

CLP SUMMARY DOCUMENT

Periodically I summarize the contents of the Creating Level Pull (CLP) workbook I authored for the Lean Enterprise Institute (LEI) into a simpler text document for distribution. The following points and questions are adapted from my workbook “Creating Level Pull” published by LEI. This summary document is not intended to be a complete and exhaustive guide covering all aspects of the workbook or the associated LEI training workshop. Instead this document is intended as a high level summary emphasizing some of the shop floor basics surrounding scheduling execution that are often either broken or under-analyzed for improvement potential. The outline for this document will follow the same twelve basic questions in the CLP workbook.

Question 0 – What is the main problem?

I tend to get a lot of phone calls and inquiries for help in implementing kanban or establishing a mechanism for Heijunka for example. I do appreciate the inquiries. My question though is generally, “What is the main problem you are facing”. There is nothing wrong with kanban or Heijunka per se but keep in mind that these are *action items or countermeasures* to specific problems that Toyota faced in its production system development history. Too many times I run into situations where companies are merely keen to implement something since it is synonymous with lean. The action item winds up becoming the goal. When that happens the companies usually don’t get the improvement or results they had hoped for since proper cause and effect has not been established. Instead of getting to the root cause on the fishbone (Ishikawa cause and effect) diagram we wind up with something akin to a “wishbone” and it does not fix the real problem.

Assuming however that some problem exists in relation to level and pull production it is healthy to state it in a problem statement up front for clarity and communication. For example we are at 80% on-time delivery and the goal is to achieve 100%. Or we are at a 20 day lead-time and we need to be at a 14 day lead-time. Or for example we are at 2 months supply of inventory and desire to cut that by 80%, etc. Any of these would suffice for a basic problem statement and help teams get off to a good start on the topic of level and pull. When the answer is to “implement heijunka” before the problem has been established we sometime fall into the trap of what I call the push system of implementations instead of the pull type of implementation. Getting this distinction right up front is a critical first step in most every implementation situation I have observed.

Question 1 – What components to make to stock and what to make to order?

The history of pull production is basically a replenishment model with inventory being held in a market location and depletion of that inventory is used to signal replenishment. If this situation is true for your business or value stream then an important issue is to decide what items to hold in inventory and how much of each. In particular for finished good components in the value stream (and also at the raw material and component level as well) the number of items in inventory needs to be carefully analyzed and established.

Essentially you need to determine what items should be made to stock and replenished and what items should be made to order in the value stream. If the situation is fairly typical then about 20% of the part numbers in a value stream will probably account for 80% of the volume and the remaining 80% of the part numbers will account for 20% of the volume. *Note: If this 80/20 rule is not true then the real issue here may actually be failure to standardize designs in engineering which is manifesting itself as a scheduling problem in manufacturing.*

At a minimum as a starting point you should put the high runner items into a market in order to provide good levels of customer service and create better schedule stability inside the four walls of manufacturing. When correctly utilized, inventory can help provide quick and reliable delivery to the customer. An equally important but often overlooked point is that inventory can also help buffer the plant against large swings in customer demand. If you don't buffer against these swings for instance the variation will pass directly through the value stream and potentially create large amounts of disruption and inefficiency. Toyota and other lean companies strive to create a more level daily schedule precisely in order to avoid this type of daily disruption. An important end goal of lean is to reduce the timeline from the time an order is received until it is delivered through the elimination of waste. Of course this means improving flow and reducing inventory as much as possible. Instead of thinking of inventory as a pure form of waste I encourage you to remember that the actual quote from Mr. Ohno of Toyota was that "inventory beyond what you need to smoothly run the process is waste". Not having inventory in the right place at the right time is often a bigger waste than having the inventory itself.

