The Best Kept Secrets of Workforce Optimization: Making the Most of Your Staffing Resources to Improve Service and the Bottom Line

Maggie Klenke and Penny Reynolds
The Call Center School

The term "workforce optimization" is getting a lot of attention in the press lately, given the impact of the slowed economy and the ripple effects of September 11. Everyone is struggling to do more with less – to make the most of the resources they have in place before making new investments in implementing technology or adding additional staff.

So where do you begin in optimizing your call center workforce? This article will outline some of the "best kept secrets" of making the most of your precious resources. It will help you focus your energies in the areas that have the most potential impact without risking long-term negative impacts. These strategies will help position your call center for both today's lean times and the better ones that will surely follow.

Secret #1: Cutting staff will not necessarily result in cost savings.

Since 65 to 70% of a total call center's operating costs are related to staffing, that is generally the first place we look to reduce costs. It is all too common to think of layoffs and reduction in staff as a way to respond to the call from senior management to tighten belts. But before you write up the pink slips, make sure you understand the implications of staff reductions.

Let's assume that you're a fairly small call center with fewer than 50 agent seats. (If you're a larger center, you can view these numbers as representative of a specialized agent group within the bigger call center structure.) Most days you're meeting your service goal of 70% in 30 seconds. The snapshot below indicates the staffing picture with varying numbers of staff during a half-hour in which you're getting 175 calls.

Number of Staff	Avg Delay (ASA)	Service Level (in 30 sec)	Staff Occupanc
30	298 sec	24%	97%
31	107 sec	46%	94%
32	54 sec	62%	91%
33	30 sec	74%	88%
34	18 sec	82%	86%
35	11 sec	88%	83%

As you can see, staffing with 33 "bodies in chairs" would enable you to meet service level fairly consistently. But the loss of one person would worsen service level from 74% to 62% (or average speed of answer from 30 seconds to 54 seconds). Eliminating another person would drop service level to 46%, and the average delay would double to 107 seconds! And reducing staffing levels by three would horribly deteriorate service level to only 24%, resulting in an average delay of 298 seconds! So those callers accustomed to waiting for only half a minute in queue would now be waiting nearly 5 minutes!

But some call centers are making the decision to let their staffing levels drop under the assumption that they "can't afford" the incremental staff. But what many don't realize is that some reductions in staff might be outweighed by the increased telephone costs associated with the longer delay times. In this example, with 33 staff in place the average delay is 30 seconds per call. Multiply that by 350 calls per hour and that's 10,500 seconds (or 175 minutes) of delay. If we apply a fully loaded telephone cost per minute to that usage of $.06 per minute, that's $10.50 for the queue time. If we try and staff with 30 staff, remember our average delay increases to 298 seconds of delay. Multiply that by 350 calls and that's 1738 minutes of delay, priced at $.06 for a total of $104.30 for the queue time that hour. In other words, by eliminating three staff to save money, we've just increased our telephone bill by $93.80 for that hour! And this doesn't even take into account the likelihood of a longer call given the poorer than expected service levels. Telephone charges would likely increase even further!

This situation is even more dangerous in a revenue-producing center. If the value of a contact is $50, and agent salaries are $20 per hour, it is easy to see that putting another agent on the phone will pay for itself even if the agent only answers one call per hour that would otherwise have abandoned from the queue. But even if the value of the call is only $5, there is clearly a trade-off in determining the staffing level that will produce the highest net bottom line. The return on appropriate staffing must be argued against budget constraints.

But if not a staff cut, then where should you look for savings opportunities in your center?

Secret #2: One of the biggest opportunities for savings comes from consolidation of workload.

One of the most important concepts to understand about call center staffing is the principle of economies of scale. Simply put, the bigger the agent group the better. As the size of the team increases to meet an increased workload, the utilization of every individual within the group improves. So, look for ways to combine smaller groups. Cross-train personnel and look for ways to use skill-based routing to tap into everyone in the site who can handle every call type. The benefits in reduced staffing can be tremendous as seen in the following example.

Number of Calls	Erlang Workload	Staff for 30 sec ASA	Staff/Work Ratio	Staff Occupancy
100	10	14	1.40	.71
500	50	56	1.12	.89
1000	100	106	1.06	.94
2000	200	209	1.04	.96

The call center agent group handling the 100 calls will require 14 staff, with each agent busy only 71% of the time. Five of these groups would require 70 agents. But if all of the work of the five groups could be consolidated and funneled into a single group, only 56 staff would be needed, with each agent busy 89% of the time.

This is a clear example of how altering the workload by cross-training and consolidating groups can result in a decreased need for staff without affecting service negatively. In fact, service may even improve, depending on the level of staff reductions you decide to make.

