WORKING PAPER

The Worst Part

Why the moment your customers hate most is the most valuable measurement in your business, and why nothing you own is collecting it.

WXI · WORST EXPERIENCE INDICATOR | PUBLISHED BY ACOUSTIC PTE. LTD., SINGAPORE | 20 JULY 2026

THE ARGUMENT IN BRIEF

WHAT THEY KEEP

A peak and an ending

That is the whole of what a customer remembers. The worst moment and the last three minutes govern what they recall, repeat and decide. The average and the duration wash out of memory entirely.

WHY YOU LACK IT

Four mechanisms erase it

The worst moment is destroyed before you see it. The customer averages it away, the scale averages it again, politeness keeps it out of the box, and the score is gamed at the point of collection.

HOW TO RECOVER IT

Five words get it back

One question, one text box, no scale: *what was the worst part?* Its grammar refuses "fine" and its premise removes the cost of saying the true thing.

01 The most valuable sentence in your business

Somewhere in your business, right now, a customer is having the worst experience your company will produce this week.

They will not tell you. They will not call. They will not escalate. If they answer your survey at all they will give you a 7 and leave the comment box empty, because the comment box asked how their experience was, and "fine" is a complete answer to that question.

Then they will leave. And in eleven months, when someone asks why churn moved, that moment will be gone. It was never written down. The only record of it is a 7.

> *"Your most unhappy customers are your greatest source of learning."*
>
> Bill Gates, ***Business @ the Speed of Thought***, 1999[1]

That line has been on the wall of every customer experience team for twenty-five years. It gets quoted in the deck, agreed with in the room, and then the meeting moves on to the score.

Here is the uncomfortable part. Almost nobody built an instrument that collects the learning. The quote is on the wall and the data is not in the building. Every mainstream CX tool takes a single reading after the fact, on a scale, from a customer who has already averaged their entire experience down to one digit. The worst moment, which is the thing Gates was pointing at, gets averaged away in the customer's own head before your software ever sees it.

Twenty-five years of agreement, and the industry built the opposite instrument.

This paper is about correcting that. The argument runs in four steps:

1. The worst moment governs what a customer remembers, tells other people, and decides about you.
2. Every instrument you currently own is designed, structurally, to destroy it.
3. There is a five-word question that recovers it.
4. And there is a way to put a price on the answer, so that somebody actually funds the fix.

The fourth step is where most CX programmes die, and we will spend real time on it.

02 What your customers actually remember

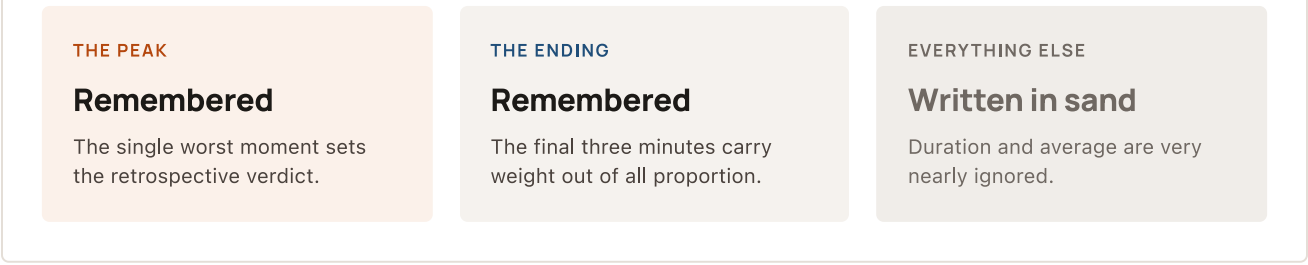
Start with the finding that should have reorganised the industry and did not.

In the mid-1990s, two researchers recorded, minute by minute, the pain of patients undergoing colonoscopy and lithotripsy: 154 patients in one study, 133 in the other. They had the real experience on tape. Then they asked the patients afterwards how bad it had been, and compared.[2]

What people remembered tracked two things. The worst moment. And the last three minutes. That was it.

