INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND INNOVATIVE STUDIES
ISSN: 2820-7157
www.ijsrisjournal.com

April 2025 **Volume 4** **Number 2** **53-60**

Received Date: 25 February 2025 **Accepted Date: 18 March 2025** **Published Date: 01 April 2025**

The Interpretive Delphi: Repurposing a Consensus Instrument for Abductive Inquiry

Maryame CHERIF OUAZZANI CHAHDI [1], Saida MARSO [2]

1. PhD student and researcher, Tangier National School of Business and Management – Abdelmalek Essaadi University, Department of Marketing, Logistics and Management
2. Lecturer and researcher, Tangier National School of Business and Management – Abdelmalek Essaadi University, Department of Marketing, Logistics and Management

Abstract

The Delphi method was built to manufacture agreement. Conceived at RAND to forecast the unforecastable, it iterates anonymous expert judgement until dispersion narrows, and reports the narrowing as its result. Six decades later, the instrument has migrated far from its origins — into health policy, education, management, planning — but its epistemology has travelled largely unexamined, and it arrives in interpretive research carrying assumptions that interpretive research explicitly rejects: that expert opinion approximates a truth, that convergence indicates validity, and that residual disagreement is noise to be reduced.

This article argues that the instrument can be kept and the epistemology discarded. It sets out the **interpretive Delphi**: a redesign in which the object of interest is not the consensus a panel reaches but the *structure of the disagreement it fails to resolve*. Three inversions are specified. Convergence is read as a diagnostic of shared institutional exposure rather than of correctness. Persistent divergence, ordinarily treated as panel failure, becomes the primary finding and the trigger for the next inferential move. And the second round is reconceived: it stops asking experts to revise towards the mean and starts asking them to *explain why they will not*.

The redesign is situated within abductive inquiry — Dubois and Gadde's *systematic combining* — where the Delphi occupies a position no other instrument fills: it is the only technique that can, in a single operation, survey a field of expert judgement and mark the places where that field is internally contradictory. Those contradictions are precisely what abduction requires as input.

The article specifies the design, states the criteria of rigour appropriate to it (which are not those of the classical Delphi), addresses the standard objections — small panels, absent statistics, unreproducible consensus — and offers a reporting protocol. A worked illustration is drawn from a study of inter-institutional governance in a peripheral investment region.

Keywords: Delphi method; abduction; systematic combining; interpretive research; expert panels; qualitative rigour.

1. Introduction

An instrument is not neutral with respect to what it is used to find. This is a commonplace of methodology, but it is a commonplace that the Delphi method has largely escaped, and the escape has cost it.

The classical Delphi does one thing with great efficiency: it moves a group of experts towards agreement while preventing them from influencing one another through status, volume, or persistence. Anonymity, iteration, and controlled feedback are its three mechanisms, and they are ingenious. Where a group is needed to produce a single judgement and no criterion exists against which to check it, the Delphi is very hard to improve on.

But the interpretive researcher does not need a single judgement, and does not believe there is one. She works from the premise that institutional actors inhabit different positions, see different portions of a phenomenon from those positions, and construct different accounts accordingly — and that these differences are not error but *data*. An instrument whose defining operation is the reduction of difference is, on the face of it, exactly the wrong tool.

And yet interpretive researchers use it, frequently. The pattern in the literature is consistent: the Delphi appears in a mixed design, is conducted more or less classically, produces convergence scores, and the scores are then discussed with a caveat noting that they are not to be taken as measurements. The caveat is sincere and it is insufficient. It disclaims an interpretation without removing what produced it. The scores were generated by a procedure that treats disagreement as failure; discussing them in interpretive language afterwards does not undo the procedure.

The argument of this article is that the incompatibility is not with the *instrument* but with the *protocol*, and that the protocol can be rebuilt. What survives is anonymity, iteration, controlled feedback, and the structured elicitation of expert judgement at a scale no interview series can match. What is discarded is the objective — consensus — and the machinery that serves it: the pursuit of narrowing dispersion, the stopping rule keyed to stability, the treatment of the outlier as a problem.

