Fulfillment Technology

Effective Queues

Their Creation And Use

Queues and work buffers are found commonly in both production and fulfillment o perations. A queue is a “temporary location” used to buffer work between processes. It is observed that there is a near universal notion among operations personnel that their current queues are not sufficiently large, and increasing those queue sizes wil l lead to greater capacity or productivity. It may be enlightening to share a “story” of a queue that we had many years ago.

We were called to visit an older production facility. The production facility was buried in or surrounded by the town in which it was located. The facility had hundreds of large pieces of production equipment that covered the floor. We were called because production requirements were growing rapidly and space for new equipment was gone. Expanding the facility was impossible due the unavailability of land. Moving to a new site was not even considered due to the cost. The company was looking to free more production floor space by more effectively using a significant space that was currently used for “work in process” (WIP). Their idea was to buy some (a lot of) ASRS equipment to hold WIP. Understanding this problem, as we were escorted through the facility, we would stop and talk to the workers at the production machines. Pointing to their input WIP queues we would ask, “How long will it take for you to finish the work in that pile?” The answer came back in the number of weeks. Likewise, pointing to the workers outbound WIP queue we would ask, “how long will it be until someone comes and picks up that completed work?” Once again the answer would once again come back in the number of weeks.

As you can imagine, we did not recommend that ASRS equipment be added. Rather we recommended that they more effectively use the queue space they already had thus recovering production space by reducing the amount of WIP.

Just how big should a queue be? This paper address that subject and you will be surprised to find some of the factors that are part of that determination!

First, let us address *do we need queues at all?* A queue provides a buffer to allow “unsynchronized” processes to be coupled together without mutual interference. Unsynchronized processes are those that start and end independently of each other. Synchronized processes are processes that are coupled in time. Good examples of synchronized processes would be those seen in an automobile production line where the line continuously moves through production zones (processes). Without queues, processes are required to be synchronized and the overall production rate is limited by (no faster than) the slowest process. Conversely, an example of unsynchronized processes is seen where there are pickers and packers working independently in a fulfillment facility. In this situation, workers start and complete work asynchronously. Unsynchronized processes coupled together without queues require one process to wait on the other reducing efficiency. **Conclusion #1:** Coupling unsynchronized processes is benefited by utilizing queues through elimination of wait times. In determining the required size of a queue between unsynchronized processes the sustained work rates of each of the two processes must be considered. If they are not balanced, the queue size must be infinite. Smaller queues will only temporarily help in coupling unsynchronize d processes with permanently unbalanced work rates. **Conclusion #2:** Queues between unsynchronized processes that are permanently imbalanced are only a temporary benefit and once filled or emptied they no longer improve efficiency by eliminating waiting.

Queues between processes that have temporary work imbalances need to be evaluated to determine their real effectiveness. Queuing temporary imbalances implies that processes have the capacity to “make up” lost time. This can mean one of two things, either you are not normally working at full capacity or you will work longer. Another related consideration is the concept we regularly encounter is a concept or a desire to "get ahead". From an overall operational perspective "getting ahead" is an illusion of productivity or capacity improvement. "Getting ahead" may have some merit in processes that are inherently unreliable and are likely to get behind later however the better solution is to improve the reliability. The entire operation has capacity and havin g one area "getting ahead" yields no improvement. Getting ahead is another way of saying that a permanent imbalance between processes exists. **Conclusion #3:** Determining or defining to what extent and how lost time is "made up" and the extent to which "g etting ahead" is tolerated are important considerations in determination of queue size.

The last and most important, as well as surprising, factor in determination of required queue size is the result or product of <u>the operating paradigm that management dictates</u> for the facility. To demonstrate this, consider the following. What would be the response if you as the operation manager were to ask a floor supervisor: "How would you feel a large (or larger) queue or buffer between X and Y would benefit you?" With rare exceptions the supervisor would state that such a buffer would be of benefit. Why would such a response be the normal situation? It is because the floor supervisors rightfully see their own individual areas of responsibility independent of the entire operation. Likewise, if you as the operation manager were to ask floor supervisors: "Who could use more labor?" You would rarely get a negative response. In common practice we do not ask that question to all floor supervisors, just those supervis ors that we see as the "bottleneck" of the operation. We "look" for bottlenecks, and normally our first inclination is to build a queue around the bottleneck. By identifying a "bottleneck" we need to realize that we are also identifying "overstaffed", "over queued" and "under utilized" areas. Normally the adding or increasing the queue to a bottleneck will not reduce the bottleneck. There is flatly a work imbalance. Balancing work eliminates the bottleneck.

So how is it that we claim that <u>the operating paradigm that you as a manger create</u> influences or even determines queue size? Having floor supervisors of the various operating areas in competition with one and another or having them want to insure that they are not the one that held responsible for a capacity or production shortfall, you are forcing them to not only hoard their own resources, but to campaign for queues. They will make certain that their area of responsibility is not seen as a problem. They will show you the "great queues" of work that they have for a downstream area of the lack of work in the upstream process queue. They are proud of their queues!!! Their queues are their protection! Their queues are a visible demonstration of their area's success. Bigger queues needed? You bet - they cannot be big enough! **Conclusion #4:** A queue can never be large enough for a floor supervisor whose success is measured by only his or her own area of operation.

So What About Queue Size?

If what you truly need is added storage space, add it. Do n ot call it a queue. The key to effectively using queues is work balancing. Create a "common measurement of success for the facility" and have all supervisors see the "big picture". Recognize that exceptions "are the rule" and imbalances will occur. Realize and acknowledge that a supervisor does not normally cause an exception. Then look to ways to make work imbalances as small and as short as possible. Focus supervisors on how to quickly identify an imbalance. Then develop a plan to quickly respond to the imbalance. A simple method for response to an imbalance may be implemented by creation of small groups of flexible workers that may be quickly deployed from the areas that are "building queues" to struggling areas. Quick response = small queues, slow response = big queues. <u>The true measurement of queue size is in minutes (or seconds) of work it will buffer</u>. Measurement of how may cartons, pallets, or items a queue contains is a meaningless fact. Balance work - don't build bigger queues. Don't create rewards for big queues or always empty queues. Recognize that overflowing or under-running an adequately sized queue is actually reducing productivity due to inefficient use of labor. Next time you walk the floor look at those queues. Queues that are normally full or empty are usually not a reflection improper queue sizing, rather a reflection of unbalanced operations. If you are at a point where it becomes nearly impossible to manually manage work balancing, we at VAS have automation methods that allow early detection and automated response to impending imbalances. Happy queuing!