



PERSONALIZATION

How Measurement is Evolving in Next-Best Action

By [Silvio Palumbo](#), [Yun Lim](#), and [Megha Ramanuja](#)

ARTICLE JULY 14, 2026 12 MIN READ

This is Part 5 of a five-part series on next-best action.

Proving value realization is the most underestimated challenge for next-best action (NBA) programs. Simplistic measurement approaches focus on activation, such as offers delivered, open rates, and win rates. But without more sophisticated incrementality measurement, organizations can't clearly distinguish between NBAs that positively influenced customer behavior and actions that had a neutral or even cannibalistic effect on the customer's intent.

Our research and direct experience indicate that when organizations adopt and deploy rigorous incrementality testing, they typically find that 20% to 40% of their active programs deliver marginal to negative lift. Although the programs elicit a customer response, they redirect demand, rather than creating it.

As NBA evolves from the marketer-defined campaign to the individually composed action, the measurement paradigm must adapt accordingly. Although this has been true for the past decade, agentic marketing demands further rigor in measuring incremental impact even as it offers powerful new levers for implementation. This article maps the evolution of measurement across three levels of maturity and introduces the frameworks needed to effect change.

Three Levels of Measurement Maturity

Most organizations’ measurement capabilities focus on campaign metrics, while a few have progressed to incrementality testing. Eventually, the best will adopt agentic measurement. (See Exhibit 1.) The progression yields better tools and more data, but it also entails a fundamental shift in the questions that measurement seeks to answer.

EXHIBIT 1
Measurement Maturity Progresses Across Three Levels

	Where most organizations are today Level 1 Campaign metrics	Where the market is moving Level 2 Incrementality testing	The emerging frontier Level 3 Agentic measurement
Core question	“What happened?”	“Did we cause it?”	“Why, and what should we do?”
What it does	Correlation	Causation	Explanation and action
Key methods	Last-touch attribution	Regression-based Big M/Little m framework	Causal explanation and agent-designed experiments
Typical insights	Email B had a 3.2% click-through rate	The holiday campaign drove 12% incremental sales vs. holdout	Offer X drove an 8% lift in segment Y due to price sensitivity
Key limitation	Can’t distinguish correlation from causation	Proves what worked, but can’t explain why or adapt autonomously	Requires thorough governance frameworks

Source: BCG analysis.

Level 1: Campaign Metrics

Today, the vast majority of organizations have access to metrics such as response rates, conversion rates, click-through rates, program-level ROI, and last-touch attribution. These indicators are necessary, simple to compute, and easy to report. Crucially, however, they identify

engagement but not incrementality: customers who received the offer converted at a higher rate than those who did not.

In fact, last-touch attribution compounds the problem by assigning full credit to the action that happened to precede the conversion, regardless of whether it actually influenced the outcome. This conflates selection effects with treatment effects. The customers selected for an offer were already more likely to convert, which is precisely why the model chose them. Level 1 measurement tells you what happened following your actions, not whether your actions caused it to happen.

Level 2: Incrementality Testing

Organizations that set up randomized holdout groups, structured A/B tests, and measure impact at the program level can begin to answer causal questions such as “Did this program change behavior relative to doing nothing?”

The challenge of getting to Level 2 is often organizational. For example, introducing randomized holdout groups means deliberately not marketing to these customers—an unavoidable opportunity cost of measurement—which creates tension with business teams that focus on short-term revenue. For similar reasons, different business units that target the same customers may suspend their marketing outreach if it proves not to be accretive. Without holdouts, organizations are flying blind. Incrementality measurement can prove that a program delivers positive ROI. But if the sample size and budget are inadequate, even incrementality testing may not conclusively attribute outcomes to individual actions within the program.

Level 3: Agentic Measurement

Incrementality testing in Level 2 contributes to Level 3, which encompasses agent-interpreted evidence, agent-generated explanations, and agent-executed optimization. At Level 2, a human analyst runs the incrementality test, interprets the results, and briefs the team on what to change. At Level 3, an agent will read the same causal evidence and perform a read-and-react cycle at a scale and speed that individual marketers can’t match, explaining why something worked (“in a retention window, customers with high lifetime value responded to value-reinforcement messaging, not discount messaging”), deciding what to change (adjust shelf weighting, update targeting parameters, and/or reallocate budget within the batch optimization layer), and acting on that decision within the system’s governance constraints. The human’s role shifts from interpreting and deciding to reviewing and governing.