What replaces it is an instrument that surveys a field of expert judgement and returns a **map of its fault lines**.

2. What the Delphi was built to do, and what it carries with it

2.1. Origins and the consensus imperative

The method emerged at RAND in the early 1950s, in a context — strategic forecasting under radical uncertainty — where no data existed and expert judgement was the only available substitute. Dalkey and Helmer's formulation established the architecture that persists: successive anonymous rounds, statistical feedback of the group response, opportunity for revision. Linstone and Turoff's (1975) survey consolidated it, and every subsequent methodological treatment has worked within it.

The architecture encodes a premise. If iteration is expected to produce convergence, and convergence is reported as the finding, then the design assumes that the experts' judgements are noisy estimates of an underlying quantity, and that averaging plus feedback will reduce the noise. This is a defensible assumption when forecasting the year a technology will mature. It is an assumption about the *world*, not merely about the *method*, and it does not travel silently.

2.2. What the migration brought with it

As the Delphi migrated into policy, health, education and management research, the forecasting rationale fell away but the machinery stayed. Contemporary applications typically retain: a target consensus threshold, expressed as a percentage of the panel within some band; a dispersion statistic, most often the interquartile range or a standard deviation; a stopping rule triggered when dispersion ceases to fall; and a results section organised around which items achieved consensus and which did not.

Each of these is a residue of the original epistemology, and each does specific damage in interpretive use.

Table 1. What the classical protocol assumes, and what interpretive research holds

Classical Delphi assumes	Interpretive research holds
Expert judgements are estimates of a single underlying state	Expert judgements are situated constructions from distinct institutional positions
Convergence indicates validity	Convergence indicates shared exposure — which may equally indicate shared blindness
Divergence is noise, to be reduced across rounds	Divergence is signal, to be *explained*
The panel's second-round revision improves the estimate	Revision under feedback may indicate reflection, or it may indicate conformity — and the classical protocol cannot distinguish these
The outlier weakens the result	The outlier occupies a position from which something is visible that the others cannot see

The last two rows are where the damage concentrates. A classical Delphi that reports movement towards the mean cannot say whether an expert revised because a colleague's argument persuaded her or because she saw that she stood alone. Anonymity removes status pressure; it does not remove conformity pressure, and the protocol is silent on which it observed. For a method whose credibility rests on the quality of the revision, this is not a minor gap.

2.3. The unexploited asset

Set the consensus objective aside, however, and the Delphi possesses a property that no other qualitative instrument offers.

It is the only technique that elicits *structured, comparable, and revisable* judgement from a set of experts on a common set of propositions, without those experts being in a room together. An interview series yields depth but not comparability: each conversation goes where it goes. A focus group yields interaction but destroys independence. A survey yields comparability but no revision, and no reasons.

The Delphi yields all four — comparability, independence, revision, and reasons — and it is the *combination* that makes it valuable to abductive inquiry. Because what abduction needs, at the moment it needs it, is a surface on which the anomalies in one's own framework become visible. A set of propositions put to a panel, and returned with the marks of where the panel splits, is exactly such a surface.

The instrument has been kept for the wrong reason. Its consensus function is the least interesting thing about it.

3. The Three Inversions

The interpretive Delphi retains the architecture and reverses the objective. Three specific inversions constitute the redesign.

3.1. Convergence read as exposure, not as truth

When a panel of institutional experts agrees, the classical reading is that the agreed proposition is probably right. The interpretive reading is different and more cautious: **agreement indicates that the panel occupies positions from which the same thing is visible.**

This is a weaker claim, and a more useful one, because it immediately raises the question the classical reading suppresses: *what is visible from all these positions, and what is visible from none of them?* A panel drawn from public agencies will converge on propositions that public agencies can see. It will converge equally strongly on the omissions that public agencies share. High convergence is therefore not, by itself, evidence of anything except homogeneity of vantage — and homogeneity of vantage is a property of the *panel*, not of the world.