The total amount of pain barely registered. The colonoscopies ran from four minutes to sixty-nine. The correlation between how long a procedure lasted and how bad the patient said it was: 0.03. The researchers had already named this in earlier work: duration neglect.[3] The experience your customer actually had, integrated over its full length, is close to irrelevant to what they walk away with.

FIGURE 1 · WHAT SURVIVES IN MEMORY



Then they ran the experiment that proves it.[4]

682 patients, randomly assigned. Half received a standard colonoscopy. Half received an identical procedure plus three extra minutes with the scope left in place, unmoving. Uncomfortable, and noticeably milder than everything before it. That group endured strictly more discomfort, for strictly longer, than the control group. Every objective measure says their experience was worse.

They remembered it as *less* unpleasant. And they were more likely to come back.

Sit with that. You can make an experience objectively worse and have it remembered better, by changing one moment. What lives in your customer's head is a peak and an ending. Everything else is written in sand.

And the peak is asymmetric. Losses loom larger than equivalent gains,[5] so the damage done by one bad moment is not repaid by one good one. This is why the delight programme did not work. You added peaks at the top while leaving the peak at the bottom exactly where it was, and the bottom one is the one with the weight.

THREE CONSEQUENCES, AND EACH ONE IS UNCOMFORTABLE

Your average is fiction

Not a rough approximation of your customer's experience. A number that describes something no customer ever had. Nobody experiences your mean.

Your worst moment is your product

The thing your customer describes at dinner is the moment it went wrong. That moment is doing your marketing. It is also doing your churn.

You are optimising the wrong thing

If your CX programme is working on the average, it is working on the part that does not survive contact with human memory.

The research has been sitting there since 1996. It is cited constantly. It is on the slide. And the instrument the industry runs on collects a summary score.

03 Why you do not have this data

You might reasonably ask why, if the worst moment is this important, nobody has it written down. The answer is that four separate mechanisms are destroying it, and three of them are inside the tool you bought.

3.1 THE CUSTOMER AVERAGES IT BEFORE YOU SEE IT

Ask "how was your experience?" and the customer performs a computation. They take a peak, an ending, a mood, and a general sense of your brand, and they collapse all of it into one number. By the time that digit reaches your database, the worst moment is gone. It contributed to the number. It cannot be recovered from the number. You are reading the output of a lossy compression that ran inside somebody's head, and you are treating it as data.

3.2 THE INSTRUMENT THEN AVERAGES IT AGAIN

The Net Promoter Score takes that already-compressed digit and compresses it further. Scores of 9 and 10 become promoters. 7 and 8 are thrown away entirely. Everything from 0 to 6 becomes a detractor, which means a customer who gave you a 6 and a customer who actively despises you are recorded as the same event, while a 6 and a 7 are recorded as opposites.[6]

	Promoters	Passives	Detractors	NPS
Firm A	60%	20%	20%	+40
Firm B	40%	60%	0%	+40

Two entirely different companies. Firm A has a fifth of its customers actively hostile. Firm B has none. Their churn profiles have nothing in common, their priorities have nothing in common, and the instrument reports the same number.

Then the number arrives at a meeting with no instruction attached. You are at 32. Nothing tells you what to change. An entire consulting industry exists to sell back, at cost, the diagnosis the instrument declined to collect.

3.3 YOUR CUSTOMERS ARE BEING POLITE

Most people will not volunteer a criticism to a company that has given no sign of wanting one. This is not customers being unhelpful. It is customers being normal. Complaining unprompted carries a social cost, it feels like an imposition, and the person doing it has no reason to believe anything will come of it. So they say nothing, tick 7, and leave. The information exists. Your customer has it. Your form gave them no reason to hand it over.

