Toward a Standard for Measuring Customer Sentiment and Worker Service Quality in Health and Human Services

A whitepaper authored by Global True North and Tandem Research

Authors: Global True North and Tandem Research

Date: February 2026

Executive Summary

Health and Human Services (HHS) programs increasingly recognize that service quality is not a decorative aspiration. It is a determinant of access, compliance, and administrative efficiency. Yet across the United States, HHS agencies often measure customer experience unevenly, using local survey instruments, informal supervisory observation, or narrow operational metrics that do not reveal how the experience of an interaction rises or falls in real time. This whitepaper proposes a practical starting standard for measuring customer sentiment and worker service quality across HHS systems. The standard draws on nationwide research into best practices from service industries with mature measurement regimes, established public-sector customer experience practice, and the health sector's validated experience-of-care measurement traditions. It offers a two-layer measurement model, a reusable criteria library, and a tiered scoring framework that agencies can adopt, compare, and improve over time.

> *The recommendations here are intended to be guideposts and suggestions based on GTN's national-level research and knowledge of best practices. Every organization is different, and the implementing organization should adapt the strategy presented here to its own needs and priorities.*

1. Why a unified sentiment standard is needed in HHS

HHS interactions are frequently high stakes, emotionally charged, and constrained by policy. A family seeking housing assistance, a beneficiary navigating Medicaid eligibility, a parent calling child support, or a participant seeking disability services often arrives with stress, distrust, and limited time. In these circumstances, the quality of an interaction influences whether a person follows through, returns with repeat contact, files a complaint, disengages, or escalates. It also influences workforce stability, because repeated conflict, emotional strain, and ambiguity erode worker morale and retention.

Despite this, many agencies lack a consistent language for describing service quality. Some track broad satisfaction, some track call volume and handle time, and some track error rates, but few link these into a coherent system that distinguishes what a customer experienced from what a worker did. A unified standard does not require agencies to become identical. It provides a comparable core so that local context can be honored while measurement remains interpretable and improvable.

2. How this standard was developed

Global True North and Tandem Research developed this proposed standard by synthesizing practices from sectors that have been compelled to measure sentiment with rigor, then translating those practices to the realities of HHS. The selection logic was pragmatic. We prioritized industries where service outcomes depend on short interactions, where performance can be observed directly, and where measurement has been operationalized into coaching and governance rather than left as a branding exercise. We reviewed the measurement families those sectors rely on, the operational rubrics they use for quality assurance, and the ways they turn measurement into improvement.

Across sectors, the same measurement architecture repeatedly surfaced. Mature programs pair a small number of outcome measures with a behavior rubric and an evidence-based review loop. They treat sentiment as a trajectory rather than a single label, and they calibrate scoring so results are comparable across evaluators. In translating these practices to HHS, we made four design choices. First, we separated experience from worker behavior to protect fairness and improve diagnostic value. Second, we constrained the criteria set to dimensions that are stable across programs and channels, including eligibility, benefits navigation, enforcement, case management, and in-person or field service. Third, we designed scoring to work with human review today, and later to be supported by analytics tooling without changing the core standard. Fourth, we embedded interpretability, so that any score can be traced to evidence and used for coaching rather than punitive surveillance.

3. What leading sectors teach about sentiment measurement

Sectors known for measuring sentiment well share a disciplined premise: measurement is only valuable when it is operational. Contact centers and customer support operations treat service quality as a system, pairing outcome metrics with calibrated quality rubrics and interaction analytics. Retail and hospitality excel at rapid feedback loops and service recovery. Financial services demonstrate how to separate empathy and compliance, then govern both. Subscription and digital support environments show the value of tracking experience across journeys rather than as isolated events. This section summarizes what these sectors do well, and why those practices translate to HHS.

3.1 Contact centers and business process operations

Contact centers are the most directly relevant comparator for many HHS environments, particularly call centers, eligibility lines, and high-volume service desks. Their measurement strength rests on disciplined quality assurance. They define observable behaviors, sample interactions, calibrate evaluators, and create closed loops where measurement leads to specific coaching and process fixes. External frameworks such as ISO 18295 and the COPC Customer Experience Standard reflect this mindset by emphasizing repeatable processes, documented quality management, and corrective action.