And the economies improve as the size of the agent group (and amount of workload) increases. But bigger is better only to a certain point….which leads us to our next secret.

Secret #3: Bigger groups are more efficient only to a certain point…then consolidation backfires.

Some call centers get carried away by the power of pooling. In our last example, to squeeze out another few headcount savings, you might assume that you'd want to take the two 106-agent groups and consolidate the workload into one group that requires only 209 staff rather than 212. Wrong. At this point, we've reached the point where efficiencies are too high. And while you might think there is no such thing as "too efficient", it is a fact that groups of this size have such high occupancies that they really can't maintain that efficiency level for very long without agents burning out or displaying some very undesirable behaviors in order to get a breather between calls.

Process Reengineering

Perhaps some of the biggest opportunities for savings lie in simply changing some of our processes and procedures, which brings us to our next few secrets.

Secret #4: Better calculations of workforce shrinkage can earn you some improvements.

Workforce shrinkage (the amount of time agents are paid to be there but are not available to take calls) includes some things that are not changeable such as vacations and coffee breaks. But there is often a fair amount of time lost in agents just not being in place and available when the schedule is expecting them to be.

And sometimes fixing the problem is as simple as a math error. For example, if you've determined that you need 150 "bodies in chairs" between 3-4pm, and you know your shrinkage for breaks, meetings, training, off-phone work, etc is 32%, then how many staff do you actually need to schedule so that 150 are actually in their seats ready to handle calls?
A common math mistake is to simply inflate 150 by 33% (150 x 1.33 = 199 staff). Wrong answer. Check it by decreasing 199 by 33% -- you only get 133 staff! The correct calculation is 150 divided by (1 minus shrinkage) or 67% for a correct answer of 224 staff.

Focusing on getting the maximum logged in time from each agent can be a significant opportunity since the lost time to unexplained unavailability ranges from 2-10% in many centers. In our example above, tightening up shrinkage just from 32% to 30% (only a 2% improvement) will reduce scheduled staff needs from 224 staff to 214 staff!

Secret #5: An expanded schedule mix can easily result in overall headcount savings.

Another place to look for efficiencies is in the variety of schedule options you use in matching your call center workforce to the workload. If most of your agents work full-time, 8-hour schedules, then schedule inflexibility is a problem. It's amazing how much more efficient your schedules will be simply by adding some part-time (3-, 4-, 5-, and 6-hour shifts) and expanding the definition of full-time to include more than just five 8-hour days.

To illustrate the savings achievable, one call center that recently implemented workforce management software expanded their schedule mix from only 4-hour part-time and 8-hour full-time schedules to include 5-hour, 6-hour and 10-hour shifts. By making these simple changes the overall FTE count went from 189 staff to 152 staff – a 20% reduction in staff requirements and actually a better, more consistent level of service!

Another call center for a major catalog company, shifted from traditional 30-minute start times to having staff begin their shifts every 15-minutes of the day. The rationale of this change was that by staggering start times, breaks and other off-phone activities could naturally be staggered throughout the day as well. And it worked! By staggering these start times, the call center went from 124 staff required to only 114 staff – an 8% savings by this simple adjustment.

Education and Tools

One of the distinctions between a "best of class" call center operation and an average one is the focus the organization has on empowering their agents and supervisors with the education and tools needed to be successful.

Secret #6: Equip agents with workforce management knowledge.

Part of the secret to reducing the shrinkage outlined earlier is to get the front-line staff to adhere more closely to their work schedules. And in many cases the strategy is simply informing them of the impact they have by not being in their seats on time. Most are unaware of the effect on service that a single person has, or why it's not ok to "just make up the time later in the day".

Take some of your own numbers and put together a simple 20-minute training session for your next team meeting on staffing to meet service goals, and how one person out of their seat will affect not just service, but how hard the rest of the group has to work to keep up. That "moment of aha" can result in immediate improvements in adherence, and in turn reduce shrinkage and headcount requirements.

Secret #7: Equip supervisors with workforce management tools.

Part of the dilemma in workforce management is to track schedule changes and exceptions in a timely fashion in order to manage service level on a real-time basis. The easier you can make it to request schedule changes and report exceptions, the better – for both front-line staff as well as supervisors. Tools like real-time adherence monitoring can substantially reduce the time supervisors must take to look at agent ACD status and then match up to schedule information to ensure everyone's doing what they're supposed to be doing. These tools result in better control of adherence with much less effort on the supervisors' part, freeing them up to do more important tasks like coaching for improved quality and performance.

Secret #8: Handling emails with the same response time as incoming voice calls may be reasonable and desirable.

