Effort, measured for the agentic era: introducing the Agentic Customer Effort Score

We placed 163 automated calls to 19 companies' support lines and reviewed 21 large online retailers' websites the way a customer's AI agent would. This is what ACES measures, what we found, and what we don't know yet.

TELEPERSON RESEARCH · TELEPERSON.COM/RESEARCH/ACES

KEY FINDINGS

- **Seven of 11 large retailers' support lines** now answer with an AI, virtual or automated assistant. None of the 8 other companies we mapped do.
- Our calls found a **route to a person on 5 of 11** retailer lines and 2 of 8 others. Where a route exists, it took about 38–74 seconds of menu, before any hold.
- **9 of 21 retailer websites** refused or challenged our automated browser; three blocked every page.
- Six of the 18 retailer sites we could view **showed no phone number** to a signed-out visitor, and Amazon sent every contact option to sign-in.
- Of ten retailers scored on all four parts, **Home Depot leads with 94** of 100. Most others land in "high effort", partly because of what we could not yet verify.

Why now

For as long as companies have measured customer service, they have measured it by asking people afterwards. How satisfied were you? Would you recommend us? How easy was it? Those questions still matter. But two things are changing who, or what, is on the customer's side of the conversation.

The first is that service itself is being automated. In our audits, seven of the 11 large online retailers whose support lines we called answered with a voice that introduced itself as an AI, virtual or automated assistant. Customers have noticed: in a 2024 Gartner survey, 64% of customers said they would prefer companies didn't use AI for customer service, and the top concern was that it would make it harder to reach a person.[3] In 2026, 87% said companies using generative AI for service must provide access to a human.[4]

The second is that customers are starting to send AI agents to deal with companies for them: software that can read a help center, fill in a form or place a call. An agent doesn't answer satisfaction surveys. When it meets a bot wall, a sign-in page or a menu that only accepts key presses, it simply fails, and the customer is back where they started, with the effort still to spend.

Effort is worth measuring because it compounds. In the research that introduced the Customer Effort Score, 57% of inbound calls came from customers who had gone to the website first.[1] Gartner found only 14% of customer service issues are fully resolved in self-service[5], and that service interactions are nearly four times more likely to lead to disloyalty than to loyalty.[6] Patience is short: in vendor surveys, 41% of people said they hang up after one to two minutes on hold[8], and 29.9% said they would switch brands rather than be put on hold.[9] Meanwhile Forrester's US CX Index fell for a fourth consecutive year in 2025, to 68.3 out of 100.[7]

We think the industry needs a measure of effort that can be taken from the outside, without a survey, and that counts the obstacles an AI agent meets as well as the ones a person meets. That is what the Agentic Customer Effort Score (ACES) is for.

A short history of service metrics

Customer satisfaction (CSAT) surveys came first and are still the most common. In 2003, Fred Reichheld introduced the Net Promoter Score in "The One Number You Need to Grow".[2] In 2010, Dixon, Freeman and Toman argued, from a survey of more than 75,000 people, that companies should stop trying to delight customers and instead reduce the effort it takes to get help, which gave rise to the Customer Effort Score.[1] Inside the contact center, operations teams track first-contact resolution, average handle time and, as automation grew, containment or deflection.

Service metrics in common use, in our words.

Metric	Origin	What it measures	What it misses
CSAT	Post-contact survey	How satisfied a customer says they were with an interaction.	Only hears from people who reached someone and answered the survey; it measures a feeling, not the work.
Net Promoter Score	2003	How likely a customer is to recommend the company, on a 0–10 scale.	A view of the whole relationship; it can't point to the step in service that went wrong.
Customer Effort Score	2010	How much effort a customer says it took to get an issue handled.	Still a survey of people, after the fact, and only of the ones who got through.
First-contact resolution	Operations	The share of issues solved on the first contact.	Counted inside the company: contacts that never arrive are invisible to it.
Average handle time	Operations	How long an agent spends on each interaction.	Efficiency for the company; a short call is not the same as an easy one.
Containment / deflection	Automation	The share of contacts kept in self-service or automation.	A customer who gives up counts as "contained".

Each of these is useful. But each is either a survey of the people who made it through, or a count kept inside the company. None tests, from the outside, whether help can be reached at all, and none was designed with a non-human customer in mind.