3.4 THE SCORE IS GAMED AT THE POINT OF COLLECTION

You already know this one. It is taught. Dealerships, hotels, support desks: explain to the customer that anything below a ten counts as a failure, ask them to raise concerns directly instead of in the survey, time the request for the moment of maximum goodwill. Note what is being gamed. A *level* can be lifted by lifting the

mood in the room at the moment of asking. That is what makes the whole thing so easy to corrupt, and it is a structural property rather than a moral failing of the people doing it.

> FOUR MECHANISMS, ALL POINTING THE SAME WAY
>
> Your most valuable data is destroyed by the customer, destroyed again by the scale, withheld out of politeness, and manipulated at the point of collection. Then the residue is averaged and put on a slide. Gates said your unhappy customers are your greatest source of learning. The industry built a machine for not listening to them.

04 THE RECOVERY MECHANISM
The question

Here is the entire recovery mechanism.

What was the worst part?

FIVE WORDS · ONE TEXT BOX · NO SCALE

It is the smallest component of the instrument and it does the most work, and every word in it is load-bearing.

WHY PEOPLE ANSWER IT

"How was it?" can be discharged in one word. It asks for a summary, and people supply "fine". The question has an exit and most customers take it, which is why your comment box is empty.

"What was the worst part?" cannot be answered with "fine". The grammar demands a thing. To answer at all, the customer has to name an incident. There is no exit.

It presupposes that a worst part exists, and that presupposition is a gift. The customer is no longer volunteering a criticism to a company that would rather not hear it. They are answering a question that has already conceded the point. The social cost that section 3.3 described just evaporated, because you paid it for them, in advance, by asking.

The superlative forces a ranking. To answer, the customer runs back through the experience and selects the worst bit. That selection is the peak. Section 2 told you the peak was the only thing that mattered and the only thing you could not get. Here it is, and the customer extracted it for you, for free.

Honesty is the low-effort path. Inventing a grievance takes work. Recalling the bit that annoyed you takes none.

And it is a trust act. The question concedes, out loud, that something went wrong. Compare it with an instrument that asks how likely you are to recommend the company to a friend, which is a request for a favour dressed up as research. One question serves the business. The other serves the customer and serves the business better as a consequence. People notice which one they are being asked.

WHAT IT COSTS YOU

Be clear about the trade. You will not learn what delighted anyone. The instrument is asymmetric on purpose, because the worst moment predicts churn and the best moment is pleasant to read and rarely tells you what to change. A business that wants its delights catalogued should buy a different product.

HOW TO BREAK IT

The question is short, which makes it fragile. Five ways to ruin it:

- **"Was there anything you didn't like?"** Restores the exit. "No" is a valid answer and most people take it. You have rebuilt the empty comment box.
- **"Why did you give that score?"** Asks the customer to justify a number. You get a paraphrase of the number.
- **"What was the best part, and what was the worst part?"** Halves the attention, invites escape into the positive, and re-averages the exact thing you just separated.
- **Multiple choice.** The moment you list the options you have told the customer what counts, and you have lost the incident you did not know to list. The entire value of this question sits in the answers you did not anticipate.
- **A character minimum.** Punishes the customer who says the true thing in four words. The four-word answers are frequently the best ones.

Placement: immediately after the experience, one box, no page transition, no obligation. A customer who is already in the frame will answer. Send them somewhere else and you lose most of them.

05 What a sentence does that a score cannot

This is where the argument stops being about measurement and starts being about whether anything changes.

A number produces a meeting about the number. Report that the score fell three points and watch. Someone questions the sample. Someone questions the methodology. Someone observes that last quarter was collected differently. Someone asks whether three points is even outside the margin of error, and nobody in the room can answer, because nobody in the room is a statistician. The debate is about the instrument. It is genuinely unresolvable. It consumes the meeting, and nothing gets fixed, because nothing specific was ever put on the table.