The methodological consequence is a reporting obligation. Every convergence must be accompanied by an account of the panel's compositional structure, and the researcher must state what the composition makes visible and what it forecloses. A convergence score reported without this account is uninterpretable.

3.2. Divergence promoted from failure to finding

In the classical protocol, an item on which the panel does not converge after the final round is a residual: it is reported, sometimes, in a table of non-consensus items, and it does not enter the analysis.

In the interpretive protocol, **it is the analysis.** Persistent divergence — divergence that survives feedback, that experts defend when invited to reconsider — is the strongest signal the instrument produces, because it identifies a proposition on which institutional position *determines* judgement. That is precisely the kind of proposition an interpretive study exists to find.

Three types of persistent divergence can be distinguished, and the distinction is analytically consequential:

Table 2. Types of Persistent Divergence and Their Interpretive Meaning

Type	Signature	What it indicates
Positional	The split tracks institutional affiliation	The proposition is contested *because* it distributes credit, blame or mandate
Experiential	The split tracks seniority or case exposure	Different periods or case types produce different realities; the phenomenon is unstable in time
Conceptual	The split tracks nothing observable; experts define the term differently	The construct is not shared — a finding about the *field*, not about the panel

Conceptual divergence is the most valuable and the most systematically discarded. When a panel splits because "collaboration" means something different in a regulatory agency than in a promotion agency, the classical protocol records a failure to converge. The interpretive protocol records the discovery that the field lacks a shared object — which is very often the most important thing there is to know about it.

3.3. The second-round stops asking for revision and starts asking for reasons

This is the operational core of the redesign, and it is the point at which the two protocols become genuinely different instruments.

The classical second round returns to each expert the group distribution and invites revision. The implicit request is: *move, or justify not moving.*

The interpretive second round returns something else. It returns to each expert **the positions of the panel and the arguments offered for them**, stripped of any indication of what the majority thinks — no medians, no distributions, no indication of where the expert sits relative to the group. And it asks a different question: not *do you wish to revise?* but *here is a reading of this proposition that differs from yours; what does it see that you do not, and what do you see that it does not?*

This move accomplishes three things at once. It eliminates conformity pressure by removing the information that produces it. It converts the second round from a revision exercise into an **elicitation of the reasons for disagreement** — which is the datum of interest. And it produces, as output, not a narrowed distribution but a set of accounts explaining why the distribution will not narrow.

A panel that has been through an interpretive second round has not agreed on more. It has explained, in its own terms, why it agrees on less than one might expect. That explanation is the result.

4. Position Within Abductive Inquiry

4.1. Where the instrument sits

Abductive research, in Dubois and Gadde's (2002) formulation of *systematic combining*, proceeds by continuous confrontation between an evolving theoretical framework and an empirical reality that keeps refusing to fit it. The engine of the process is the **anomaly**: the observation that the current framework cannot absorb, and that therefore forces its revision.

The practical difficulty of abductive work is that anomalies are hard to find on purpose. Interviews surface them, but slowly and unsystematically — one learns after twenty conversations that a category was wrong. Documents surface them only where the researcher already suspected a problem.

The interpretive Delphi surfaces them **on demand**. A framework is operationalised into propositions; the propositions are put to a panel; the panel's fault lines return a map of where the framework is contested by people who know the field. Those fault lines are candidate anomalies. Some will dissolve on inspection. Others will not, and those are the ones that redirect the inquiry.

Figure 1 shows the position.