What ACES measures

ACES is a score from 0 to 100, where higher means less effort. A score of 75 or more is *low effort*, 55 to 74 is *moderate* and below 55 is *high effort*. It has four parts:

- **Voice (30%).** From automated calls to the main published support number. A line starts at 50 if we find a route to a person and 10 if we don't. It earns up to 25 more for speed (all 25 if a person is under 30 seconds away, none at five minutes), up to 10 for a shallow menu (three fewer for each choice after the first), 5 if it understands spoken requests, and 20 if an AI voice agent can answer or finish the task itself.
- **Digital agent (30%).** The best assistant on the website: an AI agent that can hand off to a person scores 100, an AI agent without hand-off 70, live chat with a person 60, a chatbot with hand-off 50, an FAQ chatbot 25, none 0. If it only appears after signing in, it keeps three-quarters of its score.
- **Channel choice (20%).** Points for each way to get help, weighted toward two-way, real-time channels: a phone number earns 18, an AI agent with hand-off 16, live chat 14, email, text, an account portal or in-app help 8 each, and social channels 2 to 3. The total is capped at 100.
- **Access (20%).** Starts at 40. Plus 25 if the phone number is no more than one click from the home page, plus 25 if phone or chat is open around the clock, plus 10 if you can get help without an account (minus 20 if you can't), and minus 15 if there is no phone line at all.

When a company has no voice audit yet, the other three parts are reweighted. A channel counts only if we saw it; anything we couldn't verify counts as absent, so a score is a floor until it is verified.

The agentic test. Everything in ACES is observed by software, not by a person: an automated, signed-out browser on the website and automated calls on the phone. That is deliberate. The obstacles that stop our tools (a bot wall, a sign-in page, a press-1-only menu, a line that is closed) are the same ones that stop an AI agent acting for a customer. Not every obstacle is a mistake: bot protection guards against fraud and identity checks protect accounts. But each is effort that someone, human or agent, has to spend.

Preliminary findings

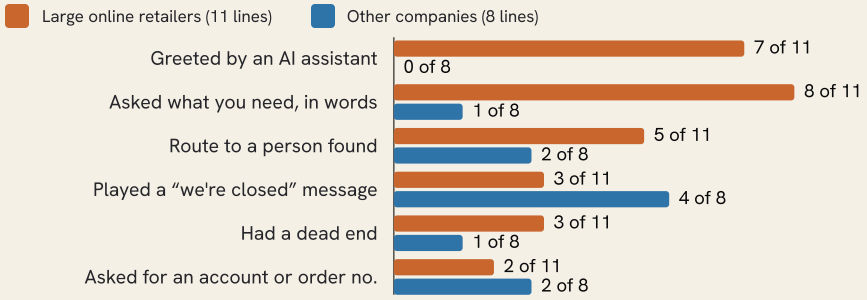
We compare two groups. The first is large online retailers: we reviewed 21 retailers' websites and have so far audited 11 retailers' support lines (Lowe's has a phone audit but no website review yet). The second is 8 other companies whose lines we have mapped: Ace Hardware, Saatva, Revlon, PetMeds, Hims & Hers, Lucid Motors, New York Life Insurance and Marchon Eyewear. In all, 163 automated calls between Nov 12 and Nov 28, 2026.

On the phone, retailers have gone conversational

Eight of 11 retailer lines invited the caller to say what they needed in their own words, and seven introduced themselves as an assistant. Apple's line said, "I'm an automated system that can handle complete sentences." Gap's said, "Hi, I'm Jean, your Gap AI assistant." Nordstrom's introduced "Nora, Nordstrom's AI assistant." Lowe's is the pattern we expect to spread: its care center was closed when we called, but its virtual store assistant still answered. Costco is the exception among retailers: a press-1 menu that we mapped to 21 menus and 65 options, with nine requests to key in an account or order number, though a person was only two choices away.

Among the other companies, six of 8 lines began with a press-1 menu, an account-number prompt or a closed recording, and only New York Life asked an open question. Saatva's line took 25 calls to map, revealing 24 menus and 45 options up to six levels deep, and we found no route to a person. Marchon Eyewear asked for an account number before anything else. Revlon's line, reached outside its hours, offered to take a message. Lucid Motors' line was closed apart from roadside assistance.