Answer Questions at Every Tier

Measurement accuracy is critical across all three tiers of marketing decisions. (See Exhibit 2.) Strategic decisions determine where to invest: “Should we shift budget from retail banking to credit cards?” Tactical decisions determine what to invest in: “Which channels should we expand?” Operational decisions determine how to optimize execution: “Does this creative outperform that one?”

EXHIBIT 2

Measurement Must Answer Questions at Every Decision Tier



Fundamental question	Decision KPIs	Example question
Strategic Where should we spend?	Portfolio-level ROI, lifetime value	How do we split budget across various life-cycle stages?
Tactical What should we spend on?	Channel-level ROI	How do we split budget across different campaigns and channels?
Operational How can we optimize it?	Tactic-level KPI, return on spending	How do we optimize performance within a channel?

Source: BCG analysis.

Organizations at Level 1 of measurement maturity are limited almost entirely to performing measurement at the operational tier: platform metrics, click-through rates, and A/B test results on creative variants. These are essential for tuning execution, but they can’t fully answer tactical and strategic questions. Measurement infrastructure needs to evolve to more comprehensively include “Big M” and “Little m” approaches. (See Exhibit 3.)

EXHIBIT 3

The Measurement Toolkit Is a Four-Legged Stool



Source: BCG analysis.

Big M measurement addresses the program-level question, “Does the entire NBA system create value relative to no intervention at all?” Developing a rigorous answer to this question requires

using a global holdout—a randomly selected group of customers who serve as a control group, receiving no agent-composed actions and thus establishing the counterfactual for the program as a whole. Big M validates the system’s financials to a degree of accuracy necessary to earn a CFO’s signoff. Unfortunately, Big M often overstates program-level impact due to a combination of the halo effect, the pull forward effect, experiment contamination, and operational errors. The global holdout also aligns incentives across marketing, operations, and analytics by providing a single, unambiguous measure of program success.

Little m measurement addresses the tactic-level question, “Does this specific action create incremental value relative to the next best alternative?” It arrives at an answer by setting up statistically significant local holdouts within the system—breaking out customers in the test audience into subgroups and randomly assigning them to different actions or to a default experience. Little m enables granular optimization by identifying which offer drives more incremental value for a specific segment and a particular construct/campaign. It also resolves the question of which channel delivers a better lift for that customer cohort. The subgroup structure permits individualized test-and-control research at scale without contamination across experiments.

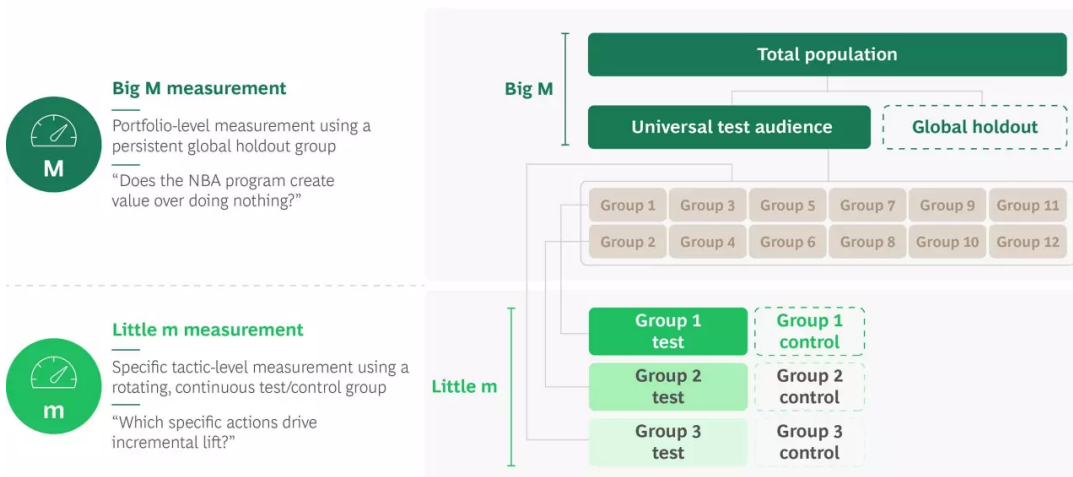
Incorporating measurement that spans strategic, tactical, and operational tiers will become increasingly critical as organizations adopt Level 3 agent-native systems. These agents make decisions that span all three tiers simultaneously, and they need highly accurate measurement to guide them.