They also pioneered speech and text analytics that quantify sentiment at scale. The most useful output of these systems is rarely the absolute score. It is the trend line: when sentiment shifts, what triggered it, and what agent behaviors restored stability. In HHS, this trajectory framing matters because customers often begin with frustration or fear, and the aim of an interaction is frequently stabilization, clarity, and trust, even when the substantive outcome is constrained.

3.2 Retail, hospitality, and service recovery cultures

Retail and hospitality measure sentiment well because service consistency and recovery after failure are central to competitiveness. Their dominant practice is rapid feedback paired with concrete recovery actions, which creates a habit of learning from dissatisfaction rather than arguing with it. Their most transferable lesson is that sentiment is shaped by perceived effort and perceived fairness as much as by warmth. A respectful tone helps, but clarity, predictability, and realistic expectations often determine whether a customer leaves calmer or more agitated.

3.3 Financial services and regulated support

Banks and insurers share important constraints with HHS: strong regulation, high stakes, and customers who may be anxious or angry. Their mature programs separate two questions that are often conflated in public service settings. One question is whether the worker was empathetic and clear. The other is whether the worker was correct and compliant. Good systems score both, then govern both. This approach is especially relevant to HHS because fairness to workers requires acknowledging that a correct answer can still feel dissatisfying, and a pleasant interaction can still fail if it is inaccurate or misleading.

3.4 Digital product support and journey measurement

Digital product and subscription businesses popularized journey-level measurement, linking experience to repeat contact, abandonment, and long-run relationship outcomes. Their contribution is methodological clarity: a few stable metrics, applied consistently, can reveal where friction accumulates and where trust is built or lost. For HHS, this suggests that sentiment measurement should not be confined to the end of a call. It should also capture whether an interaction reduces future effort by making the next step understandable and actionable.

4. Anchoring HHS measurement in validated public and health-sector practice

A proposed HHS sentiment standard should not reinvent validated measurement where it already exists. In the health domain, the Consumer Assessment of Healthcare Providers and Systems (CAHPS) family provides a widely used, survey-based approach to measuring experience of care, including instruments designed for Medicaid populations and for home and community-based services. In the broader public sector, the federal customer experience agenda emphasizes service delivery through the lens of the customer, with explicit attention to trust and human-centered design.

These traditions suggest a pragmatic posture. HHS should use survey-based outcome measures where feasible, but it should also measure interaction quality directly, because many HHS services have limited survey response rates and high variability by channel. A standardized approach therefore pairs outcomes with in-interaction evidence.

5. The proposed HHS standard: two layers, one shared language

We propose a two-layer standard that agencies can implement progressively. The first layer is an **Experience Layer**, describing what the customer is likely feeling and perceiving through the interaction. The second layer is a **Worker Behavior Layer**, describing what the worker did that plausibly shaped those perceptions.

The Experience Layer should be evaluated as a trajectory, not merely as a final label. The Worker Behavior Layer must be observable, teachable, and calibratable across supervisors. When both layers are present, agencies can separate two realities that are often confused: a customer may remain dissatisfied because of structural constraints, and a worker may still have performed skillfully and ethically. Conversely, a superficially pleasant interaction can still fail if it leaves the customer confused, misinformed, or unable to act.

6. The criteria library

A national standard requires a small set of criteria that works across programs and channels. We recommend ten Experience dimensions and nine Worker Behavior dimensions. Each criterion is scored from 1 to 5, using the guidance below. The Experience total is the sum of ten items for a maximum of 50. The Worker Behavior total is the sum of nine items for a maximum of 45.

Table 1 presents the ten **Experience dimensions** (what the customer is feeling and perceiving), with definitions, evidence signals, improvement drivers, and scoring guidance. Table 2 presents the nine **Worker Behavior dimensions** (what the worker did), with descriptions of good practice, observable signals, the reason each behavior matters, and scoring guidance.