Figure 1. The interpretive Delphi in the abductive cycle

Framework (a priori) → **[Round 1: propositions to panel]** → Field of judgement, with fault lines → **[Round 2: elicitation of reasons for divergence]** → Structured account of contested propositions → Candidate anomalies → **[Interviews: pursue anomalies in depth]** → Framework revision → *(loop)*

Documentary analysis runs alongside, supplying the institutional facts against which both panel judgement and interview accounts are checked.

4.2. Sequence matters

The instrument's value depends on where it is placed. Run *before* the interviews, the Delphi tells the researcher which propositions the field contests, and the interviews can then be

aimed at those contests rather than sweeping the ground indiscriminately. Run *after*, it becomes a member-checking exercise — legitimate, but a waste of what the instrument can do.

The recommended sequence is therefore: documentary analysis establishes the institutional facts; the Delphi maps the field of judgement and its fault lines; the interviews pursue the fault lines. Each stage narrows the next. This is not triangulation in the confirmatory sense — three methods checking one finding — but something closer to **successive interrogation**, in which each instrument is aimed by the one before it.

4.3. Why not simply use another group method?

If what the researcher wants is the structure of expert disagreement, an obvious question arises: why adapt the Delphi at all, rather than reach for a method already designed to surface divergent views? Three alternatives suggest themselves, and comparing them clarifies what the interpretive Delphi uniquely provides.

The **focus group** is the first candidate, and it fails on independence. Its defining feature — interaction — is precisely what contaminates the judgements the interpretive Delphi wants to keep separable. In a focus group, a fault line may be an artefact of the loudest participant, or it may be suppressed because dissent is socially costly in the room. The researcher cannot tell which. The Delphi's anonymity and physical separation preserve the independence of each position, so that a divergence, when it appears, is a divergence between *positions* and not between *personalities in a room*. This is not a marginal advantage; it is the difference between a fault line one can interpret and one cannot.

The **nominal group technique** improves on the focus group by structuring contribution and reducing dominance, but it still converges by design — its silent generation phase feeds into a ranking-and-aggregation phase whose purpose is to produce a group priority. Like the classical Delphi, it treats disagreement as the raw material for an agreed output, not as the output itself. Repurposing it would require the same inversion the present article performs on the Delphi, with no compensating gain.

A **series of independent interviews** preserves independence perfectly but sacrifices comparability. Because each interview follows its own course, the researcher ends with a set of rich but non-aligned accounts, from which fault lines can be reconstructed only by imposing, after the fact, a common frame the participants never faced. The Delphi imposes that common frame *ex ante* — every expert responds to the identical proposition set — so that divergence is legible as divergence on a shared question rather than inferred across conversations that were never commensurable.

The interpretive Delphi thus occupies a cell that none of the alternatives fills: independence *and* comparability *and* structured revisability. It is not that the Delphi is superior to these methods for their own purposes; it is that, for the specific purpose of mapping where a field of expert judgement is internally contradictory, the combination of properties the Delphi offers is not available elsewhere. The choice of instrument is not habit or convenience but a response to what the research question requires.

5. Rigour: which criteria, and which do not apply

The interpretive Delphi cannot be assessed by the criteria of the classical one, and it is important to say so plainly rather than to leave the reader to apply them by default.

Table 3. Criteria of rigour

Classical criterion	Status here	Interpretive replacement
Consensus threshold reached	**Does not apply** — consensus is not sought	Fault lines identified and typed (positional / experiential / conceptual)
Dispersion narrows across rounds	**Does not apply** — narrowing would indicate conformity, not learning	Reasons for non-narrowing elicited and reported
Panel size sufficient for stability	**Reframed**	Panel *composition* covers the institutional positions the phenomenon spans
Stability of response across rounds	**Reframed**	Consistency between an expert's Round 1 judgement and her Round 2 reasons
Reproducibility of the consensus	**Does not apply**	**Auditability**: the full chain from proposition to fault line to interview probe is traceable

Three criteria carry the weight.