FIGURE 1 **The phone lines, retailers versus other companies**



Share of each group's audited lines. "Closed" messages depend on when we called; we report what we heard.

Reaching a person

We found a route to a person on 5 of 11 retailer lines and 2 of 8 others. On Apple's and Home Depot's lines, as on Ace Hardware's and New York Life's, a person was about 38 seconds and one choice away. Costco and Lowe's took about 56 seconds and two choices; Nike about 74 seconds and three. These are estimates of time spent listening to the menu before the transfer, not hold time, which our calls don't measure. Dead ends were concentrated: Ace Hardware's menu had 15 recordings that end the call, Nike's 12 (along with two identity checks) and Home Depot's three.

One result needs a plain caveat. On Gap, Best Buy, Target and Nordstrom, an AI assistant answered and our calls found no route to a person. Our mapping calls are built to walk menus; faced with an open-ended assistant, they said little and hung up within one or two calls. "No route found" there means we didn't find one, not that none exists. Under the version 1 rules those lines score 20 on voice, and we have not awarded the 20 points for an AI voice agent to any line, because we haven't yet tested whether one can finish a task. Both are the first things we are fixing.

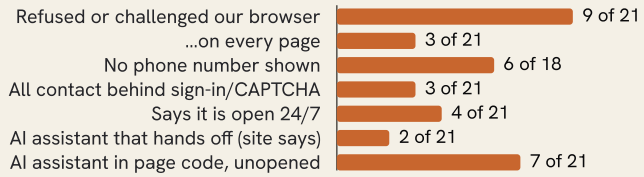
On the web, the front door is often shut to agents

Nine of the 21 retailer websites refused or challenged our automated browser on at least some pages, with an access-denied page, a rate limit, a CAPTCHA or a "Robot or human?" wall. Kohl's, Macy's and Wayfair blocked every page, so we could verify nothing about them. We count a block only when the retailer's own site refused us, not when our own network setup kept a script from loading.

Of the 18 sites we could view, six showed no phone number to a signed-out visitor: Amazon, eBay, Etsy, GameStop, Temu and Walmart. Etsy, GameStop and Temu publish no customer-service number at all; GameStop's "contact our support team" is an email link. Walmart's number appears in its page code but not on the page. Amazon sent every Contact Us option to sign-in, and eBay's contact page sat behind a CAPTCHA. Only four sites (Chewy, Home Depot, Nordstrom and Temu) say they are open around the clock.

AI assistants are arriving on retail websites faster than we could test them. Home Depot describes "Magic Apron, our AI-powered virtual assistant" with hand-off to a live associate, 24 hours a day, and Etsy's help center describes an AI support assistant that can transfer to a person, after sign-in. Page code at Best Buy, Chewy, eBay, Nordstrom, Target, Walgreens and Walmart points to AI assistants that we could not open in this round (Best Buy's is flagged for employees only; Chewy's is switched on in its app only). Nike runs a scripted menu bot that routes to its experts.

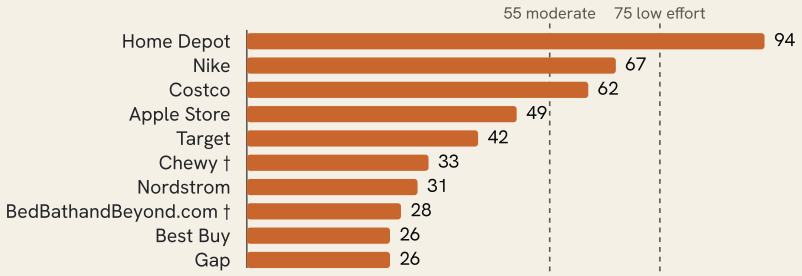
FIGURE 2 **What an automated visitor met on 21 retailer websites**



ACES for ten retailers

Ten retailers have both a phone audit and a website review, so we can score all four parts. Home Depot leads with 94: it is the only one that combines a short phone path, an AI assistant that hands off to a person, round-the-clock chat and nine ways to get help. Nike (67) and Costco (62) follow, both with live chat and a reachable person on the phone. Apple's line is one of the fastest we mapped, but its overall score falls to 49 because we found no chat entry point on its shopping pages. At the bottom, Gap and Best Buy score 26, low largely because neither their phone assistants nor their chat could be tested. Scores marked † (Chewy and BedBathandBeyond.com) rest on a review that could confirm fewer than four of 18 channels and are floors.