A sentence produces a meeting about the thing. Put this in front of the same room:

> *"The confirmation email arrived after the courier did."*

There is no methodology to attack. There is no sample to question. There is a fact, in a customer's own words, and either it is true or somebody goes and checks. Inside ten seconds the conversation has moved from whether the measurement is sound to who owns the confirmation email.

Numbers compress as they travel upward. Sentences do not. A three-point drop becomes “slightly down” becomes “broadly flat” by the third slide. A customer's sentence arrives in the boardroom in the words the customer used. It survives the hierarchy intact, and that property is worth more than most of the analytics features you are paying for.

A verdict has no owner. A sentence has one before it finishes.

A score of 32 belongs to nobody. “The confirmation email arrives after the courier” has an owner by the time you have read it aloud.

06 The problem with anecdotes

Now the objection that kills most CX programmes, and you have felt it. You already have customer complaints. There is a folder. There is a Slack channel. There is a support queue full of people telling you exactly what is wrong, in their own words, right now, and it has been there for years. Almost none of it gets fixed.

Not because anyone is stupid or indifferent. Because **anecdotes lose arguments to budgets.** Take a customer complaint to a finance director and you will be asked how many customers, how much it costs, and what happens if you do nothing. If the answer is “several people mentioned it”, the fix does not get funded, and the finance director is right.

> This is the graveyard. It is full of CX teams who collected excellent qualitative data, presented it with real conviction, and were asked for a number they did not have.

So the worst-part question, on its own, produces better anecdotes. Which produces a better folder. Which changes nothing. **The question is only half an instrument.** It tells you what happened. It cannot tell you what it costs. And what it costs is the only sentence in the room that moves money. Which brings us to the second half.

07 THE OTHER HALF
Putting a price on the worst part

To price a bad moment you need to know what the experience did to the customer. Not how they felt at the end, which includes their mood and their brand loyalty and their expectations and everything else they walked in carrying. What *the experience itself* did.

Two customers finish the same onboarding flow. Each rates it 8 out of 10. The first arrived expecting very little; a colleague had warned them it was difficult. Forty minutes later they were finished and mildly delighted. They rate it 8. The second arrived expecting excellence: the reviews, the pricing, a salesperson's promise of a fifteen-minute setup. Forty minutes later they were finished and quietly disappointed. They rate it 8.

> One is a company exceeding what a sceptic expected. The other is a company falling short of what it promised. Your instrument reports the same 8, for one reason: nobody took a reading before.

THE SECOND READING

So take one. At the start of the experience, at a moment the customer has formed an expectation and has not yet had the experience, ask them one question. At the end, ask the same customer a second. Subtract.

THE SCORE

WXI = post − pre

Displayed to the customer as a 1 to 10 slider, stored 0 to 100, giving a range of **−100 to +100**. Positive means the experience exceeded what the customer brought to it. Zero means it met expectations exactly. Negative states, in a number, by how much you fell short.

The customer's mood, loyalty, brand history and personal scoring habits are present in **both** readings. Subtract, and they cancel. What survives is closer to what you caused.

It is a control group of one: the same customer, ten minutes earlier.

WHY THIS IS NOT A NEW IDEA

It is worth saying plainly, because someone will check. Measuring expectation against outcome is forty-six years old. Richard Oliver modelled satisfaction as a function of expectation and its disconfirmation in 1980, and his results came from a **two-stage field study**: he measured people before, and he measured them again after.[7] The paradigm has been the dominant account of customer satisfaction ever since, with well over ten thousand citations.

The theory has always required two readings. The industry approximated with one, because in 1980 two readings meant fieldwork, and through the 2000s it meant two mailings and a matching problem. Every approximation abandoned the temporal separation the theory required, and the most popular one abandoned expectation altogether.[8]

A script tag on a confirmation page now collects a pre-reading in one tap, with a session identifier that solves matching for free. The constraint was technological and it has lifted. That is the whole claim: this is the old theory, measured the way it always said it should be measured, at a cost that only recently became negligible.