Compositional coverage. The panel is not a sample and makes no claim to represent a population. It is a *set of positions*, and its adequacy is judged by whether the positions from which the phenomenon is visible are occupied. A panel

of ten that covers the regulatory, promotional, elected, operational and technical vantages is stronger than a panel of thirty drawn from two of them. This must be argued explicitly, and the argument is falsifiable: a reviewer can point to a position the panel omits.

Traceability of the inferential chain. Every anomaly pursued in the interviews must be traceable back to the fault line that generated it, and every fault line back to the proposition and the round in which it appeared. This is the interpretive Delphi's substitute for reproducibility, and it is a demanding one: it requires the researcher to have recorded her own reasoning at the moment she made it, not reconstructed it afterwards.

Non-conversion of scores. Any numerical output — and the instrument does produce numbers, since experts are asked to position themselves — must be reported as *ordinal and local*: this proposition was more contested than that one, within this panel. It must not be averaged across propositions, aggregated into an index, subjected to dispersion statistics, or compared to a threshold. A composite score built from ten experts' judgements on five propositions is not a measurement of anything, and constructing one abandons the epistemology the redesign exists to protect.

6. A worked illustration

To show the design in operation rather than only in principle, consider its application in a study of inter-institutional investment governance in a peripheral region — the setting is incidental; the mechanics are the point.

A panel of institutional experts was assembled to cover the vantages from which the governance of investment attraction is visible: the regulatory, the promotional, the elected-political, the operational, and the technical. The panel was not a sample of any population; it was a set of positions, chosen so that no major vantage on the phenomenon was unoccupied, and the omitted vantages — notably the investors' own — were stated explicitly as a limit on what the exercise could see.

Round 1 put to the panel a set of propositions operationalising a working framework about what constrains investment attraction and how institutional coordination bears on it. The propositions were deliberately framed to be contestable — a proposition no one could disagree with tells the researcher nothing about where a field divides. Experts positioned themselves on each, and, crucially, gave reasons.

The interpretive reading of Round 1 attended not to how many experts endorsed each proposition but to the *shape* of the response across the panel. On some propositions the panel converged — and, per Section 3.1, this convergence was read as a signature of shared institutional exposure, prompting the question of what all these positions could see in common and what none of them could. On others the panel divided, and the divisions were typed. Where a split tracked institutional affiliation — regulatory bodies reading a proposition one way, promotional bodies another — the divergence was classed as *positional*, indicating a proposition contested because it distributed mandate or credit. Where a split tracked seniority or the era of an expert's formative experience, it was classed as *experiential*, indicating a phenomenon unstable in time. Where a split tracked nothing observable, with experts evidently defining a key term differently, it was classed as *conceptual* — the discovery that the field lacked a shared object.

Round 2 did not feedback the distribution. It returned to each expert the *reasons* the panel had given for divergent positions, stripped of any indication of majority or of where the expert stood relative to the group, and asked not "do you wish to revise?" but "here is a reading that differs from yours; what does it see that you do not, and what do you see that it does not?" The output was not a narrowed distribution but a set of accounts explaining why the distribution would not narrow — why, on the contested propositions, institutional position determined judgement.

These accounts became the input to the subsequent interview phase. The fault lines identified in Round 2 were carried forward as the anomalies the interviews would pursue in depth — not the propositions the panel agreed on, which required no further interrogation, but the ones on which it structurally divided. The Delphi had done what it was repurposed to do: it had surveyed a field of expert judgement and returned a map of its contradictions, aimed the next instrument at them, and left a documented trail from each proposition through its fault line to the probe it generated. No consensus was manufactured, and none was needed.

The illustration also surfaces the design's characteristic difficulty, which no methodological cleanliness removes: the researcher must record her own inferential reasoning *as she makes it*, because the traceability on which the design's rigour depends cannot be reconstructed after the fact without contamination. The map of fault lines is only as trustworthy as the contemporaneous record of how it was read.

