Mini Paper

SMOOTH WORK FLOW LOWERS COMPLEXITY AND RAISES YOUR OUTPUT

By Tim Fuller and Annabeth Propst

Abstract
Is your operation plagued with high inventory, low productivity, late shipments, or missed project deadlines? Do you experi-
ence high variability in the length of time it takes to complete an order or finish a project? If so, the problem may be in the way
you organize your resources to do the work. Work flow balancing may lead to significant reductions in the problems listed
above. This paper presents some basic concepts for improving work flow and shows how two different organizations have
implemented them.

Variation is the Root of all Evil
Many organizations have ignored the concept of variation when designing their work flow. Their resource capacities are
designed assuming that everything will happen according to plan. That:

⇒ Machines will not break down,
⇒ Parts will not need to be scrapped or reworked (or the level of scrap or rework is predictable), or
⇒ Unexpected events will not occur.

In short, they act as if unexpected disruptions to the work flow do not happen. It is common to see organizations set
resource capacities at the average expected workload in each step of a multi-step process.

In the real world, significant disruptions occur all the time. Shipments from suppliers are late, machines break down, key
people become ill unexpectedly, and production levels vary constantly. When capacity is maintained for the average workload,
management has a strong need to ensure this capacity is utilized so as not to lose production. In order to protect themselves
from these unexpected variations, organizations add buffers (either time buffers or inventory buffers) to allow one operation to
continue working even though other operations may be disrupted. However, buffers add significantly to the complexity of the
operation, increase cycle time, and add costs. A careful analysis of a typical manufacturing or service organization usually
reveals that the buffers are often unnecessary, too large, too small, or in the wrong place. These buffers often overload the
system's resources.

How Overloading a System Leads to Excess Complexity and Cuts Throughput
In the early 1980s, Hewlett Packard's Computer Systems Division assembled printed circuit boards for HP3000 computers.
The various assembly departments -- automatic insertion, hand component assembly, wave soldering, hand soldering, touch-up,
test, and repair -- had the capacity to handle the average work load and could use overtime in order to expand capacity and
catch up after a disruption. When bottlenecks appeared and could not be handled with overtime, capacity was added to the
operation in which the bottleneck occurred, growing the department, piecemeal, over time. Each work area in the assembly
department held several days worth of work-in-process inventory (WIP) as a buffer to ensure production workers were never
idle.

Work orders for 20 to 200 circuit boards began, based on the start date
from the MRP system. Often however, industry shortages caused one or more
components to be missing when the kit of components was given to the
workers. Standard procedure was to complete as much of the work as
possible and then hold partially assembled circuit boards until the missing
parts arrived. Supervisors spent a significant amount of time expediting late
parts to keep the work flowing so due dates could be met. Cycle time for
an order of circuit boards was 20-25 days on average. Figure 1 gives a
representation of the situation.

After a 1981 visit by Dr. Deming, managers became convinced that long
manufacturing cycles were a significant source of variation in the process and
should be shortened. Management decided to study work order cycle time to
learn more about how the process was operating and the causes of the long
cycle time. They found that for most of the 25 days of the manufacturing
cycle, partially completed boards were waiting for backordered parts or busy
resources. The managers decided that if they waited to start production until all the parts were in hand, the cycle time could be
significantly reduced. They asked the workers who made up kits to hold the kits in front of the assembly process until all
component backorders were cleared. This meant that assembly would start much later than before. Production Control was
concerned that this would result in many late deliveries. Management pushed ahead, believing that a later start would actually
improve the situation.

The new plan was implemented and incomplete kits were no longer delivered to the assembly department. The immediate
effect was a large queue of incomplete kits waiting for backordered parts. After one week, most of the work-in-process
inventory in the department had been used up. Now, many production workers were idle as they waited for complete kits to
be delivered; some were loaned to work in other parts of the division as everyone waited nervously to see what happened.

Continued on page 6
After three weeks, the system had stabilized (approximately the same number of work orders were being completed as when being started) and the following effects were observed:

1. When a complete kit was received, people started work on it immediately and kept working on it until it was done. Cycle time was reduced to 5 days or less and the variation between actual and required completion date was greatly reduced. Even though orders were started two to three weeks later there were fewer orders behind schedule.
2. A large number of incomplete kits surrounded the assembly department.
3. Expediting inside the department had stopped.
4. Empty work-in-process shelves were being dismantled and removed. An effort to create a computerized WIP tracking system was now unnecessary and was abandoned because virtually all the work in the department was being worked on.
5. Although the volume of shipments remained unchanged and 15 of 60 production workers were being loaned out to other departments, there was still significant idle time in the department. The manager performed a quick study and found that it took only half as many people as before to do the work!

Later analysis showed that the old policy of starting work with complete kits added a tremendous amount of work to the department. Some activities that had been eliminated through improved work flow included checking kits for missing parts, determining partial assembly plans, putting unfinished work aside (this was a major chore because each circuit board had to be placed into a plastic bag to protect it), re-setting up a job, and prioritizing and scheduling the backlog. The cumulative effect of the department's original policy of overloading the system (in hopes of keeping everyone busy and getting work out on time) was to reduce the productivity by 50% (see figure 2). The department now needed only 30 production workers instead of 60.

In the short term, extra people were put to work on quality improvement projects or cross-trained to help in other parts of the division. Over time these extra people were absorbed into other jobs within the division.