Table 1. Experience dimensions (customer sentiment and perception)

Dimension	Definition	Typical evidence signals	Common improvement drivers	Scoring guidance (1–5)
Valence	Overall positive, neutral, or negative affect.	Gratitude, relief, calm agreement versus complaints, hostility, sarcasm.	Clear explanations, respectful tone, credible problem-solving.	1 = openly negative/hostile; 3 = mixed/neutral; 5 = clearly positive/relieved.
Emotional intensity	Strength of emotion, regardless of direction.	Interruptions, raised voice, urgency markers, repeated exclamations.	Acknowledgment, de-escalation, expectation-setting.	1 = highly activated; 3 = moderate activation; 5 = calm, regulated.
Effort burden	How hard it feels to get help.	Repeating information, long waits without clarity, "I already told you."	Streamlined steps, fewer transfers, summaries and confirmations.	1 = exhausting/repetitive; 3 = some friction; 5 = easy, minimal repetition.
Perceived fairness	Whether the process feels consistent and just.	"That's not fair," suspicion of bias, comparisons to others.	Transparent reasoning, consistent application, respectful explanation of constraints.	1 = unfair/bias concerns; 3 = uncertain; 5 = feels consistent and justified.
Trust and credibility	Confidence in the worker and the system.	"You said last time," threats to complain, skepticism.	Reliability cues, documentation confirmation, precise next steps.	1 = distrust/suspicion; 3 = cautious; 5 = confident in information and follow-through.
Comprehension	Understanding of what is happening and why.	Confusion statements, incorrect restatements, circular questions.	Plain language, teach-back, jargon reduction.	1 = confused; 3 = partial understanding; 5 = can restate plan and rationale.
Control and agency	Sense of options and ability to act.	Helplessness, resignation, "so I can't do anything."	Bounded choices, clear owner and timeline, actionable steps.	1 = powerless; 3 = some options; 5 = clear choices and next steps.
Resolution progress	Perceived movement toward an answer.	Reduced repetition, acceptance of a plan, fewer "what now."	Structured workflow, milestones, realistic expectations.	1 = stuck; 3 = partial progress; 5 = clear movement and closure.
Respect and dignity	Feeling treated as a person, not a problem.	Defensiveness after tone shifts, appreciation for being heard.	Active listening, nonjudgment, trauma-aware language.	1 = feels disrespected; 3 = adequate; 5 = explicitly feels heard and respected.
De-escalation success	Whether intensity decreases after a peak.	Shift to collaboration, slower cadence, fewer interruptions.	Empathy plus structure, firm boundaries, clear closing.	1 = escalates/worsens; 3 = stabilizes somewhat; 5 = clearly calms by end.

Table 2. Worker behavior dimensions (service quality rubric)

Behavior	What good looks like	Observable signals	Why it matters	Scoring guidance (1–5)
Sets expectations early	States what will happen, what will not happen, and typical timing.	"Here is what I can do today," "This may take about…"	Reduces effort burden and prevents surprise-driven escalation.	1 = no expectations set; 3 = partial; 5 = clear scope and timing early.
Uses plain language	Explains terms and avoids procedural jargon.	Short sentences, definitions, concrete examples.	Improves comprehension and perceived fairness.	1 = jargon-heavy; 3 = mixed; 5 = consistently plain, accessible language.
Active listening and reflection	Names the issue and reflects emotion without blame.	"What I hear is…" "That sounds stressful."	Improves respect and lowers intensity.	1 = dismissive/interrupting; 3 = some reflection; 5 = consistent reflection and validation.
Explains constraints transparently	Connects rules to purpose and options.	"By policy we must…" with rationale and next steps.	Protects fairness and trust in constrained contexts.	1 = opaque/defensive; 3 = partial; 5 = clear rationale plus options.
Offers bounded choices	Provides realistic options without false promises.	"We can do A or B," "Here are two next steps."	Increases agency and reduces helplessness.	1 = no options; 3 = limited; 5 = clear, realistic choices.
Confirms understanding	Checks comprehension and corrects gently.	Teach-back prompts, confirmation questions.	Reduces repeat contact and later disappointment or betrayal feelings.	1 = no checks; 3 = one check; 5 = confirms key points and adjusts as needed.
Summarizes next steps and ownership	Closes with an actionable plan and who owns what.	"You will do X, we will do Y, by Z date."	Improves resolution progress and credibility.	1 = vague close; 3 = partial summary; 5 = specific actions, owner, timeframe.
Maintains respectful tone under stress	Stays calm, avoids sarcasm and defensiveness.	Even cadence, nondefensive language, clear boundaries.	Prevents escalation spirals and supports dignity.	1 = reactive/escalatory; 3 = mostly calm; 5 = consistently respectful and steady.
Documentation and follow-through cues	Signals recording, confirms receipts, clarifies future contacts.	"I am noting this now," "You will receive…"	Builds trust and reduces downstream confusion.	1 = no cues; 3 = minimal; 5 = clear documentation and follow-through statements.

