X) = \frac{\sum_i \min(X_i, Y_i)}{\sum_i Y_i}$$

Necessity coverage, the relevance of necessity, uses the sufficiency-consistency denominator instead.

$$\text{Cov}_{\text{nec}}(Y \rightarrow X) = \frac{\sum_i \min(X_i, Y_i)}{\sum_i X_i}$$

This cross-over, in which necessity consistency shares a denominator with sufficiency coverage rather than sufficiency consistency, is among the most common sources of error in applied QCA write-ups (Nikou *et al.*, 2024; Ragin, 2008; Schneider & Wagemann, 2012). Appendix F computes both pairs on the same dataset so that the distinction is anchored to a concrete example.

Standard conventions (Cangialosi, 2023; Nikou *et al.*, 2024; Pappas & Woodside, 2021; Rasoolimanesh *et al.*, 2023; Schneider & Wagemann, 2010) set the sufficiency consistency threshold above 0.80 and the necessity consistency threshold above 0.90, with substantial coverage (above 0.20) and the relevance of necessity (RoN) also inspected. Proportional Reduction in Inconsistency (PRI) should generally be reported alongside raw consistency for sufficiency, since a low PRI indicates simultaneous subset relations in both Y and $\neg Y$.

2.4 Equifinality, causal asymmetry and calibration's sensitivity

Because sufficiency does not require uniqueness, more than one combination of conditions can be sufficient for the same outcome, much as a destination can be reached by more than one route (Pappas & Woodside, 2021; see Misangyi *et al.*, 2017, for the neo-configurational statement of this logic). A solution of the form $X_1 \cdot X_2 + X_3 \cdot X_4$ states that two distinct paths lead to Y . Equifinality is, in a sense, what configurational analysis exists for, since a method that returned only a single best-fitting path would have little need of set theory at all.

The conditions that explain Y 's presence need not mirror those that explain its absence. The fact that high certification trust and high environmental concern jointly produce sustainable choice does not imply that low trust and low concern are what prevent it, since some tourists may reject the certified destination for unrelated reasons. Analyzing $\neg Y$ is, in general, a separate analysis with its own truth table and solutions, rather than a logical by-product of analyzing Y . Both the toy example (Appendix F) and the pilot (Section 5) analyze Y and $\neg Y$ for this reason.

This asymmetry recurs across published fsQCA applications, from early strategic-management demonstrations (Greckhamer, Misangyi, Elms, & Lacey, 2008) onward. Hossain, Rahman, Quaddus, Hooi, and Olanrewaju (2021), comparing SEM and fsQCA on 198 open-government-data cases, found that configurations sufficient for low performance were not the mirror opposite of those sufficient for high performance. Galeazzo and Furlan (2018), studying lean-manufacturing bundles across 19 firms, report the same pattern, since no lean bundle met the necessity threshold, yet two equifinal configurations were sufficient for high financial performance while three distinct and nonmirrored configurations were sufficient for poor performance.

Finally, calibration is not a neutral technical step, since fsQCA results are highly sensitive to the anchors chosen. Krogslund *et al.* (2015) found the method “extraordinarily sensitive” to the three calibration anchors and susceptible to “built-in confirmation bias,” and Palacios-Marqués *et al.* (2017) obtained contradictory necessity results on identical hotel data under two

different calibration procedures. [Section 3.1](#) examines this evidence in detail, because it is the reason instrument design matters upstream of calibration.

3. The measurement gap in tourism fsQCA

3.1 Mechanical and distribution-driven calibration

Mechanical calibration occurs when thresholds are chosen only after inspecting the sample distribution, for example setting the crossover at the sample median simply because that is where the data cluster. This is calibration in name only, re-importing the sample-relative logic that fuzzy-set calibration exists to reject ([Ragin, 2008](#)).

Distribution-driven threshold repair is a related and more specific case, in which the crossover point is moved because the initial anchor produced too few cases on one side. Here, skew is treated as a calibration problem to be tuned away rather than evidence that the instrument needs redesigning.

The evidence that these choices matter is strong. [Krogslund et al. \(2015\)](#), replicating three published fsQCA studies and running Monte Carlo simulations, report three findings. First, fsQCA results were “extraordinarily sensitive” to the three calibration anchors, such that “slight changes in the values of the three fuzzy-set anchors” produced “strikingly different and often contradictory results.” Second, the method exhibited a “built-in confirmation bias,” making it “highly likely to identify causal combinations containing even randomly generated variables as sufficient.” Third, calibration itself was described as “half science, half art,” requiring judgments “underdetermined by prior theory.”

This severity is not confined to political science. [Palacios-Marqués et al. \(2017\)](#), analyzing 209 Spanish hotels, calibrated the same three conditions two different ways, namely direct calibration of averaged Likert scores versus percentile-based calibration of multiplied scores, and obtained contradictory necessity findings on the same 209 cases. The first found all three conditions necessary for innovation performance, and the second found none, leading the authors to conclude that fsQCA results “may show to be very different” depending on calibration procedure. This is [Krogslund et al.’s \(2015\)](#) sensitivity finding restated independently, in a hospitality dataset.

If calibration alone is this sensitive to anchor choice, the upstream instrument, whose response distributions determine which anchors are even plausible, matters considerably more than the literature has generally acknowledged. This article’s framework does not solve the calibration sensitivity problem as such. It addresses the condition that makes it worse, namely arriving at calibration with an instrument never designed to produce scores to which a theory-driven anchor could meaningfully be applied.

3.2 Missing counterfactual reasoning

Reporting a parsimonious or intermediate solution without stating the directional expectations behind it is exactly the gap [Appendix F](#) (Sections F.6–F.7) walks through by hand. Without a stated expectation, a reader cannot tell whether a dropped condition was excluded because theory suggests it does not matter or simply because the researcher never checked.

This looks like an analysis-stage mistake, but we argue that its root lies further upstream. A researcher can only state a directional expectation for a condition whose membership meaning was defined before data collection, rather than improvised at write-up. A condition measured as “I support sustainable tourism” generates a different directional expectation than one measured as “I would pay 20%+ more for a certified option,” even though both might be labeled “sustainability attitude” in a methods section.

3.3 Weak robustness reporting

Many published fsQCA applications report no sensitivity analysis at all, that is, no check of whether the solution changes under a different consistency threshold, PRI threshold, crossover

anchor, frequency cut-off or case-selection rule. In the absence of such checks, a single solution is presented as though it were the only one the data could plausibly support. [Krogslund et al. \(2015\)](#) recommend presenting results graphically across the full parameter space, not only the values that produce the preferred solution. The pilot reported later ([Section 5.9.5](#)) follows this recommendation, reporting robustness across a grid of incl.cut, PRI and n.cut values, including the thresholds at which no solution survives, making the fragility or robustness of a solution directly inspectable rather than merely asserted.

4. Measurement before calibration as a five-step framework

The framework sequences instrument design ahead of calibration in five steps. The first two steps are conceptual, covering what the set is and what membership means. The third step is constructional, covering how items are built. The fourth step is diagnostic, covering whether the instrument worked. Only then, in the fifth step, is calibration undertaken.

4.1 Step 1. define the set

The first step is to name the set precisely enough that its boundary functions as a claim rather than a mood.

“High Willingness to Choose a Higher-Cost Certified Sustainable Destination” is a set with a boundary of this kind, since a respondent either crosses the threshold of accepting a real price premium and inconvenience for verified sustainability or does not. “Positive Attitude toward Sustainability” is not, because it invites the symbolic-agreement response pattern discussed in [Section 1.3](#), given that almost everyone can honestly claim a positive attitude while differing enormously in what they would actually do.

The distinction is between a set whose membership can be empirically determined and a mood whose endorsement is close to universal. A boundary that everyone clears is not, in any meaningful sense, a boundary at all.

4.2 Step 2. define membership meaning

Before a single item is written, it is worth stating explicitly what full membership, crossover and full nonmembership represent in substantive terms. An example is drawn from the pilot ([Section 5.3](#)), in which full membership means choosing Destination B even when its price premium and inconvenience are pushed well past the stated terms (e.g. 30%+ more, noticeably less convenient), crossover means genuine indifference between A and B, since the sustainability features exactly offset the stated trade-off, and full nonmembership means choosing A even after B’s premium and inconvenience are removed entirely.

This is not merely a formality. It is what makes a directional expectation possible later ([Sections F.7 and 5.9](#)), and it is, in a real sense, what a necessity or sufficiency claim is actually about. A researcher who cannot say what crossover means for a construct is not yet ready to calibrate it.

4.3 Step 3. design measurement for gradation

The third step is to build items around trade-offs rather than agreement. A trade-off forces a respondent to reveal how much of something else they would give up, producing a gradation that a symbolic-agreement item cannot. The willingness-to-pay literature has long shown that elicitation format materially affects the estimates obtained ([Miller, Hofstetter, Krohmer, & Zhang, 2011](#); [Voelckner, 2006](#)), and the forced-choice trade-off ladders used here follow the stated-choice tradition of discrete-choice measurement ([Louviere, Hensher, & Swait, 2000](#)).