ONE MORE THING THE SUBTRACTION BUYS

A level can be gamed. A difference is much harder to game, because you would have to manipulate two moments in opposite directions, and the first one happens before your staff member knows how the interaction is going to go.

You can beg a customer for a ten. It is considerably harder to beg a customer for a delta.

The same logic handles the cross-border problem. Response styles differ systematically by country: a 26-country study found power distance, collectivism, uncertainty avoidance and extraversion all predicted how people use scales, with Japan showing the lowest acquiescence and the highest midpoint responding of all 26. Run the survey in English to non-native speakers and you add a further artefact that varies with their English.[9] Your regional comparison is partly a comparison of national scoring conventions. A difference between two readings from the same person cancels a stable response style. Someone who avoids the extremes avoids them twice.

AND NOW THE TWO HALVES WORK

Put the question and the delta together and you get a thing neither produces alone.

FIGURE 2 · FROM VERBATIM TO WORK ORDER — ONE POINT OF EXPERIENCE

THEME: Delivery window given as a day, not a time

SHARE OF WORST-PART RESPONSES	WXI AT THIS POX	MONTHLY VOLUME
34%	−11	12,000/mo

That is not an anecdote. It is a named problem, holding a measured share of a measured loss, across a known volume. It is a business case, and it was assembled out of a slider and a text box. A finance director will act on that. A finance director will not act on a folder of complaints, and is right not to.

THIS IS THE WHOLE THING

The delta says **where**, and **how much**.

The sentence says **what**.

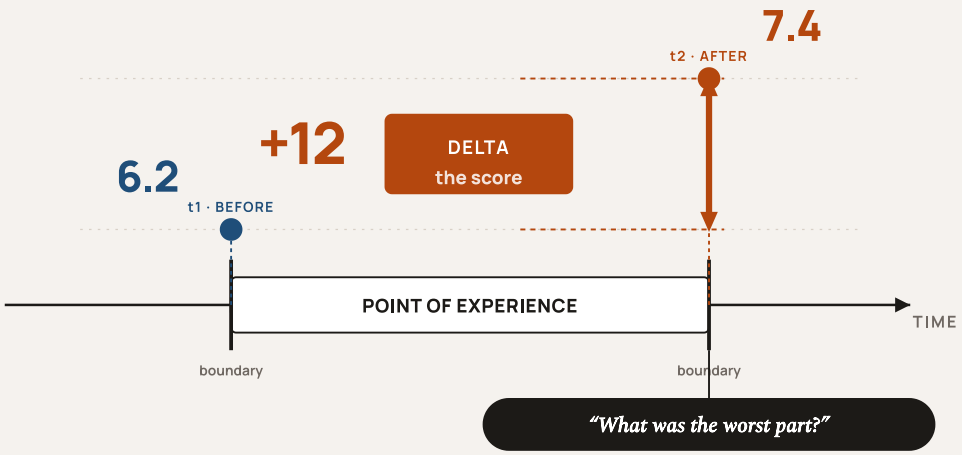
Together they produce a ranked work order with a price on it. Neither one moves a budget alone.

That is what a CX programme was always supposed to produce, and it is what none of them produce, because they were built on an instrument that collected half of the input and destroyed the other half on the way in.

08 The instrument, briefly

The Point of Experience is the unit: a bounded interaction with an identifiable beginning and end. A support ticket. An onboarding flow. A delivery. A booking. A trial. An appointment. Boundedness matters, and it disqualifies things. The qualifying test is concrete: *does a page or message already exist that the customer sees immediately before the experience begins?* Confirmation screens, booking acknowledgements, ticket receipts, dispatch notifications, appointment reminders, trial welcome pages. Where one exists, the pre-reading costs nothing to place. Ambient, unbounded relationships with a brand are out of scope, and we would rather say so than sell you something that will not work.