FIGURE 3 **ACES for ten large online retailers (0–100, higher is less effort)**



† Review confirmed fewer than four of 18 channels: the score is a floor.

The four parts behind each score in Figure 3.

Retailer	Voice	Digital agent	Channel choice	Access	ACES	Band
Home Depot	89	100	84	100	**94**	Low effort
Nike	80	60	51	75	**67**	Moderate effort
Costco	85	60	40	50	**62**	Moderate effort
Apple Store	89	0	34	75	**49**	High effort
Target	20	60	40	50	**42**	High effort
Chewy †	20	0	34	100	**33**	High effort
Nordstrom	20	0	26	100	**31**	High effort
BedBathandBeyond.com †	20	0	34	75	**28**	High effort
Best Buy	20	0	26	75	**26**	High effort
Gap	20	0	26	75	**26**	High effort

Comparing the two groups

On the voice part alone, the retailers average 50 and the other companies 35; the median of both groups is 20, because so many lines in each produced no route to a person. The retailers' advantage comes from more lines with a reachable person and fewer after-hours recordings (3 of 11, against 4 of 8). The other companies more often still depend on press-1 menus and business hours: precisely what an agent calling at night cannot get past. The groups are small, 11 and 8 lines, and were not chosen at random, so these are differences to test, not conclusions.

All 19 audited support lines. Voice is the voice part of ACES (0–100). ACES includes all four parts and is shown only for retailers whose website we also reviewed. "Menu to a person" is the estimated listening time and choices before the transfer; hold time is not included.

Company	Audited	Calls	How it answered	Menu to a person	Dead ends	ID checks	Voice	ACES
LARGE ONLINE RETAILERS								
Apple Store	Nov 28	5	AI assistant	~38 s, 1 choice	0	0	89	49
Home Depot	Nov 27	5	AI assistant	~38 s, 1 choice	3	0	89	94
Costco	Nov 27	25	Press-1 menu	~56 s, 2 choices	0	9	85	62
Lowe's	Nov 20	14	AI assistant, closed on a path	~56 s, 2 choices	0	0	85	–
Nike	Nov 28	18	Open question, closed on a path	~74 s, 3 choices	12	2	80	67
BedBathandBeyond.com	Nov 28	1	Closed recording	None found	1	0	20	28 †
Best Buy	Nov 28	1	AI assistant	None found	0	0	20	26
Chewy	Nov 27	1	Named greeter	None found	0	0	20	33 †
Gap	Nov 28	2	AI assistant	None found	0	0	20	26
Nordstrom	Nov 27	1	AI assistant	None found	0	0	20	31
Target	Nov 27	2	AI assistant	None found	0	0	20	42
OTHER COMPANIES WE'VE MAPPED								
Ace Hardware	Nov 20	24	Press-1 menu	~38 s, 1 choice	15	0	89	–
New York Life Insurance	Nov 16	25	Open question, closed on a path	~38 s, 1 choice	0	0	89	–
Hims & Hers	Nov 16	1	Named greeter	None found	0	0	20	–
Marchon Eyewear	Nov 12	1	Account number first	None found	0	1	20	–
Revlon	Nov 15	1	Closed recording	None found	0	0	20	–
Saatva	Nov 17	25	Press-1 menu	None found	0	0	15	–
PetMeds	Nov 12	10	Closed recording	None found	0	2	14	–
Lucid Motors	Nov 16	1	Closed recording	None found	0	0	11	–

How to read these results

A score is only as good as the evidence behind it. Before using any number in this report, check four things:

- **How much was verified.** A score built on a full phone audit and a website review that confirmed most channels is firm. A score marked † rests on a review that could confirm fewer than four of 18 channels; treat it as a floor that is likely to rise. A score without a phone audit is reweighted from three parts and is less certain still.
- **Whether a gap is real.** With one snapshot per company and rules that award points in steps (a channel is there or it isn't), a difference of a few points is within what a re-test could change. Treat gaps under about ten points as ties, and read the parts table rather than the total to see where a difference comes from.
- **Whether "not found" means "absent".** Our tools count only what they saw. A route to a person we didn't find, or a chat window that didn't load, scores as missing. That makes the scores conservative: they can understate a company, not overstate it.
- **What the group comparisons can carry.** The groups are small and hand-picked, so differences between retailers and other companies are hypotheses for the next round, not findings about industries.