Three design principles follow. First, trade-offs come before agreement, which applies when the construct is a choice under constraint, such as willingness to pay, acceptance of inconvenience or deference to the preferences of others, and the primary item should present the constraint directly and ask the respondent to reveal the point at which their choice would

change, whereas agreement items serve as a fallback rather than a default. Second, behavioral anchoring applies where scales cannot be avoided, since for a disposition such as environmental concern, which has no single discrete trade-off, items should be anchored in actual past behavior rather than identity or values, for example “I have changed a travel plan. . . to reduce my environmental impact” rather than “I support environmental protection”. Third, domain separation applies between conditions and outcomes, since a condition measured in the same purchase domain as the outcome risks circularity, in which willingness-to-pay simply predicts willingness-to-pay. Price Premium Acceptance, as a condition, is accordingly measured in an unrelated everyday-purchase domain, while the outcome is measured in the destination-choice scenario.

The toy example in [Appendix F](#) calibrated its conditions directly, since its purpose was to teach analysis mechanics rather than instrument design. The pilot in [Section 5](#) demonstrates what Step 3 looks like when an instrument must be designed from scratch, using a forced-choice scenario, mirrored trade-off ladders and behaviorally anchored scales.

4.4 Step 4. validate before calibration

Before calibration, it is worth inspecting the data for the problems cataloged in [Section 3](#), diagnostically rather than as post-hoc rationalization. Three checks seem essential. First, distribution inspection asks whether responses are clustered at ceiling or floor. If so, the instrument has failed to produce gradation, no calibration anchor will recover what was never captured and the response is to return to Step 3. Second, exact crossover cases are counted, since for any condition whose raw score lands precisely on the calibration crossover, the resulting fuzzy score will be exactly 0.50 and that case will be excluded from the truth table silently and often unnoticed. Counting exact crossovers before calibration surfaces this problem while there is still time to adjust the anchor or the instrument.

Where instrument redesign is not an option, for instance when reanalyzing an existing dataset, the field’s standard fallback is a small and arbitrary offset. [Galeazzo and Furlan \(2018\)](#) added 0.001 to every membership score to move exact-0.5 cases off the crossover, following [Fiss \(2011\)](#), [Greckhamer \(2011\)](#) and [Ragin \(2008\)](#), and [Cangialosi \(2023\)](#) recommends the same fix at the tutorial level, suggesting it is common enough to have become routine practice rather than an occasional expedient. This is a legitimate analysis-stage patch, but a patch is what it remains, since it resolves the symptom after the fact rather than the underlying cause. The instrument-design response in Step 3, together with the pre-registered anchor check in [Appendix A](#)’s second precondition, is intended to make this kind of patch unnecessary rather than routine.

The third check concerns ladder monotonicity and comprehension, since for trade-off ladders responses should be verified as monotonic and comprehension checks should be verified as passed.

This step functions as a quality-assurance gate. If the instrument fails it, calibration probably should not proceed, and the response is to redesign the measure rather than to move the anchor until the truth table looks balanced. The pilot ([Section 5.8](#)) demonstrates this gate in operation, since the PRICE condition produces 22 exact crossover cases, surfaced there as an instrument-design problem rather than a calibration problem to be quietly tuned away.

4.5 Step 5. calibrate using theory-driven anchors ([Ragin, 2008](#))

Once Steps 1–4 have produced an instrument capable of discriminating meaningful degrees of membership, calibration can proceed much as Ragin describes. Anchors for full membership, crossover and full nonmembership are set from external theoretical or substantive knowledge, never from the sample’s own distribution. [Section 3.1](#)’s mechanical and distribution-driven calibration mistakes are, within this framework, best understood as symptoms of skipping Steps 1–4 and arriving at Step 5 with an instrument that was never going to support a theory-driven anchor.

The framework extends, rather than replaces, existing calibration guidance ([Cangialosi, 2023](#); [Nikou et al., 2024](#); [Pappas & Woodside, 2021](#); [Rasoolimanesh et al., 2023](#)) by adding the instrument-design preconditions those guides assume but do not themselves provide. It also addresses the sensitivity problem [Krogslund et al. \(2015\)](#) identify, by ensuring the instrument-feeding calibration is itself capable of producing scores to which a stable and theory-driven anchor can be applied.

5. Pilot illustration and the application of the framework

5.1 Purpose and scope

[Sections 1–4](#) made the argument at the level of principle. Theory-driven calibration ([Ragin, 2008](#)) presupposes an instrument capable of producing scores to which an external anchor can meaningfully be applied, and the five framework steps describe how such an instrument might be built. This section applies the framework to itself, within the tourism decision named at the outset, namely tourists choosing between destinations that differ in price, convenience and verified sustainability credentials.

The pilot was not designed to test a substantive theory of sustainable tourism choice, and the sampling plan described below could not really support such a claim. Its purpose is narrower. It shows what an instrument looks like when Steps 1–3 are actually followed rather than skipped, and it produces a real dataset on which the analysis workflow rehearsed in [Appendix F](#), including truth table, necessity, sufficiency, three solutions and robustness, can be run on genuine cases. A larger and theory-testing extension of this design belongs to a separate empirical paper, sketched briefly in [Section 6.9](#).

5.2 Participants, inclusion and exclusion criteria

The analyzed dataset contains 129 complete cases. Because this is an instrument pilot rather than a systematic review, the inclusion and exclusion criteria below apply to *respondents*, not to a corpus of studies.

For the inclusion criteria, respondents were required to meet all four of the following conditions, namely being at least 18 years old, being able to read and answer the survey in English, having traveled within the past 12 months or intending to travel within the next 12 months, and personally participating in choosing where they travel. The last two criteria matter directly, since the outcome measure asks respondents to place themselves inside a destination-choice scenario, and a respondent who does not make this kind of decision in real life would have little basis for the trade-off judgments the items ask for. Additionally, respondents were excluded from the analysis if they met any of the following criteria: failing to complete the survey, failing the attention check or either comprehension check, reading the vignette faster than the flagged minimum reading time, exhibiting a straightline response pattern or showing an excessive number of window-focus losses indicative of multitasking. [Section 5.8](#) details the application of these criteria and the corresponding data-quality checks.

The sample is treated throughout as appropriate for an instrument and analysis pilot, not as a basis for population-level inference. A convenience sample is a reasonably accepted limitation for an instrument pilot. It would not be accepted for the theory-testing study sketched in [Section 6.9](#), which is why the two are kept separate here.

5.3 Scenario and outcome set definition

Imagine you are planning a short vacation and choosing between two beach destinations you have not visited before.

Destination A costs less and is easier to reach directly from your home city. It provides no information about sustainability practices.

Destination B costs about 20% more and involves a slightly longer or less convenient trip. It carries independently verified sustainability certification, and part of its tourism revenue funds local community programs.

In every other respect (weather, safety, general appeal), the two destinations are equivalent.

Holding weather, safety and general appeal equal makes this a vignette about the sustainability trade-off specifically, rather than about destinations in general. Vignette designs of this kind follow the factorial-survey tradition, in which scenario features are held constant so that the manipulated contrast carries the interpretive weight (Auspurg & Hinz, 2015).

Applying Step 1, the outcome set is “High Willingness to Choose a Higher-Cost Certified Sustainable Destination,” a set with a genuine boundary rather than a mood nearly everyone would endorse. For Step 2, full membership means choosing B even when its price premium and inconvenience are pushed well past the stated terms (30%+ more, noticeably less convenient). Crossover means genuine indifference between A and B. Full nonmembership means choosing A even after B’s premium and inconvenience are removed entirely.

Naming the set this way, “willingness to choose” rather than “positive attitude toward,” is what Step 1 means by a boundary that functions as a claim rather than a mood, and it is why the instrument below is built as a two-stage trade-off task rather than a single agreement item.

5.4 The two-stage design of outcome measurement

Stage 1, the forced-choice stage, asks “If you had to choose one destination for this trip, which would you choose?” with the response options A, B and Undecided. This fixes which side of the 0.5 crossover a respondent falls on.

Stage 2, the degree stage, branches by the Stage 1 response. This fixes distance from the crossover. It is written as a trade-off ladder rather than a Likert item, with each rung naming concrete terms rather than asking for agreement, which applies Step 3 directly and is the mechanism that avoids the ceiling-effect problem discussed in Section 1.3. An item such as “I would prefer B” tends to collapse every B-leaning respondent onto the same point, whereas a ladder forces each respondent to reveal the specific terms at which their choice would flip.

The seven-rung ladder runs from weakest to strongest commitment. The weakest rung for a B-chooser, for example, reads “I would choose B only if it cost the same and were equally convenient as A,” and the strongest reads “. . . even if it cost 40%+ more and was significantly less convenient.” A mirrored ladder applies for A-choosers. The full seven-rung wording for both directions, and the undecided soft-lean follow-up used to resolve what would otherwise be an exact-crossover dropout problem, are reported in [Appendix B](#).

Stage 2 deliberately varies both price and convenience together, rather than price alone, so that the outcome measure captures overall support for the destination rather than price sensitivity specifically, a distinction that matters once price sensitivity reappears as its own condition ([Section 5.6.2](#)).

5.5 Calibration and anchor specification

[Table 1](#) converts each ladder rung into an illustrative fuzzy score. The anchors at the strongest and weakest rungs and the undecided branches are fixed by the substantive meaning assigned in [Section 5.3](#), while intermediate rungs are filled in by mirroring each B rung’s distance from 0.5 onto the corresponding A rung, which is why the two columns compress by roughly the same amount as they approach their respective poles.