It is clear that more customers are choosing to communicate with our companies through email. Some centers are already handling this additional workload and the rest will likely add it in the near future. Typical goals for responding to emails range from a few hours to a few days. With the automated response system acknowledging that the company has received the request and letting the customer know when to expect a reply, delays of hours or days seem acceptable. But is it the best choice?

What do you do if you send an email to a company and you don't get a response in the time you require? Do you send another email? Maybe you just call. But whichever you do, what happens to the original email? It just sits in the queue until an agent gets to it and they process it, even if the question was answered already (unless a very sophisticated contact management system informs the agent it has been handled). This creates incremental work for the center and the risk that the answers given the customer don't match.

The reason centers generally set longer response time goals for emails than phone calls is that they believe it will reduce costs. But the workload is the same whether the work is done now or later (assuming the customer does not email or call again while waiting). The only potential saving is in the difference between staffing for a sequential workload and a randomly arriving one. So assuming the 56 person workgroup in Secret #2's chart above, the incremental staff for calls is 6 people over what it would be if 50 people did 50 hours of email work from the backlog over the same period. But that assumes that the email team can be worked at 100% occupancy and that is not realistic. If you make a more reasonable assumption of 90% occupancy, you are back to needing the same staff for the emails as you do for the phone calls.
So why not answer emails as they arrive just like calls? It would amaze your customers and confound your competition!! Your costs would likely be the same or possibly less. Of course, the emails will come at all hours whether the phone lines are open or not. So if the center is not already a 24 hours shop, be aware of this potential increase in hours of operation.

Secret #9: Scheduling blocks of time for other types of work may be the best way to accomplish workload blending

Now that the center is handling both calls and emails, it is time to decide how best to distribute the work among the agent team. Let's assume your center has the same 56 staff handling telephone calls mentioned in the chart in Secret #2. This is the staffing needed to meet a 30 second ASA. Now the email load must be handled. The occupancy level of those

phone agents is 89%, leaving only 6.6 minutes per hour available for any other activity (or just taking a breath between calls). And these 6.6 minutes will not come all at once, but will be in several between-call units ranging from a few seconds to a minute or two.
You can set up a separate team for the email handling or you can blend the email work with the telephone calls, fitting in it between calls. (Of course, blending assumes that the agents are capable of doing the written work as well as the telephone work.) The question is whether blending on a contact-by-contact basis is the best choice, or if a schedule that includes dedicated blocks of time for each kind of contact would be more effective.

If concentration is required to complete an email response effectively, then blending the contacts is risky. The agent may be interrupted during email processing by a new incoming phone call, resulting in the need to rethink the email later extending AHT, and in potential errors. Some agents may be able to switch media quickly and effectively, while others need to focus on one thing at a time. So most centers that have chosen to handle multi-media in their centers have elected to schedule blocks of time that an agent works on each type of contact they can handle. The blocks can be an entire shift or week, but more commonly are for an hour or two within a shift. This allows the agents to keep their skills active and get some variety in the workday.

Secret #10: Blending workload is a much better opportunity for small call centers than large ones

Blending various types of workload holds a promise for efficiency that was first introduced by the automated dialer vendors touting the glories of blending inbound and outbound calls. Now we are getting suggestions that blending calls and other media should work well too, since many of our agents have idle time on their hands that needs to be filled. There are centers that have scheduling restrictions that make it necessary to have more people than needed in lulls just to cover peaks. These centers are the best candidates for blending. But with reasonable scheduling flexibility, the lulls are used as times for breaks & training, and part-time personnel give us the chance to reduce staff during low volume periods as well. So few centers have staff with enough idle time to make blending effective or even viable.
This is another situation where economies of scale have an effect. In a large center, the occupancy of staff is quite high as noted in Secret #2 above. But in a smaller center, or in a small, dedicated agent team in a big center, the occupancy levels can drop to allow room for some “filler” work. So while it seems like only a big center could afford the technologies needed to make blending workable, it is smaller situations where it can be the most useful.

Summary

The secrets we’ve outlined here are nothing new. They’re proven techniques and strategies that are working well in “best of class” call center operations around the globe. These are strategies that will not only serve to optimize your workforce today in these tough economic times, but will position your center for continued success in the good times that are just ahead.

About the Authors…

Penny Reynolds and Maggie Klenke are co-founders and Senior Partners with **The Call Center School**, a Nashville, Tennessee based consulting and education company. The company provides a wide range of educational offerings for call center professionals, including traditional classroom courses, web-based seminars, and self-paced e-learning programs. For more information, see www.thecallcenterschool.com or call 615-812-8400.