Of course, in other cases however, production items are either made (engineered) to order, are too numerous, or are too expensive to hold in inventory. In these cases although everything may truly be unique at the final assembly stage often many things are more common at the sub-assembly, component or raw material stage. Look for where it is important to hold inventory in the value stream. Can you consolidate standard types at a mid point somewhere? Or can you at least hold the raw material and purchase parts you need in inventory in order to simplify production? If you can't do any of this you are probably looking at what is known as a sequential pull system that is unfortunately much harder to operate (more detail on this later).

Key Points: Look for the "80/20 rule" on finished goods, sub-assembly components, and raw materials in a value stream. If the rule applies then it is easier to set up simple replenishment pull systems. If not a more complex sequential pull system may be required and is much more difficult in many cases. Also it often helps to think in terms of lean "A,B,C logic" for components as well. A items are high runner items that should be made daily, B items weekly, and C items monthly for example (in some environments it might instead be weekly, monthly, quarterly). You might for example hold your A or B items in inventory and make the infrequent C items to order as needed.

Question 2 - How much of each item to hold in inventory?

Companies often fail to establish the right level of inventory to hold in market locations (this same logic basically applies to finished goods, component inventory and raw material). Even if this amount was once calculated properly it is often not revised over time and can lead to problems and shortages downstream. Careful attention at the part level is required to review current demand and any known forecast information to set inventory. (And yes unfortunately this is not an exact science since perfect information is not available and if you have a gazillion part numbers it is not possible to do this manually). For the items put in stock however some statistical method of estimating what is known as cycle stock, buffer stock, and safety stock needs to be properly calculated and put in place. Cycle stock is based upon actual historical demand, buffer stock is an estimate put in place to cover variability of the customer fluctuation, and safety stock is the amount required to make up for any process unreliability that might additionally be required (downtime, scrap, rework, etc.). If these effects are not properly taken into account then the internal logistical system often will not work regardless of the effort applied.

Key Points: Analyze a few part numbers in inventory (FG, component WIP, Raw Material) and determine if they are indeed held in the right quantities. See how often these items are reviewed and updated as well and if they contribute to delays in delivering parts to downstream customers. Make sure the lead-time assumptions for purchase parts are correct as well.

Question 3 – How is material stored and organized?

There is no magic dust or formula here. In general companies which both overproduce and allow random storage of inventory are prone to problems in merely locating the right part even if it exists in raw material, component WIP, or finished goods. The rules of 5S and visual control were developed in Toyota to highlight and identify problems such as these. Proper workplace organization, parts storage, and visual control are critical enablers for logistical efficiency. Otherwise material handlers and shop floor coordinators spend endless amounts of time looking for parts rather than making and delivering them to the downstream customer.

Key Points: Apply good shop floor 5S principles and visual control techniques to inventory storage, tool storage, schedule information, and anything else critical. If there is an element critical to performance then devise a method so that abnormal conditions become visual and noticeable by the person responsible for the area.

Question 4 – Where to schedule the value stream and what type of pull system?

The answer to this question actually depends upon the nature of your production environment and what type of pull system best applies to your situation. High volume discrete parts typically operate under a replenishment pull system where finished good inventory (or component level inventory) is held and replenished as the customer consumes this product. In this case the end of the line (some type of final assembly

operation normally) is scheduled and everyone upstream reacts to this processes instruction.

In low volume high mix shops or custom order environments however the opposite is often true. Instead the production instruction has to go upstream to an earlier point in the line and signal the need to make something. This part then flows downstream through succeeding processes and must arrive in assembly on time in order to aid the customer. Typically this latter type of pull system is called a “sequential pull” and is more difficult to operate. The reason for the difficult is because this type of system approximates a make to order signal and there is no extra inventory to buffer the system – it must be made right the first time and be delivered on time. Furthermore maintaining clear visual control over first in first out (FIFO) is extremely critical. The only buffer that exists in this type of pull system is the notion of a time buffer and releasing things slightly early (technically this is early production and a form of waste) in order to compensate for operational inefficiency, mistakes, or delays. Additionally, adjusting end customer delivery dates (rather than having a fixed delivery promise) is the only way to maintain a level load of work in the factory. Otherwise the factory production is roughly only as level as the incoming orders.