For transparency and reproducibility, [Table 2](#) summarizes the calibration anchor points, namely full membership, crossover and full nonmembership, for the outcome and for each of the four conditions, as specified before data collection. This mapping should be read as

Table 1. Illustrative calibration of the outcome measure

Response	Score	Response	Score
B7	0.95	A7	0.05
B6	0.90	A6	0.10
B5	0.80	A5	0.20
B4	0.70	A4	0.30
B3	0.65	A3	0.35
B2	0.60	A2	0.40
B1	0.55	A1	0.45
Undecided, somewhat leaning B	0.60	Undecided, somewhat leaning A	0.40
Undecided, barely leaning B	0.55	Undecided, barely leaning A	0.45
Undecided, truly no lean	0.50		

Note(s): Referenced in [Section 5.5](#). Maps each of the seven outcome-ladder rungs ([Appendix B.1](#)), plus the undecided soft-lean branches, onto an illustrative fuzzy membership score. The same mirroring pattern is applied to the CERT (X/Y) and NORM (EXPECT/OWN) ladders

Table 2. Calibration anchor points for the outcome and all conditions

Set	Full membership (1.0)	Crossover (0.50)	Full nonmembership (0.0)
Y, high willingness to choose B	Choosing B even at a 30–40%+ premium and noticeably or significantly less convenient (rungs B6–B7, score 0.90–0.95)	Genuine indifference between A and B, undecided with truly no lean (score 0.50)	Choosing A even when B’s premium and inconvenience are removed entirely (rung A7, score 0.05)
CERT	Independently audited certification worth a 30%+ premium, where the audit itself is what matters (rung X7, score 0.95)	Audited and self-certified claims trusted equally, “no difference” with truly no lean (score 0.50)	No differentiation or outright distrust of the “independent” claim, independent audit of no value regardless of price (rung Y7, score 0.05)
PRICE	Routinely paying a 30%+ premium for certified or ethical everyday purchases (rung 7, score 0.90)	Paying up to about 15% more for certified or ethical everyday purchases (rung 4, score 0.50)	Price is decisive regardless of domain, nothing extra (rung 1, score 0.10)
ENV	Environmental protection a standing priority actually acted upon, including in travel plans (mean of the three behaviorally anchored items = 7, score 0.90)	General concern without clear behavioral commitment (item mean = 4, the scale midpoint, score 0.50)	Not a live concern (item mean = 1, score 0.10)
NORM	Choice bends toward important others’ expectations even against strong personal preference (rung 7 of the deference ladder, score 0.95)	Personal preference and social expectation weighted about equally (“depends”, truly torn, score 0.50)	Personal preference dominates regardless of social input (rung 7 of the own-preference ladder, score 0.05)

Note(s): Referenced in [Section 5.5](#). Anchors were specified from substantive meaning (Step 2) before data collection. Rung-to-score mappings for the ladder-based sets follow the mirroring pattern shown in [Table 1](#). PRICE maps its seven rungs linearly onto 0.10–0.90, so that rung 4 falls on the crossover. ENV maps the three-item mean linearly onto 0.10 (item mean 1), 0.50 (item mean 4) and 0.90 (item mean 7). All values were reproduced exactly from the field data by the analysis script ([Appendix C](#))

illustrative rather than as a finding in its own right. Consistent with Step 5, the final anchors would need to be justified on substantive grounds, for example why a 30%+ premium under noticeably reduced convenience constitutes full membership rather than merely strong membership, and pressure-tested by the robustness stage rather than simply adopted because they produce a usable distribution.

5.6 The design and justification of the conditions

The pilot fields four conditions, namely Certification Credibility (CERT), Price Premium Acceptance (PRICE), Environmental Concern (ENV) and Subjective Norm (NORM), reserving Community Benefit for the follow-up study (Section 6.9). Four considerations drove this choice. First, respondent burden matters, since CERT and NORM are each full two-stage tasks, and a tired respondent by the fourth vignette would likely produce the same flat and low-variance pattern Section 1.3 warns against, only caused by fatigue rather than symbolic agreement. Second, truth-table tractability matters, since four conditions generate 16 logical corners rather than 32, which is more tractable for a pilot of 80–120 cases. Third, conceptual coverage without redundancy matters, since the four conditions span trust in a specific sustainability signal, general price sensitivity measured outside the travel domain, a standing environmental disposition and susceptibility to social expectation. Fourth, continuity with the planned follow-up matters, since the follow-up will manipulate Certification, Community Benefit, and Price Premium experimentally rather than measure them correlationally.

Each condition states its membership meaning before its item, applying Steps 2 and 3 to the pilot instrument itself. CERT and NORM use the same two-stage architecture as the outcome (Section 5.4), PRICE is a single-stage ladder in a different purchase domain and ENV is a behaviorally anchored scale, following Step 3’s stated fallback. Full item wording for all four conditions is reported in Appendix B, and Table 3 summarizes the design.

5.6.1 Certification credibility (CERT). CERT is built around the specific distinction between independently audited and self-certified sustainability claims, rather than certification in the abstract, since a signal’s credibility appears to hinge on exactly this distinction, the same distinction the greenwashing literature draws between independent verification and self-declaration (De Freitas Netto, Sobral, Ribeiro, & Soares, 2020). Experimental evidence from community-based tourism marketing found that greenwashing suspicion sharply reduced engagement with a promotion bundling community and commercial claims when a moral identity was salient, but had little effect under a deal-seeking identity (Wattanawaraporn & Manosuthi, 2025), suggesting that trust in a sustainability claim depends on how credible the underlying signal is judged to be. Full membership means independently-audited certification is seen as clearly more trustworthy

Table 3. Summary of pilot conditions

Condition	Crossover meaning	Item format	Raw code examples
CERT	Trusts audited and self-certified claims equally	Two-stage, forced choice plus mirrored 7-rung ladder	X6, Y2, LEAN_X_BARELY, NO_LEAN
PRICE	Pays up to about 15% more for certified or ethical everyday purchases	Single-stage 7-rung ladder	1–7
ENV	General concern, no clear behavioral commitment	Three-item behaviorally anchored scale	1–7 per item
NORM	Personal preference and social expectation equally weighted	Two-stage, forced choice plus mirrored 7-rung ladder	EXPECT5, OWN3, LEAN_OWN_SOMEWHAT, NO_LEAN

Note(s): Referenced in Section 5.6. Full item wording for all four conditions is reported in Appendix B

than self-certification and that belief changes willingness to pay. Crossover means trusting both equally. Full nonmembership means no differentiation, or outright distrust of the “independent” claim itself. A two-stage forced-choice-plus-ladder task, structurally identical to the outcome measure, contrasts an independently audited option against a self-certified one. Full wording is reported in [Appendix B](#).

5.6.2 Price premium acceptance (PRICE). Full membership means routinely paying a 30%+ premium for certified or ethical options across everyday purchases. Crossover means paying up to about 15% more, the ladder’s midpoint rung, which the calibration anchors fix at 0.50 ([Table 2](#)). Full nonmembership means price is decisive regardless of domain. To capture this construct, a single-stage, seven-rung ladder asked respondents to state the maximum premium they would typically pay for certified or ethical options in nontravel everyday purchases (e.g. coffee, clothing, groceries). This item was deliberately contextualized outside the travel domain to avoid directly restating the outcome measure. Full item wording is provided in [Appendix B](#). However, as noted in [Section 5.9.6](#), this single-stage format emerged as the sole clear measurement weakness in the pilot, prompting the transition of PRICE to a two-stage architecture in the subsequent iteration.

5.6.3 Environmental concern (ENV). Full membership indicates that environmental protection is an active priority translated into behavior, rather than mere verbal endorsement. Crossover reflects general concern lacking consistent behavioral commitment, while full nonmembership signifies that environmental protection is not a salient consideration. Because this condition represents an underlying disposition rather than a discrete choice, forcing a binary trade-off would have been artificial. Instead, a conventional three-item battery served as the framework’s designated fallback option. To enable reliability assessment, items were behaviorally anchored rather than symbolically phrased (e.g. “I have changed a travel plan, paid more, or taken a less convenient route specifically to reduce my environmental impact”). Although agreement-based items risk a ceiling effect, this compromise remains effective in practice. Anchoring item content in concrete past behaviors and self-directed effort, rather than abstract identity or values, prevents the scale from devolving into generic statements like “I support sustainable tourism.” The complete item wording is provided in [Appendix B](#).

5.6.4 Subjective norm (NORM). Full membership means actual choice bends toward what important others expect, even against personal inclination. Crossover means personal preference and social expectation are weighted about equally. Full nonmembership means personal preference dominates regardless of social input. Next, a two-stage forced-choice-plus-ladder task is phrased without reference to Destination A or B, a deliberate change driven by the administration-order design ([Section 5.7](#)), since an item that only makes sense once the vignette has been presented cannot sensibly be dropped into a block that sometimes runs before it. Full wording is reported in [Appendix B](#).

5.7 Administration order and field implementation

Fielding raised a problem the design sections above do not address, namely the order in which the vignette and outcome and the four conditions are administered. If conditions were always measured after the destination choice, responses could partly reflect choice-supportive rationalization rather than a pre-existing disposition, inflating exactly the condition-to-outcome relationships the analysis depends on, the standard common-method-variance concern in survey research ([Podsakoff, MacKenzie, Lee, & Podsakoff, 2003](#)).