Breaking Down the Measurement Toolkit

Addressing strategic, tactical, and operational decisions requires a combination of four complementary approaches: in-platform metrics, modeling, customer insights, and experimentation. Each approach answers a different question on a different time horizon. (See Exhibit 4.)

EXHIBIT 4

Big M Validates the System, and Little m Optimizes It



Source: BCG analysis.

Note: NBA = next-best action.

In-Platform Metrics

In-platform metrics provide real-time monitoring. Response rates, engagement scores, and delivery metrics are necessary for operational health because they provide critical insight into engagement and real-time interaction. But in isolation they are insufficient for making fully informed decisions and measuring incremental impact.

Modeling

Modeling provides estimates of incremental short-term ROI. For example, media mix modeling quantifies the contribution of each marketing channel to overall outcomes—input that can help inform optimal budget reallocation decisions. Uplift models assess the true incremental impact of specific actions on individual customers. Multitouch attribution models distribute credit across touchpoints in the customer journey.

Among the most rigorous approaches for estimating incremental effects are causal/uplift models such as T-learners (separate outcome models for treatment and control groups), S-learners (single models with treatment as an input feature), and doubly robust learners (models that combine outcome modeling with propensity weighting to reduce bias from either component). The doubly robust approach is particularly valuable because it provides consistent estimates even when one of its two components is misspecified.

Customer Insights

Customer insights measure long-term impact. Brand tracking, customer satisfaction surveys, and cohort analyses capture effects that short-term metrics miss. Is the NBA program building brand equity or eroding it? Is personalization increasing customer lifetime value or just pulling purchases forward?

Experimentation

Experimentation provides robustness to incrementality measurement. Randomized holdout groups, structured A/B tests, and geo-experiments provide causal evidence that modeling alone cannot guarantee. Design of experiment provides a broad range of options—matched markets, synthetic controls, stratified sampling, and so on—to definitively determine whether an action can change behavior. Experimentation can be technically complex in specific situations (for example, how do you experiment with different variants during a live broadcast?) and organizationally challenging with regard to coordination and alignment on issues such as the opportunity cost of not exposing select customers to actions for the sake of performance measurement.

Most organizations struggle to harmonize all four elements of the toolkit, and many lean heavily on in-platform metrics. But it is imperative to operate all four continuously and triangulate the results to adjust short-term and long-term strategies.

Common Measurement Pitfalls

Operational errors in measurement are fairly common and addressable. The most challenging obstacles tend to involve governance and coordination. Even organizations that invest in incrementality testing encounter pitfalls that can undermine the credibility and utility of their results. Five hazards are especially noteworthy:

- **Holdout Contamination.** The global holdout operates on the assumption that the holdout group receives no treatment. In practice, holdout customers often receive communications through channels that the NBA system doesn't control, such as branch staff making outbound calls, retail associates offering promotions, and paid media campaigns running in parallel. Every uncontrolled touchpoint dilutes the holdout and biases the measurement toward underestimating program impact. Organizations need to rigorously establish and enforce who can contact holdout customers and through which channels.
- **Novelty Effects.** New programs tend to show inflated results early on, because customers often respond to the novelty of a new type of communication or offer, rather than to the

underlying value proposition. As a result, organizations that measure only during the launch window are likely overestimate long-term program value. Effective measurement requires letting programs run past the novelty period—typically 8 to 12 weeks, depending on purchase cycle—before drawing conclusions.

- **Survivor Bias.** Long-running programs accumulate survivor bias. The customers who remain in the program are disproportionately those who respond well to it. Over time, customers who are annoyed by overcommunication or poorly targeted offers churn out of the relationship entirely. Consequently, measuring only among survivors overstates program effectiveness. The global holdout can correct survivor bias—but only if the organization maintains the holdout (and finds proper refresh approaches) from the program’s inception.
- **Political Factors.** The most underestimated measurement challenge is organizational. When incrementality testing reveals that a flagship program delivers limited incremental lift, program owners have a strong incentive to challenge the methodology rather than accept the finding. This is the most common backlash following first-time incrementality testing. Leaders who anticipate this dynamic can look for ways to build credibility for the measurement framework before it delivers uncomfortable news. For example, the team could start by measuring a program where it expects positive results, thereby demonstrating that the methodology works before applying it to politically sensitive programs.
- **Budget Tension.** Every percentage point of traffic allocated to exploration (testing new actions or maintaining holdouts) is a percentage point excluded from being optimized for short-term revenue. Business leaders who are compensated on quarterly metrics will push to minimize holdout sizes and reduce exploration budgets. An organization’s chief data and analytics officer (CDAO) and chief marketing officer (CMO) must frame exploration as an investment in information, not a cost to revenue. Organizations that explore more thoroughly will learn faster, and the compounding returns of faster learning will overcome the short-term cost within quarters.