FIGURE 3 · THE POINT OF EXPERIENCE



Two measurements, an experience bounded between them, and the gap that is the score. A reading is taken before the Point of Experience begins and again after it ends. Their difference is the delta, the identity motif of the instrument. At the exit, one qualitative question: what was the worst part? The number says where and how much; the sentence says what.

THE JOIN IS THE CRITICAL PATH

The two readings must be joined to the same person, differenced per person, and aggregated only afterwards. Compare the mean pre-reading against the mean post-reading across a population and you have a between-groups difference that quietly reintroduces every confound the method exists to remove, because the people who answered at the end are not the people who answered at the start. This is the single point at which a careless implementation silently turns the instrument back into the thing it replaced.

THEMES HAVE TO HOLD THEIR NAMES

Free text at volume needs structuring, and a classification pipeline that regenerates its categories nightly will turn “Delivery window confusion” into “Unclear delivery timing” into “Scheduling communication”, at which point the trend line means nothing and the pricing table in section 7 cannot be built at all. Themes are persistent named entities with identity across runs. New responses join existing themes where they fit. New themes appear only where they genuinely must.

That last one looks like an engineering detail. It is the mechanism. Theme stability is what converts a pile of sentences into a priced work order, and it is the reason section 7 works.

09 Where this could be wrong

Any paper that omits this section is marketing. Four objections are serious, and if we are asking your customers what the worst part was, we can manage the same about our own instrument.

9.1 DIFFERENCE SCORES HAVE A BAD REPUTATION, AND IT IS EARNED

This is the strongest attack that exists on this method and it comes from the literature. A well-known 1993 review concluded that difference scores suffer from problems of reliability, discriminant validity, spurious correlation and variance restriction, and should generally not be used in consumer research.[10] It landed hardest on SERVQUAL, the service-quality instrument built on an expectation-minus-perception gap. A performance-only alternative was found superior, and replicated.[11] The gap model lost that argument.

The mechanism. Difference the two measures and the errors add while the true-score variance partially cancels, and the cancellation is worst when the two components correlate highly. In SERVQUAL, expectation and perception were captured on one instrument, in one sitting, from a respondent who already knew the outcome. Of course they correlated. The reliability collapsed.

Why this instrument is different. The pre-reading here is taken before the outcome exists. It cannot be contaminated by an event that has not happened. That lowers the correlation between the two components, which means less cancellation and a more reliable difference. The critique's own central mechanism is weakened by the one property that separates this from SERVQUAL.

And the claim is different. That literature attacks difference scores used as *measures of a construct*, where a direct measure exists and works better. This instrument offers the difference as a *measure of change*, where subtraction is the definition of the quantity rather than a proxy for it.

> THE CONCESSION, WHICH IS REAL
>
> The reliability penalty does not vanish. Two consequences, both binding. **First, individual deltas are noise.** One customer's WXI of −22 says very little about that customer. Delta is an aggregate statistic at a Point of Experience, and any interface that presents per-customer deltas as diagnostics is inviting people to over-read noise. **Second, you need more data than a single-reading tool needs** to say the same thing with the same confidence. That is the price of the design. It is worth paying, and you should hear it from us before you buy rather than discover it in month three.

9.2 ASKING FOR AN EXPECTATION MIGHT CHANGE THE EXPERIENCE

Stating a number may anchor the judgement that follows,[12] or it may sensitise the customer to disconfirmation, or it may simply focus attention on evaluation and change what they notice. The direction is not obvious and we do not know the answer. It is also cheaply testable: randomise the pre-reading across arriving customers, then compare the post-reading distribution of those who gave one against those who did not. Matching

distributions mean the effect is immaterial. Diverging ones are a finding worth publishing. Running that test and publishing the result, whatever it shows, is an obligation of anyone advancing this method, and it is on our roadmap rather than in our marketing.