7. Objections

"Ten experts is too few." Too few for what? The objection assumes a sampling logic that the design rejects. Ten is too few to estimate a population parameter, and the interpretive Delphi estimates none. The relevant question is whether ten experts occupy the positions from which the phenomenon is visible, and that question is answered by composition, not by count. The objection should be met head-on rather than

deflected: state the positions, state which are covered, and state which are not.

"Without statistics, this is just a group interview conducted by email." No — and the difference is precise. A group interview lacks independence: participants hear each other and adjust. A series of individual interviews lacks comparability: each goes its own way. The interpretive Delphi preserves independence *and* imposes a common set of propositions, which is what makes fault lines visible as fault lines rather than as an assortment of unrelated opinions. Remove the propositions and the comparison collapses.

"You have removed the mechanism that makes the Delphi a Delphi." Partly true, and worth conceding. Feedback towards the mean is removed. Anonymity, iteration, structured elicitation and controlled feedback of *reasons* are retained. Whether the result deserves the name is a question about labels. What matters is whether it produces knowledge the classical protocol destroys — and it does: the classical protocol's central operation is the elimination of exactly the material this design treats as its finding.

"Persistent divergence might simply mean the propositions were badly written." It might, and the design must be able to tell the difference. This is what the typology in Section 3.2 is for. A badly written proposition produces *conceptual* divergence with no positional structure — experts split, but the split tracks nothing. That is a warning, and it should be reported as one. Positional and experiential divergence, by contrast, exhibit structure, and structure is not an artefact of bad wording.

8. Reporting Protocol

The following minimum is proposed for interpretive Delphi studies. Its purpose is to make the design auditable and to prevent the silent reimportation of consensus logic.

1. **Panel composition table**: each expert by institutional position, seniority, and the vantage that position affords. Not names; positions.
2. **Statement of coverage**: which vantages the panel occupies and — explicitly — which it does not, with the consequences for what the study can and cannot see.
3. **Proposition set**, verbatim, with the theoretical source of each.
4. **Round 1 field**: the distribution of judgement per proposition, reported ordinally, without composite scores.
5. **Fault-line table**: contested propositions, typed by divergence category, with the panel's own reasons quoted.
6. **Round 2 instrument**: the exact wording of the elicitation, demonstrating that no majority information was fed back.
7. **Anomaly trace**: which fault lines were carried into the interview phase, and why those and not others.
8. **Negative reporting**: fault lines that dissolved on further inquiry, and propositions the panel agreed on that the interviews subsequently contradicted. This last is the most informative and the most frequently omitted.

9. Conclusion

The Delphi has been used in interpretive research for decades under a borrowed epistemology, and the borrowing has been mostly unexamined. Researchers have retained an instrument built to eliminate disagreement while writing, in their limitation's sections, that disagreement is what they were interested in.

The resolution is not to abandon the instrument. It is to notice that its most valuable output has always been the residue it was designed to discard. A panel of experts who will not agree, and who can say precisely why, has told the researcher something that no amount of convergence could: that the phenomenon under study looks different from different institutional positions, and that the difference is not error.

The interpretive Delphi formalises this. It keeps the anonymity, the iteration, the structured elicitation — everything that made the method valuable — and replaces the objective. Where the classical protocol asks *what do the experts agree on*, this one ask *what will they not agree on, and what does the shape of their disagreement reveal about the field they inhabit*.

The second question is harder to answer. It is also, in interpretive research, the only one worth asking.

References

- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458
- Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research, 55*(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8

- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219–245. https://doi.org/10.1177/1077800405284363

- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

- Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, *78*(9), 1695–1704. https://doi.org/10.1016/j.techfore.2011.04.005

- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

- Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. Addison-Wesley.

- Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, *30*(3), 167–186. https://doi.org/10.1177/0735275112457914

- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE. *Avant soumission : vérifier tous les DOI sur Crossref et adapter le style bibliographique au gabarit exact de la revue cible.*