Looking Ahead: From Human Measurement to Autonomous Optimization

Most organizations don’t need to leapfrog to Level 3 immediately. Understanding how measurement can evolve is essential for building infrastructure that scales as the agentic future emerges. The transition from Level 2 to Level 3 will involve three major shifts:

- **From Human-Designed Experiments to Agent-Designed Experiments.** At Level 2, a marketing specialist or data scientist must define the holdout groups, the test/control structure, and the subgroup segmentation. This works for a finite set of actions and segments. When agents compose thousands of unique action sequences from a shelf of hundreds of components, however, specialists can't scale the design of predefined experiments for every possible combination. Traditional A/B testing also begins to break down under the sheer volume of permutations, as traffic and sample sizes become too fragmented to generate statistically reliable results across all combinations. The experimental design itself must therefore become autonomous, with agents dynamically allocating exploration traffic on the basis of where the highest-value information lies and increasingly relying on adaptive techniques such as multi-armed bandits and causal learning approaches rather than static test/control frameworks.
- **From Interpretation to Explanation.** Today, an analyst looks at incrementality results and interprets what they mean for the business. An agentic measurement system will generate that explanation directly, identifying not just that an action worked, but why it worked, which customer segments drove the lift, and what that finding implies for future targeting. Techniques such as Shapley values (which decompose credit across a sequence of actions) and structural causal models provide the analytical foundation. The agent will then translate these results into natural-language explanations that marketing leaders can act on.
- **From Insight to Action.** The most consequential shift involves closing the loop between measurement and decisioning. At Level 2, there is a handoff between teams: the measurement team produces a finding and presents it to the marketing team, and the marketing team decides what to change. That handoff introduces days and in many cases weeks of latency. At Level 3, the agent acts on evidence as it accumulates: adjusting shelf composition, updating targeting parameters, and reallocating budget across actions, all within governance boundaries set by the marketing team. The time from insight to action compresses from weeks or days to hours. This is where the vibe marketer described in the previous article meets the measurement. The marketer sets the objectives and guardrails, and the agent handles the decisioning and the measurement-driven optimization in a continuous loop.

The Measurement-Led Organization

Elevating the role and sophistication of measurement pays off immediately by revealing which programs deliver incrementality and which are harvesting demand that would have materialized anyway.

The minimum starting point for this inquiry is a combination of in-platform metrics and Big M measurement. Setting up a global holdout to validate the broader NBA program establishes the rigor needed to focus on incremental value. This approach does not entail new technology or investment, but it does require organizational discipline and internal coordination across campaigns, across BUs, product lines, and channels. The resulting backbone sets the right tone for the organization to continue building the four-legged stool and the subgroup structure for Little m experimentation.

The result is a strategic advantage. By developing a measurement-led organization, CDAOs and CMOs enable the entire organization to make better decisions, faster. When the CMO asks, “Should we increase marketing spending next year?” the answer comes from causal evidence, not extrapolated dashboards. Measurement becomes the connective tissue between strategy and execution.

This is Part 5 of a five-part series on the future of next-best action. [Part 1](#) explores why most NBA programs underdeliver and identifies four structural gaps. [Part 2](#) explores the shift from marketer-orchestrated journeys to agent-composed actions. [Part 3](#) unpacks the decisioning science, from propensity models to contextual bandits to agent-based reasoning. [Part 4](#) examines how marketing organizations restructure around the new operating model, introducing the vibe marketer archetype.

Authors



Silvio Palumbo

Managing Director & Senior
Partner
New York



Yun Lim

Associate Director, Marketing
Chicago



Megha Ramanuja

Managing Director & Partner
Los Angeles



ABOUT BOSTON CONSULTING GROUP

Boston Consulting Group bridges the gap between ambition and outcomes for the world's leading companies and organizations. We are built for this era of unprecedented change — bringing strategic clarity rooted in over 60 years of deep domain knowledge, combined with applied AI shaped by our practitioners. BCG works shoulder-to-shoulder with CEOs across industries and geographies to deliver transformative impact at scale: stronger returns, transferred capabilities, and change that sticks. For more information, visit bcg.com.