9.3 A STRONG BRAND IS MECHANICALLY PENALISED

A customer arriving at 95 out of 100 has a maximum possible delta of +5. One arriving at 30 has 70 points of headroom. Effective marketing raises expectations, which consumes headroom, which suppresses delta. A luxury hotel and a budget airline cannot be compared on raw delta, and the comparison would flatter the airline. Raw delta is comparable within a firm over time, and across Points of Experience in a similar baseline band. Across firms with different brand positions it is an error.

> Always report the baseline distribution alongside the delta. A +2 against a mean baseline of 88 and a +2 against a mean baseline of 41 are different results. Reporting the delta alone would reproduce exactly the information-destroying move we spent section 3.2 criticising, and we would deserve everything we got.

9.4 THE ORPHANED BASELINE

The most dangerous one, and the one the method turns into an asset. Only customers who complete both readings produce a score. Customers who abandon midway never produce a post-reading and vanish from the aggregate. Abandonment correlates with things going badly. **An instrument built to find the worst experiences is exposed to losing exactly those experiences to attrition.** That has to be said out loud.

The recovery is that leaving leaves a trace. A pre-reading with no matching post-reading is an **orphaned baseline**, and it is data:

- **Its existence is a signal.** Somebody came in and did not leave through the door you were watching.
- **Its rate is a metric.** Orphaned baselines per Point of Experience is an abandonment measure that costs no additional instrumentation.
- **Its value carries information.** A cohort of orphans clustered at *high* pre-readings describes customers who arrived hopeful and left before the end. That population is worth more than most of the reports a CX platform will sell you.

A single-reading instrument is blind to this by construction. It has nothing to be orphaned from. A customer who abandons simply fails to respond, and a non-response is indistinguishable from indifference.

9.5 A LIMITATION WITH NO FIX

There is no benchmark corpus. Everyone in the room knows what an NPS of 30 means, and that familiarity is the incumbent's real and underrated moat. A director presenting a WXI of +4 has to teach the room a new scale first, and executives have limited patience for that. There is no clever answer. It resolves with accumulated data and time.

10 Monday

If you take one thing from this paper, take the question. It costs nothing and you do not need us to run it.

STEP ONE

Add the question to something

One touchpoint. One text box. No scale, no multiple choice, no character minimum, no “was there anything you didn't like”. Ask the actual question.

STEP TWO

Read the answers yourself

Not a summary. Not a sentiment score. The sentences, in the customer's words, for a fortnight.

STEP THREE

Watch the first meeting

Read one aloud. Notice that nobody argues with it. That is the whole mechanism, and you can feel it working before you have bought anything.

Then you will hit the wall in section 6. You will have a folder of true, specific, unarguable statements about what is wrong with your business, and you will not be able to get any of it funded, because you cannot say what it costs. That is the point at which you need the second reading. And that is what we build.

11 What it takes to run

The method is straightforward. Running it is an engineering problem, and the engineering is where it either works or quietly reverts to the thing it replaced.

FIVE THINGS, EACH WITH A SILENT FAILURE MODE

- The join has to hold across two moments and two systems.
- The pre-reading has to cost exactly one tap on a page somebody else built.
- Themes have to keep their names across runs, for years.
- Orphaned baselines have to be preserved rather than swept up as errors.
- Capture has to survive traffic without dropping a row, on a public endpoint, under tenancy isolation strict enough that one company's candour never reaches another company's screen.

Every one of those is a place where a well-intentioned implementation produces numbers that look plausible and mean nothing, and tells nobody. That is the part we build. A companion technical paper sets out the reference implementation: the capture snippet, the join, the classification pipeline, theme persistence and the reporting model.

12 The argument in one page

Your customers remember a peak and an ending. Everything else washes out, including the duration, including the average. The peak is usually the worst moment, and losses weigh more than gains, so that moment is doing your churn and your word of mouth regardless of what else you shipped.

You do not have it written down. The customer averaged it away before your form loaded. Your scale averaged it away again. Politeness kept it out of the comment box. Your staff were coached to keep it out of the survey.