7. Tiering and reporting

A standard becomes usable when it produces scores that are comparable, interpretable, and actionable. We recommend computing two separate totals for each reviewed interaction: an Experience total (maximum 50) and a Worker Behavior total (maximum 45). Each total is mapped into three tiers by dividing the maximum into thirds. Tiering reduces the temptation to treat a score as an objective truth about a person, while still supporting operational decisions and clear communication to leadership.

Table 3. Recommended tier thresholds for Experience and Worker Behavior totals

Tier	Experience total (10 items, max 50)	Worker behavior total (9 items, max 45)	Experience interpretation	Worker behavior interpretation
Tier 3: At-risk	1–16	1–15	High effort, low trust, confusion, or perceived unfairness; sentiment trajectory worsened or failed to stabilize.	Core stabilizing behaviors absent or inconsistent; risk of repeat contact and escalation is high.
Tier 2: Stable	17–33	16–30	Partial clarity and progress; negative sentiment may be present but bounded and somewhat stabilized.	Most core behaviors present, with gaps that limit improvement in sentiment trajectory.
Tier 1: Strong	34–50	31–45	Clarity, dignity, and credible progress; sentiment improved or de-escalated even under constraints.	Behaviors consistently present, including expectation-setting, clarity, and actionable closing; performance supports trust and reduces repeat contact.

Tiering should be paired with evidence. Scores should be accompanied by the two or three moments that most influenced the result, expressed as brief excerpts or time stamps. This practice improves calibration, makes coaching concrete, and protects fairness by ensuring that a score is never divorced from context.

8. Implementation guidance for HHS agencies

Implementation should proceed in stages. First, establish a calibrated quality assurance process using the Worker Behavior rubric on a sampled set of interactions. Second, add Experience scoring as a trajectory, beginning with transcript-based review and expanding to audio cues where appropriate. Third, integrate outcome measures where feasible, including short post-interaction questions that mirror validated domains of satisfaction and effort.

Governance is essential. Agencies should define how scores will be used, who can see them, and how to prevent misuse. High-performing sectors emphasize that sentiment scoring is a tool for improvement, not a substitute for supervision or a blunt instrument for discipline. HHS should also define checks, ensuring that language access needs, disability accommodations, and differing cultural communication norms are considered in calibration.

9. Why a unified framework can be useful to HHS agencies

A unified framework is useful because it makes quality visible in a way that is comparable without being reductive. When agencies share a common vocabulary for experience and behavior, they can benchmark across regions, programs, and channels without pretending that every program is identical. This comparability supports practical governance: leadership can distinguish systemic friction from isolated performance issues, and can prioritize process improvements that reduce customer effort at scale.

The framework is also a workforce tool. A stable rubric supports clearer onboarding, more consistent coaching, and fairer evaluation because workers are assessed on observable behaviors rather than on diffuse impressions. Tiering simplifies communication for managers while preserving nuance through the accompanying evidence excerpts. Over time, agencies can track whether training investments move interactions from at-risk to stable and from stable to strong, and can see where specific behaviors are driving sentiment improvement.

Finally, a unified standard can strengthen public accountability. HHS systems are often asked to demonstrate that they deliver services with dignity and reliability. A shared measurement approach creates a credible starting point for that demonstration. It can also support procurement and technology governance by giving agencies a clear set of requirements for tools that claim to measure experience, ensuring that any analytics or automation remains anchored to an intelligible, human-reviewable standard.

References and foundational sources

Centers for Medicare & Medicaid Services (CMS). *Consumer Assessment of Healthcare Providers & Systems (CAHPS)*. https://www.cms.gov/data-research/research/consumer-assessment-healthcare-providers-systems

Medicaid.gov. *CAHPS Home and Community-Based Services Survey*. https://www.medicaid.gov/medicaid/quality-of-care/quality-of-care-performance-measurement/cahps-home-and-community-based-services-survey

Performance.gov. *Federal Customer Experience (CX)*. https://www.performance.gov/cx/

Executive Order 14058 (December 13, 2021). *Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government*. https://www.federalregister.gov/documents/2021/12/16/2021-27380/transforming-federal-customer-experience-and-service-delivery-to-rebuild-trust-in-government

ISO 18295-1:2017. *Customer contact centres, requirements for service provision*. https://www.iso.org/standard/64739.html

COPC Inc. *COPC Standards for customer experience operations*. https://www.copc.com/copc-standards/