Ask the Expert:
Dave Brown on Customer Service

A reader asks: We just completely changed the way we run our support department. We changed the phone menus and just about everything else about the way we do things. I guess you'd call it a reengineering. What metrics should we monitor in order to know whether the changes are working?

Dave's answer: To answer that correctly, I'd need to know what the problems were that you were trying to solve by reengineering. However, I'll take the approach of describing the performance metrics that should be monitored by most support centers. But, I need to state a few assumptions and put my recommendations in perspective. I'll assume that your support center has multiple queues. That is, customers listen to a phone menu and then, based on their menu selections, they are routed to the proper queue, which is staffed by a group or an individual agent. I'll also assume that if nobody is available to take the call, the customer will be put on hold and then will be answered by the next available agent (with the correct skill). I'm assuming that you would not have calls roll to voicemail or to an "operator-type" to take a message…they would hold and be answered in the order received. If that's not your environment, then some modification of these metrics will be required. More important, if that's not your environment, then you might want to rethink that reengineering…you may have more opportunity to improve.

My recommendations fall into two broad categories: those that should be monitored in real time and those that should be reviewed after the fact (his-torical reports). The purpose of the real-time performance measurement is to avoid problems or correct problems before they get out of hand. On the other hand, historical reports are used for more in-depth analysis, trend identification, and planning. Both are important to effective management of a support center.

First, I'll talk about the real-time performance metrics. The goal is to recognize when there is a problem, or a situation that is about to become a problem, so that you are able to react quickly. A support center that monitors performance in real time is able to make on-the-spot adjustments for things like random call spikes… adapting to the situation rather than explaining it the next day.

Let's define "real-time metrics." In most support centers, one of the primary objectives is to achieve a certain service level. A typical real-time metric would be current service level or current average speed of answer. Either of those two variations might be calculated based on so far today or for a rolling period, such as for the past hour. But that's not truly "real-time," is it? It's the recent past—I call it "fresh history." If it's history and you can't change it, then it's not real-time…at least not by my definition. Real-time metrics give you the current situation, allowing you to shape the future based on that knowledge.

The right way to monitor real-time performance is to observe the current state of calls holding and then relate that to the number of agents and their current state. That will allow you to predict what is about to happen, and it provides an opportunity to make adjustments, thereby influencing the historical reports.

The most important thing to monitor in real time is the status of the call queues. You must be able to view each queue individually, not just total calls holding, unless any call can go to any agent. For each queue, you need to see how many calls are holding and the longest hold time. Preferably, you will be able to customize the view and group queues by product family or site or other logical groupings, based on the way you manage your operation.

Next, we must be able to view all agents and their current state. This point requires clarification because there are two ways to approach it. Let's say your support center schedules people on the phones half the day and gives them the other half off to do their research, testing, and other follow-up activities. In Approach #1, we would only want to see the agents that are scheduled on the phones, not those that are off doing research. Using Approach #2, we would have everyone logged in (so that you know your total potential staff), but some of those people (those that are doing their research) would be shown in some other "state" so that you recognize that they are not on phone calls—and also not expected to be on phone calls. This, by the way, is a common approach in a call center environment. The first approach is more common in a technical support center environment.

Okay, back to the agents and their state. By state, I mean they are either active (talking to a customer), available (waiting for a call), or unavailable (not ready to take a customer call). The first two states should be self-explanatory. The unavailable state usually equates to wrap-up or other after-call activities.

Additionally, we need to know how long they've been in their current state, and we need to be able to see which skills they have or with which queue they are associated. This agent status information, combined with the "calls in queue" information, gives you a complete, real-time view.

Assuming you have that information available to you in real time, then you have the visibility necessary to effectively manage in real time. You should constantly (or regularly) monitor the relationship between the calls in the queue and the availability of agents. With experience, you'll be able to predict what is about to happen. For instance, if you have 10 agents logged in and talking to customers and there are five customers in the queue, and you know your average talk time is 15 minutes, then you might estimate that hold times will reach seven and a half minutes (queue/agents x average hold time = predicted speed of answer). Is that acceptable? If not, you can take action, such as finding another agent or two and getting them to log in. What if you saw that a couple of agents were in unavailable status and had been for quite a while, as calls start to back up in the queue? Maybe, you'd go find out what was keeping them from taking a call! The point is that the real-time data I've described provides you with the visibility necessary to foresee problems while they are still manageable and make appropriate decisions.

The other, more typical group of performance measurements is the historical reports. By definition, these are reports of what has occurred already. You can't change it, but there's a lot to learn from studying the data, and it is absolutely essential to good planning. There are five overlapping categories: workload, service level, service effectiveness, customer satisfaction, and agent productivity. I'll talk about three of them here.

In the workload category, you are basically looking at volume and average handle times, and you need to see daily, weekly, and monthly reports. But, you also need to be able to view these in hourly increments (or smaller). The small increments are essential for staff planning (all staffing formulas require hourly or smaller increments). Typically, you'd want to categorize the volume in the same manner that you sort the calls into queues or some other logical fashion that coincides with the way you view your agents' skills and/or schedule them. For each time period, track the total calls offered, actual calls answered, and calls abandoned.

Next, is the service level category. The traditional method is to measure "average speed of answer" (ASA) and "average time to abandon" (ATA). That's acceptable. However, the preferred method would be by time steps, such as the percentage of calls answered within less than one minute, one to two minutes, two to five minutes, and greater than five minutes. The same thing for abandoned calls. I've also found it very helpful to know the longest call held before answer and the longest call held before abandon.

It's important to meet initial service level targets by answering calls quickly. But service effectiveness or call closure rate is equally important. You should track the number and percentage of calls closed on first contact. Then, several additional thresholds should be monitored (based on your environment), such as closed same day, next day, within five business days, and greater than five business days. In some environments, it's also valuable to track the number of issues that were escalated to product development.

The final two categories are agent productivity and customer satisfaction. However, those are topics in themselves, so I won't address them here. Suffice it to say that both are important and should be measured and monitored thoroughly. ◈

Author's Note: The question asked was specifically about telephone support. In most environments, where e-support plays an ever-increasing role, you should create comparable performance measures for your e-mail and Web-based support inquiries.

Dave Brown is a management consultant, teacher, and writer. Dave teaches management training programs for Support Center University (www.SupportCenterU.com). He also consults with selected clients to establish world-class service operations and is considered an expert in the area of process improvement, staffing models, and change management. Reach Dave at 303-494-4932 or dave.brown@SupportCenterU.com.

Have a tough challenge? Submit your question to Dave at dave.brown@SupportCenterU.com. He will respond to all inquiries, and if your question is selected for publication, you'll receive a complimentary copy of his book, *Optimizing Support Center Staffing.*