Goldratt's Theory of Constraints Model

In the mid-1980s, The Goal and other works by Eli Goldratt helped people focus on constraint management as a way to maximize throughput and minimize costs and inventories given existing variation in the production system. Goldratt’s models are helpful in explaining how such dramatic results were achieved in the previous example with a minor change in operating policy.

According to Goldratt, the goal of any production operation should be to increase throughput (if you want to increase profits, you should focus on raising throughput rather than on reducing costs or cutting inventories. Goldratt’s system helps by placing buffers in the right places, making them the right size, and eliminating the complexity associated with overloading the system. His model also identifies the specific points in a process that should be improved first to provide the greatest impact on the system. Companies that follow Goldratt’s principles of constraint management report decreases in cycle time and more consistent delivery times. In The Goal Goldratt presents his five focusing steps to improve work flow.

1. **Identify** the system’s constraint(s).
2. Decide how to **exploit** the system’s constraint(s)
3. **Subordinate** everything else to the above decision.
4. **Elevate** the system’s constraint(s).
5. **Warning**!!! If in the previous steps a constraint has been broken, go back to step 1, but do not allow INERTIA to cause a system’s constraint.

Identification of the constraint requires finding the resource that is limiting the throughput of the operation. A key indicator is a constant backlog of ready-to-process work (recall from the previous example that under the old policy, much of the WIP was waiting for components and was not ready to process). To exploit the constraints, ensure that the constraint is always busy processing “real” work. Subordination of the other resources may require additional capacity in some departments or protective buffers. A meter or artificial constraint matching the capacity of the real constraint may be placed in the process to avoid overloading if the level of incoming work varies significantly. (Metering ramps on freeways are an example of this principle.)

Elevation of the constraint may be achieved by purchasing more of the resource, reducing non-productive time, or off-loading some tasks to non-constraint resources. In some cases it may be desirable to always operate with the same constraint. In these cases, non-constraint capacities should be adjusted prior to adjusting the constraint. In the printed circuit board example, the new policy elevated the capacity of all the resources so there was no constraint inside the department.

If you implement Goldratt’s model, you create a production system that balances the flow of the work rather than balancing the capacity of the various resources. The effects of designing your system in this way can be dramatic – production up, costs and assets down. You can easily construct simple simulations to demonstrate the effectiveness of Goldratt’s principles in increasing throughput and decreasing inventories and costs.

Continued on page 7
Theory of Constraints Applied at Malloy Lithographing

Malloy Lithographing, a book printer in Ann Arbor, Michigan, has adopted these principles to improve their service to customers. They have selected the expensive presses as a permanent, logical bottleneck and organized the other resources to supply a consistent flow of work to keep the presses busy. In the press room, the watchword is “keep the presses busy printing high quality pages.” In other departments, the goal is to keep the work moving along. A WIP backlog is maintained in front of the presses to ensure they are always busy. There is no focus on keeping non-bottleneck resources busy. If there is no work to process in these departments, employees switch to back-up projects or help out in other areas. In addition, Malloy has added a sophisticated metering system at the beginning of the operation to control the flow through the entire operation and avoid overloading the system during busy periods.

The change to this new system was no easy task and required cooperation from everyone in the plant. Staff members needed education about the purpose of the new system and the underlying principles. Project teams in each non-constraint department had to figure out how to process all the work they received within 24 hours to keep the work flowing to and from the constraint without large buffers. Press operators worked on projects to reduce set-up time, reduce scrap, and take other actions to keep press output as high as possible. Many discussions were held and experiments performed to develop a metering system that served customer needs and worked well internally.

Constraint management techniques have helped Malloy make a significant reduction in the time it takes to complete all the steps necessary to prepare, print, bind, and ship a book. Cycle time has decreased by one third; on-time delivery performance has nearly doubled. Dave Booth, Manager of Quality and Systems at Malloy, feels their new system is helping and makes these comments about their experience:

“In order for people in non-bottleneck areas to adjust in response to what's happening in the system, they need both good information and a good understanding of how the whole system works. Recognizing and providing this has been management's biggest challenge.”

“Our ability to understand what's going on in our system is better because work moves faster and feedback is more apparent. When we make a change, we see the results more clearly and quickly. This allows us to make better use of experimentation, which lets us improve more rapidly.”

“The impact of problems is greater now because there is less work in process to absorb the resulting variation. On the other hand, this fact serves to increase our commitment to find and eliminate the sources of problems rather than work around them.”

Final Thoughts

This article does not cover all of the possible techniques for work flow management and only briefly covers implementation of constraint management principles. Goldratt’s model helps to maximize throughput with minimum costs at a given level of process variation. Variation reduction in every intermediate process and process input quality will provide further gains. (This includes variation in product quality, process time, and production rates.) This will allow you to reduce the size of your buffers and, in certain situations, help you increase throughput.

Another technique, alluded to in the Malloy case study, is to design a logical constraint into your organization, rather than allowing it to happen. You then design your operation around the desired constraint and grow or shrink the organization in a coordinated way. In organizations where a particular resource is very expensive, it may be advantageous to consider this approach.

Although the two case studies discussed in this paper are about manufacturing, these techniques work just as well in non-manufacturing settings. The need for work flow management may, in fact, be even greater in the service industries because the flow of work is “invisible.” However, the lack of visibility makes the identification of constraints much more challenging.

About the Authors

Tim Fuller is a work flow consultant with Fuller Associates and Annabeth Propst is a business improvement consultant with Quality Transformation Services.