The fielded instrument addresses this by randomizing, per respondent, whether the four conditions are answered before or after the vignette and outcome block, balanced close to 50/50, with the assigned arm logged (`order_arm`) so that any order effect can be checked directly in the robustness stage. This is also why NORM ([Section 5.6.4](#)) was reworded to drop its reference to Destination A/B, since the other three conditions were already self-contained in this respect. The four condition blocks are also independently shuffled per respondent, with that order logged as `condition_order`.

The instrument is implemented as a survey web application. Every two-stage item logs a raw response code rather than a calibrated score, since calibration is deliberately left to the analysis stage (Section 5.5). Data-quality instrumentation is captured automatically, including time per screen, a minimum-reading-time flag on the vignette, a tab and window focus-loss count, a straightline-response flag and pass and fail flags for the attention check and two comprehension checks, feeding directly into the screening step (Section 5.8).

5.8 Data screening, quality checks and membership distributions

Applying Step 4. Before calibration, the data were inspected for the problems Section 3 catalogs, not as post-hoc rationalization but as a genuine diagnostic gate. This section combines two activities, namely the formal application of the inclusion and exclusion criteria of Section 5.2 and the distribution diagnostics of Step 4.

The automated instrumentation described in Section 5.7 was applied as the first stage of screening. Attention checks and response-consistency indicators are the standard means of identifying careless or satisficing responding in online surveys (Krosnick, 1991; Meade & Craig, 2012). Table 4 reports the data-quality summary, including the attention-check pass rate, comprehension-check pass rates, vignette reading-time violations, focus-loss counts, straightline flags and the number of cases removed by each exclusion criterion, together with the final retained sample. The attention check was passed by 123 of the 129 respondents (95.3%). The six failures were retained in the reported analysis for two reasons. They passed the remaining checks, and the results are invariant to their exclusion, since re-running the analysis at $N = 123$ leaves the Y solution unchanged (the conjunction of CERT, PRICE, ENV and NORM, $\text{incl} = 0.902$) and the $\neg Y$ solution substantively unchanged (Section 5.9.5;

Table 4. Data-quality screening report

Criterion	Criterion type	Pass/detected	Cases removed
Survey completion	Inclusion (required)	129 complete	0
Minimum age (≥ 18)	Inclusion (required)	129	0
English proficiency	Inclusion (required)	129	0
Travel within past/next 12 months	Inclusion (required)	129	0
Personal participation in destination choice	Inclusion (required)	129	0
Attention check (ATTN1 = 3)	Exclusion	123 pass (95.3%)/6 fail (4.7%)	0
Comprehension check 1 (COMP1)	Exclusion	n/a	n/a
Comprehension check 2 (COMP2)	Exclusion	n/a	n/a
Vignette minimum reading time (≥ 20 s)	Exclusion	n/a	n/a
Straightline response pattern	Exclusion	n/a	n/a
Excessive focus loss (tab/window switches)	Exclusion	n/a	n/a
Retained sample			129

Note(s): Referenced in Section 5.8. Criterion definitions follow Section 5.2, and the automated instrumentation is described in Section 5.7. The six attention-check failures were retained in the reported analysis for two reasons. They passed the remaining checks, and the results are invariant to their exclusion, since the Y solution at $N = 123$ is identical, namely the conjunction of CERT, PRICE, ENV and NORM, $\text{incl} 0.902$ (Section 5.9.5). Rows marked “n/a” depend on the full field log, which records reading time, focus loss and comprehension flags. The excerpted dataset made available for this revision contains only the attention flag, so those counts will be reported when the complete log is assembled

[Appendix E](#)). Reading-time, focus-loss and comprehension counts from the full field log will be appended here when the complete log is assembled, since the excerpted dataset used for this revision records only the attention flag.

Table 5 reports the membership-score distributions for the outcome and four conditions ($N = 129$), including the count of exact-0.5 crossover cases per set.

These distributions suggest the instrument avoids the most severe form of ceiling effect. No set has a mean or median above 0.60. The outcome (Y) has cases distributed across essentially the full range, 0.05–0.95, with a mean of 0.512, despite measuring a socially desirable construct. It seems reasonable to attribute this variance to Step 3, namely trade-off ladders rather than symbolic agreement items. Exact crossover cases are fairly frequent, especially for PRICE (22 cases, 17% of the sample), which is both a finding and a diagnostic. PRICE's single-stage ladder produces noticeably more crossover clustering than the two-stage items (CERT = 11, NORM = 9), suggesting the single-stage design should be revised to a two-stage format in the next iteration, a concrete finding that would likely have gone unnoticed had calibration been applied mechanically (see [Appendix D](#)).

For truth-table minimization, cases with an exact 0.5 in any causal condition were excluded, producing a primary truth-table sample of 88 cases out of 129. This is not ordinary missing-data deletion. These cases are substantively ambiguous by calibration design, and the soft-lean fallback's partial success, with only 5 cases at $Y = 0.5$ versus 22 at PRICE = 0.5 where no such fallback exists, is itself evidence the design mechanism works where it was applied. [Section 5.9.6](#) examines the implications of this exclusion for the PRICE result in detail.

5.9 QCA results

The analysis used R package QCA version 3.21 under R 4.3.2 ([Duşa, 2019](#); [Thiem & Duşa, 2013](#)), following the same sequence [Appendix F](#) rehearses by hand, namely necessity for Y and $\neg Y$, truth tables, complex, parsimonious and intermediate solutions for both outcomes, directional expectations stated before intermediate minimization and robustness checks. The primary truth table used $\text{incl.cut} = 0.80$, $\text{PRI} = 0.70$ and $\text{n.cut} = 1$. Directional expectations for Y assumed the presence of CERT, PRICE, ENV and NORM each contributes positively, and for $\neg Y$ that the absence of each does.

5.9.1 Necessity. No single condition reached the conventional 0.90 necessity-consistency threshold for Y. **Table 6** reports necessity consistency (inclN), the relevance of necessity (RoN) and coverage (covN) for each single condition and its negation, for Y ($N = 129$).

The strongest candidates, NORM ($\text{inclN} = 0.845$) and CERT ($\text{inclN} = 0.840$), still fall short. The pilot therefore does not support the claim that any single condition is strictly necessary for high willingness to choose the certified sustainable destination.

For $\neg Y$, the strongest single condition was $\neg \text{ENV}$ ($\text{inclN} = 0.894$, $\text{RoN} = 0.738$, $\text{covN} = 0.747$), close to but not quite the 0.90 threshold.

Table 5. Membership-score distributions ($N = 129$)

Set	Min	Q1	Median	Mean	Q3	Max	Exact 0.5
Y	0.05	0.35	0.55	0.512	0.65	0.95	5
CERT	0.35	0.40	0.50	0.545	0.60	0.95	11
PRICE	0.10	0.233	0.367	0.457	0.633	0.90	22
ENV	0.10	0.233	0.367	0.416	0.589	0.856	9
NORM	0.30	0.40	0.60	0.583	0.70	0.95	9

Note(s): Referenced in [Section 5.8](#). “Exact 0.5” reports the count of respondents whose calibrated score lands precisely on the crossover for that set. All values reproduced exactly from the field data by the analysis script ([Appendix C](#))

Table 6. Necessity of single conditions for Y ($N = 129$)

Condition	inclN	RoN	covN
CERT	0.840	0.798	0.789
¬CERT	0.650	0.817	0.731
PRICE	0.698	0.846	0.783
¬PRICE	0.698	0.710	0.657
ENV	0.711	0.919	0.876
¬ENV	0.735	0.667	0.645
NORM	0.845	0.735	0.742
¬NORM	0.561	0.818	0.689

Note(s): Referenced in [Section 5.9.1](#). inclN = necessity-inclusion (consistency), RoN = relevance of necessity and covN = necessity coverage. No condition or negated condition reaches the conventional 0.90 necessity threshold. Values reproduced from the field data by the analysis script ([Appendix C](#))

This contrasts with the toy example in [Appendix F](#), where Knowledgeable Guide was fully necessary (consistency = 1.000) by design, illustrating that necessity is, in the end, an empirical claim rather than an assumption one can build in. Extended necessity analysis confirms no conjunction of conditions reaches 0.90 either, since the highest value observed is the conjunction of CERT, PRICE, ENV and NORM at inclN = 0.542.

5.9.2 Truth table for Y. After excluding exact causal-condition crossovers ($N = 88$), 15 of 16 logically possible configurations were observed. [Table 7](#) reports the full truth table prior to Boolean minimization, namely all observed configurations with their membership counts, raw inclusion (incl) and PRI scores, so that the minimization step can be audited against the raw rows.