The firmest results are the obstacles we observed directly, such as a bot wall, a hidden phone number or a menu full of dead ends: facts about a company on the day we looked. The rankings are a starting point, not a verdict.

Methodology and limits

- **Dates.** Phone audits finished between November 12 and 28, 2026. The website review was done on November 27, 2026. Websites and phone menus change often.
- **How calls were placed.** Each audit placed short automated calls to the company's published support number, chose menu options and spoke simple requests ("order status", "returns"), and hung up as soon as a transfer to a person began, so no one's time was used. Time to a person is estimated from the greeting and the listening time per menu level; hold time is not measured. Calls were placed at different times of day, several outside business hours.
- **Automatic classification.** Prompts are transcribed and classified by software and can be wrong. On PetMeds' line, for example, the recording after "press 2 to connect to an agent" sounds as if someone may have picked up, but it was not classified as reaching a person. We could not tell whether the conversational greeters at Chewy and Hims & Hers were people or voice agents.
- **What "unknown" means.** The website review used a signed-out automated browser and asked no questions of any assistant. Where a page or widget didn't load, the channel is "unknown", which the score treats as absent. Some failures came from our own network setup blocking third-party scripts; we don't count those against the retailer.
- **Not yet scored.** No line received AI voice agent points, because we didn't test whether an agent can finish a task. Retailers greeted by an AI assistant with no route found are likely understated.
- **Sample and selection.** Ten retailers have a full score, 11 retailer lines and 8 other lines have a phone audit, and the companies were chosen for relevance, not at random. Nothing here describes an industry as a whole.
- **One snapshot.** Each company was observed once, at one time of day, from the United States, in English, as a signed-out visitor. Results can differ by hour, by region, for signed-in customers and for customers with an order in progress.
- **What we did not measure.** Hold time, whether a person or an assistant actually solved a problem, the quality of answers, email and social response times, and the mobile app.
- **An automated caller.** Our calls used a synthetic voice and short scripted phrases, which some systems may treat differently from a person, and our mapper was built for touch-tone menus.
- **The weights are a judgment.** The four parts and their points reflect our view of what costs a customer, or their agent, effort. They have not yet been tested against customer outcomes such as satisfaction, repeat contact or churn; that validation is part of the next version.
- **Preliminary.** Phone audits are still running for the other 8 reviewed retailers that publish a phone number; Lowe's website review is still to come. The companies are not a random sample. We will publish updates and corrections as audits complete, and any company can ask us to re-run its audit.

What's next

This first edition has clear limits. It covers a small, hand-picked set of companies, observed once, by tools that can miss what a person would find, and it scores channels without yet testing whether they solve anything. Several scores are floors, and the weights are not yet validated against customer outcomes. We publish it now because the obstacles we saw are real.

The next version of ACES will hold real conversations: with AI voice agents, to see whether they resolve a task or hand off, and with website assistants, using the same standard questions for every company. We will finish the phone audits for the rest of the retailers, repeat the study regularly so changes show up, and extend it to more industries. We will keep publishing the method alongside the scores, so anyone can check the arithmetic.

> **Get your company's ACES audit.** We'll map your support line and website the way a customer's agent would, free: teleperson.com/ivr-audit

SOURCES

1. Matthew Dixon, Karen Freeman and Nicholas Toman, "Stop Trying to Delight Your Customers," Harvard Business Review, July–August 2010.
2. Frederick F. Reichheld, "The One Number You Need to Grow," Harvard Business Review, December 2003.
3. Gartner, press release, 9 July 2024: survey on customer preferences about AI in customer service.
4. Gartner, press release, 4 August 2026: survey on access to a human in generative-AI customer service.
5. Gartner, press release, 19 August 2024: survey on self-service resolution.
6. Gartner, "The Effortless Experience" research on customer service and loyalty.
7. Forrester, US Customer Experience Index, 2025.
8. CallRail, consumer survey, September 2025 (vendor survey).
9. Parloa, Consumer Patience Index 2026 (vendor survey).

Phone and website data: Teleperson AI audits and website review, November 2026. Audit reports for each company are in the public audit library at teleperson.com/ivr-audit.