There is a question that recovers it, and it is five words long, and it works because its grammar refuses "fine" and its premise removes the social cost of saying the true thing. The answers are worthless until priced. A folder of complaints has never been funded by anyone. Take a reading before the experience and one after, subtract, and the customer's mood and loyalty and expectations cancel out, leaving what you caused. Now the theme has a number attached, and the number is what moves money.

The delta says where and how much. The sentence says what. Together they are a work order.

Bill Gates was right in 1999, and the instrument the industry built does the opposite of what he said. That is a twenty-five-year-old error, and it is fixable this quarter.

> *An instrument that asks customers what the worst part was should be able to say what the worst part of the instrument is. Section 9 is our attempt. If we have missed one, we would like to hear it, and we will publish it.*

CONTINUE THE CONVERSATION

For more information on WXI, visit wxi.app.

Notes

1 Gates, B. (1999). Business @ the Speed of Thought: Succeeding in the Digital Economy. Warner Books. Attribution confirmed via Oxford Reference.

2 Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain, 66(1), 3–8. DOI: 10.1016/0304-3959(96)02994-6. Retrospective judgements correlated with peak intensity (P < 0.005) and with intensity during the final three minutes (P < 0.005).

3 Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. Journal of Personality and Social Psychology, 65(1), 45–55.

4 Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. Pain, 104(1–2), 187–194. DOI: 10.1016/S0304-3959(03)00003-4. See also Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: adding a better end. Psychological Science, 4(6), 401–405.

5 Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. Econometrica, 47(2), 263–291.

6 On the founding claim: Reichheld, F. F. (2003). The one number you need to grow. Harvard Business Review, 81(12), 46–54. On its replication record, two independent teams: Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of Net Promoter and firm revenue growth. Journal of Marketing, 71(3), 39–51. DOI: 10.1509/jmkg.71.3.039. Using 21 firms and more than 15,500 interviews from the Norwegian Customer Satisfaction Barometer, and working inside the industries Reichheld cited as exemplars, they failed to replicate the assertion of clear superiority. The paper won the 2007 Marketing Science Institute / H. Paul Root Award. And Morgan, N. A., & Rego, L. L. (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. Marketing Science, 25(5), 426–439. DOI: 10.1287/mksc.1050.0180. Using ACSI data linked to COMPUSTAT and CRSP measures across 1994–2000, average satisfaction and top-two-box measures carried predictive weight while their net-promoter construct did not reach significance. Two caveats we would rather raise ourselves: a version of the Keiningham analysis was later republished under a different title, so the 2007 and 2008 papers are not independent confirmations of each other; and Morgan and Rego built their net-promoter construct from ACSI items rather than the canonical likelihood-to-recommend wording.

7 Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460–469. DOI: 10.2307/3150499.

8 On the gap-model approximation: Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12–40.

9 Harzing, A.-W. (2006). Response styles in cross-national survey research: a 26-country study. International Journal of Cross Cultural Management, 6(2), 243–266. DOI: 10.1177/1470595806066332.

10 Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. Journal of Consumer Research, 19(4), 655–662. DOI: 10.1086/209329.

11 Cronin, J. J., & Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. Journal of Marketing, 56(3), 55–68. DOI: 10.1177/002224299205600304. Replicated in Brady, M. K., Cronin, J. J., & Brand, R. R. (2002). Performance-only measurement of service quality: a replication and extension. Journal of Business Research, 55(1), 17–31.

12 Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science, 185(4157), 1124–1131.

Citations verified against primary or authoritative secondary sources during drafting, including volume, issue, pages and DOI, for Redelmeier and Kahneman (1996), Redelmeier, Katz and Kahneman (2003), Keiningham et al. (2007), Morgan and Rego (2006), Oliver (1980), Cronin and Taylor (1992), Peter, Churchill and Brown (1993) and Harzing (2006). The Gates attribution is confirmed via Oxford Reference. The remaining references are standard and widely cited.