Only one configuration passed the primary sufficiency rule (incl.cut = 0.80, PRI = 0.70, n.cut = 1). That single sufficient row combines CERT, PRICE, ENV and NORM,

Table 7. Truth table for Y prior to Boolean minimization ($N = 88$)

CERT	PRICE	ENV	NORM	OUT	n	incl	PRI
0	0	0	0	0	21	0.659	0.175
0	0	0	1	0	12	0.740	0.275
0	0	1	0	0	2	0.825	0.257
0	0	1	1	0	1	0.851	0.319
0	1	0	0	0	4	0.761	0.108
0	1	0	1	0	8	0.808	0.300
0	1	1	0	0	0		
0	1	1	1	0	2	0.889	0.391
1	0	0	0	0	3	0.715	0.209
1	0	0	1	0	10	0.758	0.351
1	0	1	0	0	4	0.837	0.404
1	0	1	1	0	1	0.857	0.445
1	1	0	0	0	4	0.791	0.352
1	1	0	1	0	1	0.842	0.467
1	1	1	0	0	4	0.896	0.650
1	1	1	1	1	11	0.902	0.725

Note(s): Referenced in [Section 5.9.2](#). OUT = outcome assignment under the primary sufficiency rule (incl.cut = 0.80, PRI = 0.70, n.cut = 1), incl = raw inclusion (sufficiency consistency) and PRI = proportional reduction in inconsistency. Row 0 1 1 0 (¬CERT · PRICE · ENV · ¬NORM) is the only logical remainder, with no observed cases. Only the all-present row (1 1 1 1) passes the sufficiency rule. Reproduced exactly from the field data by the analysis script ([Appendix C](#))

with $n = 11$, $\text{incl} = 0.902$ and $\text{PRI} = 0.725$. Several other rows showed high raw inclusion but a noticeably lower PRI, for example $\text{CERT} \cdot \text{PRICE} \cdot \text{ENV} \cdot \neg \text{NORM}$ ($n = 4$, $\text{incl} = 0.896$, $\text{PRI} = 0.650$), $\neg \text{CERT} \cdot \text{PRICE} \cdot \text{ENV} \cdot \text{NORM}$ ($n = 2$, $\text{incl} = 0.889$, $\text{PRI} = 0.391$) and $\text{CERT} \cdot \neg \text{PRICE} \cdot \text{ENV} \cdot \text{NORM}$ ($n = 1$, $\text{incl} = 0.857$, $\text{PRI} = 0.445$), illustrating why PRI should be inspected alongside raw consistency rather than in its place, since these low PRI values indicate simultaneous subset relations in both Y and $\neg Y$, a pattern the toy example's deliberately designed high-PRI rows did not exhibit. Only one corner, $\neg \text{CERT} \cdot \text{PRICE} \cdot \text{ENV} \cdot \neg \text{NORM}$, is a logical remainder with no observed cases.

5.9.3 Sufficiency solution for Y . The complex, parsimonious and intermediate solutions all converged, since the single sufficient path combining CERT, PRICE, ENV and NORM yielded a consistency of 0.902, a PRI of 0.725 and a coverage of 0.534. Table 8 presents this solution in standard fsQCA notation, with core and peripheral conditions distinguished.

This differs structurally from the toy example, where the parsimonious solution diverged from the intermediate by dropping the necessary condition. Here, with only one sufficient row and no simplifiable remainders, the parsimonious algorithm has nothing to drop, so all three solutions are identical by necessity, a methodological finding in its own right, since the three-solution comparison is genuinely diagnostic only when the truth table contains multiple sufficient rows together with logical remainders.

Substantively, within this pilot, high willingness to choose the certified sustainable destination was observed most consistently when all four memberships were jointly present. This is a conjunctural result restricted to the present sample and analysis parameters. It does not imply any one condition carries an independent net effect (Section 5.9.1 confirms none does), and it does not establish that all four conditions are *required*, a stronger claim the pilot's single row of 11 cases could not support. A claim of that strength would require a larger sample in which several sufficient rows and their counterfactual relations could be compared.

5.9.4 Sufficiency solution for $\neg Y$. The negated outcome produced a different solution, demonstrating causal asymmetry clearly. Table 9 presents the four absence-based paths sufficient for the negated outcome $\neg Y$ ($N = 88$) in standard notation.

Four absence-based paths are sufficient for low willingness, with an overall consistency of 0.853, a PRI of 0.693 and a coverage of 0.721. High willingness (Y) is associated with the joint presence of all four conditions along a single path, whereas low willingness ($\neg Y$) arises through several configurations in which three of four conditions are absent, equifinality for the negated outcome.

Notably, the complex, parsimonious and intermediate solutions for $\neg Y$ all coincide on these four paths in the analysis output (Appendix C). The parsimonious minimization as implemented in the R package does not shorten the formula, and the intermediate solution, which could in principle drop conditions whose absence is not needed by the stated directional expectation, retains all four paths. An alternative hand minimization can merge the two paths that share $\neg \text{CERT}$ and $\neg \text{NORM}$ into a single term (the conjunction $\neg \text{CERT} \cdot \neg \text{NORM}$ combined with $\neg \text{CERT} \cdot \neg \text{PRICE} \cdot \neg \text{ENV}$ and $\neg \text{PRICE} \cdot \neg \text{ENV} \cdot \neg \text{NORM}$), but every

Table 8. Sufficiency solution for Y ($N = 88$)

Path	CERT	PRICE	ENV	NORM	Consistency	PRI	Coverage
1	•	•	•	•	0.902	0.725	0.534
Overall solution					0.902	0.725	0.534

Note(s): Referenced in Section 5.9.3. The complex, parsimonious and intermediate solutions converge on the single path $\text{CERT} \cdot \text{PRICE} \cdot \text{ENV} \cdot \text{NORM}$. • = core causal condition present (large circle in standard notation). All four conditions are core because the parsimonious and intermediate solutions coincide. Blank cells would indicate conditions that do not appear in a path (don't care), and no such cell occurs here. Consistency and PRI figures follow Table 7, and coverage is computed on the 88-case truth-table sample

Table 9. Sufficiency solutions for $\neg Y$ ($N = 88$)

Path	CERT	PRICE	ENV	NORM	Consistency	PRI	Coverage
1	⊗	⊗	⊗		0.873	0.712	0.624
2	⊗	⊗		⊗	0.914	0.797	0.534
3	⊗		⊗	⊗	0.927	0.830	0.579
4		⊗	⊗	⊗	0.892	0.758	0.564
Overall solution					0.853	0.693	0.721

Note(s): Referenced in [Section 5.9.4](#). Four absence-based paths sufficient for the negated outcome ($\neg Y$), demonstrating causal asymmetry relative to the single conjunctural path found for Y ([Table 8](#)). $\otimes$ = core causal condition absent (large crossed circle in standard notation), and a blank cell means the condition does not matter for that path (don't care). All conditions are core because the parsimonious, complex and intermediate solutions coincide on these four paths. Reproduced exactly from the field data by the analysis script ([Appendix C](#))

formulation of the solution retains all four conditions and the same absence-based structure. In other words, the pilot's asymmetry finding, namely one presence-based path for Y and four absence-based paths for $\neg Y$, does not depend on the choice of minimization option, which makes it harder to dismiss as an artifact of the algorithm. This is a different diagnostic role for the three-solution comparison than in the toy example ([Appendix F](#), Section F.6), where the parsimonious solution diverged because a genuine simplification was available. Here the convergence itself is the finding.

5.9.5 Robustness. The Y solution was tested across a grid of alternative incl.cut, PRI and n.cut values (full grid in [Appendix E](#)). Three findings emerge. First, the solution is stable across inclusion thresholds, since with $PRI \geq 0.65$ and $n.cut = 1$ the solution is invariant to incl.cut across 0.75–0.85. Second, it is somewhat sensitive to the frequency threshold, since at $n.cut = 2$ single-case rows are reclassified as remainders and the solution reduces to a three-condition conjunction with two equally valid alternatives, $CERT \cdot PRICE \cdot NORM$ and $CERT \cdot ENV \cdot NORM$, where CERT and NORM survive in both while one of PRICE or ENV drops, indicating that PRICE and ENV are the less robustly implicated conditions and CERT and NORM the most robust. Third, the solution is limited by PRI, since at $PRI \geq 0.75$ no row is selected for Y at all, because the all-present row has a PRI of only 0.725, a meaningful but not especially strong sufficiency claim, reported rather than obscured. The full parameter grid, including the thresholds at which no solution survives, is reported in [Appendix E](#) rather than cherry-picked, consistent with [Kroglund et al.'s \(2015\)](#) recommendation for full parameter-space reporting and with earlier guidance on robustness assessment in QCA ([Skaaning, 2011](#)).

An additional screening check addresses the six attention-check failures retained in the sample ([Section 5.8](#)). Re-running the analysis at $N = 123$ leaves the Y solution unchanged (the conjunction of CERT, PRICE, ENV and NORM, incl = 0.902, PRI = 0.726, coverage = 0.542) and the $\neg Y$ solution substantively unchanged (three absence-based paths, overall incl = 0.842, PRI = 0.676, coverage = 0.769). All substantive conclusions are invariant to this exclusion.

5.9.6 The PRICE crossover problem, its impact, rationale and redesign. Impact on the truth table. The 22 PRICE cases at exactly 0.50 (17% of the sample) are the largest single contributor to the reduction from 129 retained cases to the 88-case truth-table sample. Their exclusion is deliberate and substantive, since a respondent whose calibrated score is exactly 0.50 is, by the anchor definitions in [Section 5.5](#), at maximum ambiguity with respect to price tolerance, and membership judgments for such cases would be arbitrary. Excluding them keeps the truth table free of cases whose corner assignment is a coin flip. The cost is statistical, since the rows of the truth table involving PRICE rest on a reduced case base, which compounds the PRI sensitivity reported in [Section 5.9.5](#) and counsels particular caution in interpreting PRICE's role.

The two-stage architecture used for the outcome, CERT and NORM resolves the crossover problem at the design stage. Stage 1 forces a side of the crossover (choose X, choose Y or

undecided), and Stage 2 measures distance from it on a ladder whose rungs deliberately avoid landing on the ambiguous midpoint. The soft-lean fallback then converts “undecided” into a graded lean wherever a lean exists. The evidence that this mechanism works is the crossover counts themselves, since the sets with a forced first stage produced 5 (Y), 11 (CERT) and 9 (NORM) exact crossovers, whereas PRICE, the only single-stage ladder, which effectively asked “which side, and how far?” in one step, produced 22. The single-stage ladder also made the measurement more vulnerable to response clustering, since with no forced choice first, respondents could sit on the rungs adjacent to the crossover (rung 4, “up to about 15% more”) without ever committing to a side.

The next instrument iteration replaces the PRICE single-stage ladder with the two-stage forced-choice-plus-ladder format and re-checks the anchor positions so that no raw response value lands exactly on the calibration crossover ([Appendix A](#), precondition 2). This is the framework’s intended division of labor, since Step 4 surfaced the problem at screening, where it can be addressed by redesign, rather than at calibration, where it would likely have been hidden by distribution-driven threshold repair.

5.10 Interpretation

First, a single conjunctural path emerged, since within this sample high willingness was observed only in the configuration that combines all four conditions jointly, with no single condition necessary ([Section 5.9.1](#)), a conjunctural pattern of exactly the kind configurational methods exist to detect and one consistent with prior single-recipe findings in tourism fsQCA ([Salman et al., 2023](#); see [Section 6.3](#)). Second, no necessary condition appeared, since, unlike the toy example, where Guide was fully necessary by design, real choice data show no condition reaching the 0.90 threshold, suggesting sustainable destination choice appears multiply realizable rather than constrained by a single prerequisite. Third, causal asymmetry emerged, since Y was observed under the joint presence of all four conditions while $\neg Y$ arises through several absence-based paths, meaning the conditions that prevent sustainable choice are not simply the mirror image of those that enable it. Fourth, the PRICE crossover problem counts as a diagnostic success, since the 22 PRICE exact-crossover cases are not a pilot failure but evidence the framework works as intended, because Step 4 surfaces the problem at screening, where it can be addressed through redesign, rather than at calibration, where it would likely be hidden by distribution-driven threshold repair. That the single-stage PRICE ladder produced more crossovers than the two-stage CERT and NORM ladders suggests that the two-stage architecture is the more effective instrument-design pattern. Fifth, the robustness limits qualify the primary result, since the Y solution does not survive a stricter PRI threshold (≥ 0.75), and at the higher frequency cut ($n.cut = 2$) it loses one of the four conditions (either PRICE or ENV), which places the strength of the sufficiency claim at “meaningful but modest” rather than “strong.”

6. Discussion

6.1 What the framework achieved

The pilot is consistent with the Measurement Before Calibration framework in four respects. First, the instrument avoided ceiling effects, since despite measuring a socially desirable construct, the outcome and conditions show variation across the membership scale rather than uniformly high membership, which appears attributable to the trade-off ladder design (Step 3), because each rung names a concrete and escalating cost, forcing respondents to locate the point at which they would no longer choose B. Second, ambiguity was made visible rather than hidden, since the 22 PRICE exact-crossover cases are reported openly ([Table 5](#)) rather than recalibrated away, and the soft-lean fallback reduced exact crossovers considerably where it was applied ($Y = 5$, $NORM = 9$) compared with where it was not ($PRICE = 22$). Third, the results demonstrate conjunctural and asymmetric reasoning directly, since Y was observed

under joint presence of all four conditions while $\neg Y$ arises from multiple absence combinations, a set-theoretic finding rather than a list of independent net effects. Fourth, every methodological decision is traceable, since calibration anchors, directional expectations, truth-table thresholds (reported with PRI alongside inclusion), excluded crossover cases and robustness limits are all documented explicitly, the operational definition of “transparent QCA” the framework exists to enable.

Because the pilot is small and its purpose illustrative, these outcomes should be read as demonstrations of the framework’s intended behavior rather than as evidence that the framework improves results relative to conventional practice. A formal comparison against a conventional instrument is the next step, as [Section 6.9](#) notes.

6.2 What the pilot revealed about the instrument

First, PRICE needs a two-stage redesign, as argued in [Section 5.9.6](#), since its single-stage ladder clustered 22 respondents at the exact crossover, versus 11 (CERT) and 9 (NORM), both two-stage, and the forced first-stage choice appears to anchor which side of the crossover a respondent falls on before the ladder asks how far, whereas the single-stage PRICE ladder effectively asked both questions at once. Second, the ENV behavioral anchoring worked reasonably well, since the three-item scale (mean = 0.416, median = 0.367) avoided the ceiling pattern typical of symbolic environmental-concern items, consistent with Step 3’s fallback principle, though the robustness grid cautions that ENV, like PRICE, is among the less robustly implicated conditions at the higher frequency threshold ([Section 5.9.5](#)). Third, the NORM reword, which dropped its reference to Destination A/B to allow pre-vignette administration, did not obviously degrade the measure, since NORM shows the widest range (0.30–0.95) and highest median (0.60) of the four conditions, and CERT and NORM are the two conditions that survive in every $n.cut = 2$ alternative ([Section 5.9.5](#)).

6.3 Comparison with extant findings

[Salman et al. \(2023\)](#), studying ecotourism stakeholder management, likewise found only one causal recipe survived sufficiency analysis (all three conditions jointly required, consistency 0.968), with no alternative path emerging. The present pilot replicates this structural pattern, a single conjunctural path, in a different setting (destination choice rather than stakeholder management) and with more conservative fit (0.902) and explicit robustness limits. That both studies, using quite different instruments, arrive at a single conjunction suggests the conjunctural structure may be characteristic of sustainability-relevant decisions rather than an artifact of either instrument.

[Palacios-Marqués et al. \(2017\)](#) showed that two defensible calibration procedures produced contradictory necessity findings on identical hotel data. The present pilot does not solve this problem, but it illustrates the framework’s response to it, since anchors were specified from substantive meaning before analysis ([Section 5.5](#)) and pressure-tested across a parameter grid ([Section 5.9.5](#); [Appendix E](#)), the reader can see exactly which claims survive which assumptions. The PRI and $n.cut$ sensitivity found here is the pilot’s own instance of the sensitivity phenomenon, reported rather than suppressed.

[Wattanawaraporn and Manosuthi \(2025\)](#) found that bundling a credible pro-community signal with commercial incentives can backfire among the travelers a sustainability claim is meant to attract, if the combination reads as symbolically inconsistent. Consistent with that experimental evidence, CERT was never sufficient on its own in this pilot, since it appears in the Y solution only in combination with PRICE, ENV and NORM. Neither study supports treating certification credibility as a standalone driver, and the wider sustainability-marketing literature has long documented the paradoxes of communicating sustainable tourism products ([Font & McCabe, 2017](#)).

The pilot’s asymmetry finding, one presence-based path for Y and several absence-based paths for $\neg Y$, echoes the nonmirrored structures reported by [Hossain et al. \(2021\)](#) and

Galeazzo and Furlan (2018) in nontourism domains, and extends the pattern to sustainable destination choice.

6.4 Theoretical implications

The pilot was not designed as a theory test, and its results should not be over-interpreted, but three patterns seem worth noting as hypothesis-generating. First, sustainable destination choice appears conjunctural in this sample, since no condition alone was necessary and the sole sufficient configuration required all four jointly, meaning credible certification, price tolerance, environmental concern and social reinforcement may operate as a package rather than as independent drivers, consistent with tourism-specific evidence that pro-environmental travel behavior draws on attitudes, norms and behavioral control jointly (Han, 2015). Second, the causes of not choosing seem more varied than the causes of choosing ($\neg Y$ equifinal, four paths and Y conjunctural, one path), suggesting the process that suppresses sustainable choice is less constrained than the process that produces it. Third, certification credibility alone does not appear sufficient, since CERT appears in the Y solution only in combination with PRICE, ENV and NORM, consistent with the experimental evidence discussed in Section 6.3. These patterns are observations from a pilot, not established findings and the theory-testing study sketched in Section 6.9 is designed to test them.

6.5 Methodological implications

Krogslund *et al.*'s (2015) demonstration that fsQCA results are highly sensitive to calibration anchors carries an implication worth drawing out. If calibration is this sensitive, the upstream instrument feeding it, whose distributions determine which anchors are even plausible, matters correspondingly more. This framework is, in effect, a direct response, since it does not solve the sensitivity problem as such but addresses the condition that makes it worse.

Existing guidelines (Cangialosi, 2023; Nikou *et al.*, 2024; Oana *et al.*, 2021; Rasoolimanesh *et al.*, 2023; Thomann and Maggetti, 2020) are necessary but probably not sufficient, since a researcher who skips their calibration procedures and reliability checks will likely produce unreliable results, but a scale with a Cronbach's alpha of 0.90 can still produce ceiling effects no calibration anchor can recover from. The five steps proposed here are not a new calibration algorithm, nor a revision of Ragin's (2008) logic, but a sequence that operationalizes what his framework presupposes but does not itself supply, namely an instrument capable of producing scores to which an external anchor can meaningfully be applied.

6.6 Practical implications

First, for researchers applying fsQCA to survey data, the framework translates into a short pre-fielding checklist, namely define the set and its three anchors in behavioral terms before writing items, build trade-off ladders rather than agreement scales wherever the construct is a choice under constraint, count exact crossover cases before calibration and report a robustness grid rather than a single solution. Second, for destination marketers, the pilot suggests certification credibility works in combination, since consistent with evidence that travelers will pay more for verified green options only within bounds (Kang, Stein, Heo, & Lee, 2012), a certified offer priced well above the market, marketed without reference to social expectations and targeted at travelers without activated environmental concern would lack three of the four memberships observed in the only sufficient configuration. Third, the ENV and NORM results suggest measurement of "sustainability attitude" that asks about actual past behavior, and of social influence that is measured independently of the choice scenario, can discriminate where conventional agreement items collapse, which is relevant to any tourism instrument intended to support calibration.

6.7 Policy recommendations

Policy recommendations must be stated cautiously, since they rest on a single small pilot. Two tentative points seem worth making. First, the CERT results give some weight to policies that strengthen third-party auditing and public verification, since respondents distinguished sharply between independently audited and self-certified claims, and certification credibility appeared in the sufficient configuration only in this verified form, a view consistent with hospitality managers' own preference for audited certification schemes (Geerts, 2014). Second, the conjunctural result implies that interventions aimed at increasing sustainable destination choice may need to target all four memberships simultaneously, namely price acceptability, credible certification, environmental concern and social reinforcement, whereas interventions aimed at reducing barriers could be more targeted, since $\neg Y$ was reachable through several absence-based paths. Both suggestions are hypothesis-generating and require a confirmatory study in Section 6.9.

6.8 Limitations

The pilot relies on a convenience sample of 129 respondents, not a representative sample of tourists or sustainable travelers. Population-level generalization is not claimed, and would not be especially defensible even had the results been stronger. The Y solution rests on just 11 cases within a single truth-table row, sufficient for an instrument demonstration but not a substantive theory claim. A larger sample would likely populate more truth-table corners and could reveal additional sufficiency paths. At $PRI \geq 0.75$, no solution survives for Y at all. The primary result ($PRI = 0.725$) is meaningful but not robust enough to survive a stricter threshold, a limitation reported transparently rather than hidden.

The single-stage PRICE ladder produced 22 exact crossover cases (17% of the sample), indicating the measure needs redesigning before the next study (Section 5.9.6). Both Y and $\neg Y$ were analyzed, but no separate theoretical model for $\neg Y$ was specified in advance, so the $\neg Y$ results should be treated as exploratory. The pilot demonstrates that the proposed instrument can produce discriminating and calibratable scores. It does not yet demonstrate that it performs better than a conventional instrument on the same respondents. A formal benchmark comparison, which would field a conventional Likert operationalization alongside the trade-off instrument, is the immediate next step and was not available at the time of this revision.

6.9 Avenues for future research

The pilot's companion study is envisaged as a between-subjects vignette experiment manipulating Certification, Community Benefit and Price Premium directly, at approximately 400 respondents, analyzed at the respondent-configuration level. The combination of experimental manipulation with configurational analysis follows established precedents (Salonen, Zimmer, & Keränen, 2021). Community Benefit re-enters as an experimental factor, manipulated rather than measured, the more defensible way to isolate its causal role. The study would reuse the same two-stage outcome measure and calibration logic, test whether the conjunction of CERT, PRICE, ENV and NORM replicates in a larger and more representative sample, examine whether experimental assignment alters the configurational patterns observed, and retain the same transparency standards throughout, namely published anchors, stated directional expectations, a robustness grid and crossover reporting.

The most direct answer to the question "is this instrument better than the alternative?" is a within-subject comparison, in which the same respondents complete both the trade-off instrument and a conventional Likert battery, scored under the calibration practices typical of the literature (e.g., percentile-based direct calibration), with the two pipelines compared on distributional quality, crossover prevalence, truth-table content and solution stability. This comparison is a priority for the next data collection.

Rasoolimanesh (2026) shows fsQCA and Necessary Condition Analysis (Dul, 2016) can be used complementarily, since NCA identifies necessary but not sufficient conditions. The

framework applies to both, since an instrument that fails to produce gradation compromises NCA's ceiling-line estimation just as much as fsQCA's sufficiency analysis. Relatedly, combined symmetric (PLS-SEM) and asymmetric (fsQCA) analyses are increasingly recommended in hospitality research (Rasoolimanesh, Ringle, Sarstedt, & Olya, 2021), and the instrument-design discipline proposed here applies upstream of both.

The framework is not specific to sustainable destination choice. Constructs vulnerable to symbolic agreement, including authenticity, loyalty, satisfaction, experience quality and resident support, would likely benefit from the same discipline, including the gastronomy and wine tourism gap identified by Eluwole *et al.* (2024).

Step 4 can incorporate conventional reliability and validity checks (DeVellis, 2017; Nikou *et al.*, 2024; Rasoolimanesh *et al.*, 2023; Väättäinen & Dickenson, 2019) without contradiction, since these checks are necessary but not sufficient, because a perfectly reliable instrument can still produce distributions to which a theory-driven anchor cannot meaningfully be applied.

AI contribution disclaimer

This manuscript was initially drafted in Thai. Generative artificial intelligence was subsequently used to translate and assist in writing the full manuscript in English. The authors reviewed, revised and verified the AI-assisted text and take full responsibility for the final content.

Statements and declarations

Human participants were reviewed and approved by the Institutional Review Board of Khon Kaen University (IRB No. HE663166).

Supplementary material

The supplementary material for this article can be found online.

References

- Andrade-Valbuena, N. A., & Dos Santos, M. A. (2026). Cognitive and contextual configurations in entrepreneurial opportunity evaluation: Insights from fsQCA. *BMC Psychology*, 14(1), 203. doi: [10.1186/s40359-025-03940-1](https://doi.org/10.1186/s40359-025-03940-1).
- Auspurg, K., & Hinz, T. (2015). *Factorial survey experiments*. Sage Publications.
- Cangialosi, N. (2023). Fuzzy-set qualitative comparative analysis (fsQCA) in organizational psychology: Theoretical overview, research guidelines, and a step-by-step tutorial using R software. *The Spanish Journal of Psychology*, 26, e21. doi: [10.1017/SJP.2023.21](https://doi.org/10.1017/SJP.2023.21).
- De Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Soares, G. R. D. L. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences Europe*, 32(1), 19. doi: [10.1186/s12302-020-0300-3](https://doi.org/10.1186/s12302-020-0300-3).
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.
- Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of 'necessary but not sufficient' causality. *Organizational Research Methods*, 19(1), 10–52. doi: [10.1177/1094428115584005](https://doi.org/10.1177/1094428115584005).
- Duşa, A. (2019). *QCA with R: A comprehensive resource*. Springer International Publishing.
- Eluwole, K. K., Lasisi, T. T., Parvez, M. O., & Cobanoglu, C. (2024). Application of fuzzy-set qualitative comparative analysis (fsQCA) in hospitality and tourism research: A bibliometric study. *Journal of Hospitality and Tourism Insights*, 7(5), 3032–3054. doi: [10.1108/JHTI-08-2023-0572](https://doi.org/10.1108/JHTI-08-2023-0572).

- Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393–420. doi: [10.5465/amj.2011.60263120](https://doi.org/10.5465/amj.2011.60263120).
- Font, X., & McCabe, S. (2017). Sustainability and marketing in tourism: Its contexts, paradoxes, approaches, challenges and potential. *Journal of Sustainable Tourism*, 25(7), 869–883. doi: [10.1080/09669582.2017.1301721](https://doi.org/10.1080/09669582.2017.1301721).
- Galeazzo, A., & Furlan, A. (2018). Lean bundles and configurations: A fsQCA approach. *International Journal of Operations & Production Management*, 38(2), 513–533. doi: [10.1108/IJOPM-11-2016-0657](https://doi.org/10.1108/IJOPM-11-2016-0657).
- Geerts, W. (2014). Environmental certification schemes: Hotel managers' views and perceptions. *International Journal of Hospitality Management*, 39, 87–96. doi: [10.1016/j.ijhm.2014.02.007](https://doi.org/10.1016/j.ijhm.2014.02.007).
- Geremew, Y. M., Huang, W.-J., & Hung, K. (2024). Fuzzy-set qualitative comparative analysis as a mixed-method and analysis technique: A comprehensive systematic review. *Journal of Travel Research*, 63(1), 3–26. doi: [10.1177/00472875231168619](https://doi.org/10.1177/00472875231168619).
- Gligor, D., & Bozkurt, S. (2020). FsQCA versus regression: The context of customer engagement. *Journal of Retailing and Consumer Services*, 52, 101929. doi: [10.1016/j.jretconser.2019.101929](https://doi.org/10.1016/j.jretconser.2019.101929).
- Greckhamer, T. (2011). Cross-cultural differences in compensation level and inequality across occupations: A set-theoretic analysis. *Organization Studies*, 32(1), 85–115. doi: [10.1177/0170840610380806](https://doi.org/10.1177/0170840610380806).
- Greckhamer, T., Misangyi, V. F., Elms, H., & Lacey, R. (2008). Using qualitative comparative analysis in strategic management research: An examination of combinations of industry, corporate, and business-unit effects. *Organizational Research Methods*, 11(4), 695–726. doi: [10.1177/1094428107302907](https://doi.org/10.1177/1094428107302907).
- Han, H. (2015). Travelers' pro-environmental behavior in a green lodging context: Converging value-belief-norm theory and the theory of planned behavior. *Tourism Management*, 47, 164–177. doi: [10.1016/j.tourman.2014.09.014](https://doi.org/10.1016/j.tourman.2014.09.014).
- Hossain, M. A., Rahman, S., Quaddus, M., Hooi, E., & Olanrewaju, A.-S. (2021). Factors affecting performance of open government data initiatives: A multi-method approach using SEM and fsQCA. *Journal of Organizational Computing and Electronic Commerce*, 31(4), 300–319. doi: [10.1080/10919392.2021.2018258](https://doi.org/10.1080/10919392.2021.2018258).
- Kang, K. H., Stein, L., Heo, C. Y., & Lee, S. (2012). Consumers' willingness to pay for green initiatives of the hotel industry. *International Journal of Hospitality Management*, 31(2), 564–572. doi: [10.1016/j.ijhm.2011.08.001](https://doi.org/10.1016/j.ijhm.2011.08.001).
- Kraus, S., Ribeiro-Soriano, D., & Schüssler, M. (2018). Fuzzy-set qualitative comparative analysis (fsQCA) in entrepreneurship and innovation research — the rise of a method. *The International Entrepreneurship and Management Journal*, 14(1), 15–33. doi: [10.1007/s11365-017-0461-8](https://doi.org/10.1007/s11365-017-0461-8).
- Krogslund, C., Choi, D. D., & Poertner, M. (2015). Fuzzy sets on shaky ground: Parameter sensitivity and confirmation bias in fsQCA. *Political Analysis*, 23(1), 21–41. doi: [10.1093/pan/mpu016](https://doi.org/10.1093/pan/mpu016).
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. doi: [10.1002/acp.2350050305](https://doi.org/10.1002/acp.2350050305).
- Kumar, S., Sahoo, S., Ali, F., & Cobanoglu, C. (2024). Rise of fsQCA in tourism and hospitality research: A systematic literature review. *International Journal of Contemporary Hospitality Management*, 36(7), 2165–2193. doi: [10.1108/IJCHM-03-2023-0288](https://doi.org/10.1108/IJCHM-03-2023-0288).
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge University Press.
- Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, 2(4), 245–264.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. doi: [10.1037/a0028085](https://doi.org/10.1037/a0028085).
- Miller, K. M., Hofstetter, R., Krohmer, H., & Zhang, Z. J. (2011). How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, 48(1), 172–184. doi: [10.1509/jmkr.48.1.172](https://doi.org/10.1509/jmkr.48.1.172).

- Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*, 43(1), 255–282. doi: [10.1177/0149206316679252](https://doi.org/10.1177/0149206316679252).
- Nikou, S., Mezei, J., Liguori, E. W., & El Tarabishy, A. (2024). FsQCA in entrepreneurship research: Opportunities and best practices. *Journal of Small Business Management*, 62(3), 1531–1548. doi: [10.1080/00472778.2022.2147190](https://doi.org/10.1080/00472778.2022.2147190).
- Oana, I.-E., Schneider, C. Q., & Thomann, E. (2021). *Qualitative comparative analysis using R: A beginner's guide*. Cambridge University Press.
- Olan, F., Liu, S., Neaga, I., Chen, H., & Nakpodia, F. (2019). How cultural impact on knowledge sharing contributes to organizational performance: Using the fsQCA approach. *Journal of Business Research*, 94, 313–319. doi: [10.1016/j.jbusres.2018.02.027](https://doi.org/10.1016/j.jbusres.2018.02.027).
- Olya, H. G. T., & Akhshik, A. (2019). Tackling the complexity of the pro-environmental behavior intentions of visitors to turtle sites. *Journal of Travel Research*, 58(2), 313–332. doi: [10.1177/0047287517751676](https://doi.org/10.1177/0047287517751676).
- Olya, H. G. T., & Altinay, L. (2016). Asymmetric modeling of intention to purchase tourism weather insurance and loyalty. *Journal of Business Research*, 69(8), 2791–2800. doi: [10.1016/j.jbusres.2015.11.015](https://doi.org/10.1016/j.jbusres.2015.11.015).
- Palacios-Marqués, D., Roig-Dobón, S., & Comeig, I. (2017). Background factors to innovation performance: Results of an empirical study using fsQCA methodology. *Quality and Quantity*, 51(5), 1939–1953. doi: [10.1007/s11135-016-0414-2](https://doi.org/10.1007/s11135-016-0414-2).
- Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, 58, 102310. doi: [10.1016/j.ijinfomgt.2021.102310](https://doi.org/10.1016/j.ijinfomgt.2021.102310).
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. doi: [10.1037/0021-9010.88.5.879](https://doi.org/10.1037/0021-9010.88.5.879).
- Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.
- Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.
- Rasoolimanesh, S. M. (2026). Applying fsQCA, NCA and NConfA in tourism research. *Annals of Tourism Research*, 116, 104069. doi: [10.1016/j.annals.2025.104069](https://doi.org/10.1016/j.annals.2025.104069).
- Rasoolimanesh, S. M., Ringle, C. M., Sarstedt, M., & Olya, H. (2021). The combined use of symmetric and asymmetric approaches: Partial least squares-structural equation modeling and fuzzy-set qualitative comparative analysis. *International Journal of Contemporary Hospitality Management*, 33(5), 1571–1592. doi: [10.1108/IJCHM-10-2020-1164](https://doi.org/10.1108/IJCHM-10-2020-1164).
- Rasoolimanesh, S. M., Valaei, N., & Rezaei, S. (2023). Guideline for application of fuzzy-set qualitative comparative analysis (fsQCA) in tourism and hospitality studies. In *Cutting edge research methods in hospitality and tourism*. doi: [10.1108/978-1-80455-063-220231009](https://doi.org/10.1108/978-1-80455-063-220231009).
- Salman, A., Jaafar, M., Mohamad, D., & Khoshkam, M. (2023). Understanding multi-stakeholder complexity & developing a causal recipe (fsQCA) for achieving sustainable ecotourism. *Environment, Development and Sustainability*, 25(9), 10261–10284. doi: [10.1007/s10668-022-02488-z](https://doi.org/10.1007/s10668-022-02488-z).
- Salonen, A., Zimmer, M., & Keränen, J. (2021). Theory development in servitization through the application of fsQCA and experiments. *International Journal of Operations & Production Management*, 41(5), 746–769. doi: [10.1108/IJOPM-08-2020-0537](https://doi.org/10.1108/IJOPM-08-2020-0537).
- Schneider, C. Q., & Wagemann, C. (2010). Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. *Comparative Sociology*, 9(3), 397–418. doi: [10.1163/156913210X12493538729793](https://doi.org/10.1163/156913210X12493538729793).
- Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences*. Cambridge University Press.
- Skaaning, S.-E. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods and Research*, 40(2), 391–408. doi: [10.1177/0049124111404818](https://doi.org/10.1177/0049124111404818).

- Smithson, M., & Verkuilen, J. (2006). *Fuzzy set theory: Applications in the social sciences*. Sage Publications.
- Thiem, A., & Duşa, A. (2013). Qca: A package for qualitative comparative analysis. *The R Journal*, 5(1), 87–97. doi: [10.32614/RJ-2013-009](https://doi.org/10.32614/RJ-2013-009).
- Thomann, E., & Maggetti, M. (2020). Designing research with qualitative comparative analysis (QCA): Approaches, challenges, and tools. *Sociological Methods and Research*, 49(2), 356–386. doi: [10.1177/0049124117729700](https://doi.org/10.1177/0049124117729700).
- Väätäinen, M., & Dickenson, P. (2019). (Re)examining the effects of athlete brand image (ABI) on psychological commitment: An empirical investigation using structural equation modelling (SEM) and fuzzy set qualitative comparative analysis (fsQCA). *European Sport Management Quarterly*, 19(2), 244–264. doi: [10.1080/16184742.2018.1508242](https://doi.org/10.1080/16184742.2018.1508242).
- Voelckner, F. (2006). An empirical comparison of methods for measuring consumers' willingness to pay. *Marketing Letters*, 17(2), 137–149. doi: [10.1007/s11002-006-5147-x](https://doi.org/10.1007/s11002-006-5147-x).
- Wagemann, C., Buche, J., & Siewert, M. B. (2016). QCA and business research: Work in progress or a consolidated agenda?. *Journal of Business Research*, 69(7), 2531–2540. doi: [10.1016/j.jbusres.2015.10.010](https://doi.org/10.1016/j.jbusres.2015.10.010).
- Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research*, 43(3), 476–496. doi: [10.1080/00273170802285941](https://doi.org/10.1080/00273170802285941).
- Wattanawaraporn, R., & Manosuthi, N. (2025). When discounts undermine good intentions: Identity salience and symbolic misalignment in CBT marketing. *ABAC Journal*, 45(4), 545–564. doi: [10.59865/abacj.2025.53](https://doi.org/10.59865/abacj.2025.53).

Further reading

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. doi: [10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t).
- To, C. K. M., Au, J. S. C., & Kan, C. W. (2019). Uncovering business model innovation contexts: A comparative analysis by fsQCA methods. *Journal of Business Research*, 101, 783–796. doi: [10.1016/j.jbusres.2018.12.042](https://doi.org/10.1016/j.jbusres.2018.12.042).

Corresponding author

Prarawan Senachai can be contacted at: prarse@kku.ac.th