There are of course many shops that have to operate under both types of pull systems. In other words there is a combination of replenishment pull (make to stock) and sequential pull (make to order). This is called a mixed pull system and unfortunately it means that more than one signaling mechanism needs to exist in a production line. For example a replenishment trigger needs to exist for make to stock items and a special signal indicating make to order items needs to exist as well. Often these signals have to go to more than one location and unless they are thought through carefully then can “collide” at some mid point in the value stream and force the process to decide what to run on their own.

Key Points: At the value stream level the exact type of pull system and where the line is scheduled needs to be made crystal clear. For replenishment pull it is often the end of the line. For sequential pull it might be the first operation or somewhere in the middle. If a mixed system exists then how are the inevitable problems of mixed signals handled in the middle of the line? Create clear rules for which parts run first when there is a conflict.

Q5 – How level can you produce at the pacemaker?

Production leveling consists of two dimensions – quantity and mix. Normally emphasis is only put on the former element as companies attempt to smooth the workload in the factory based upon some notion of capacity or revenue. In order for production systems to work most efficiently at the value stream level however both the dimensions of quantity and mix need to be analyzed and considered carefully. For example companies often may think that parts are late from an upstream components shop due to scheduling delays when in reality the upstream process is overburdened by the actual demand quantity placed upon the process. (Remember – traditional MRP systems assume infinite capacity). This situation can also result in some companies as leveling is only thought of

in terms of dollar amounts or quantity of units to produce in final assembly shops. The reality however is that not all dollar amounts or work units are equal – especially in upstream operations. Some components have more work content associated with them than other. Five parts today might be the same as ten parts tomorrow if half the work content is involved the second day. Revenue can be equally misleading at the component level. For this reason leveling should consider the amount of work content involved and how the pacemaker process (and of course down stream) is affected.

Beyond the issue of leveling *quantity* companies also often fail to figure out the right *mix* to produce at a pacemaker process. In other words for the components made at the pacemaker process the right Every Part Every (EPE) Interval needs to be calculated. (In interested please refer to separate handout on Sikorsky Helicopter: Backbone of Lean in the Component Shops which is also available on the LEI website). In general the more frequent the EPE interval the more likely you are to make the right mix of components and deliver the right part on time downstream. In other words increasing batch sizes conversely reduces the probability of having the right part at the right time. Combining the right EPE interval with better understanding of capacity at supplying processes will improve delivery performance. Unfortunately traditional EOQ logic drives an EPE interval that is optimal for some false notion of “product cost” and ignores the need and inherent changing requests of customers downstream. Toyota Motor Corporation figured this problem out in the 1950’s and since that time has worked for several decades on shortening set up and changeover times as well and designing machines differently.

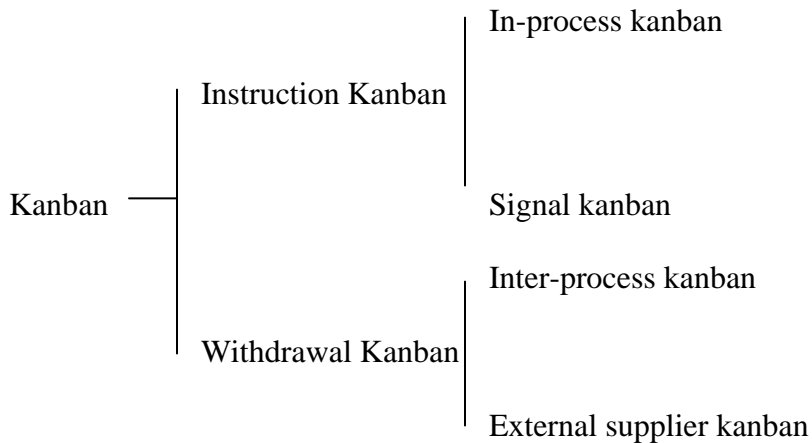
Key Points: Regardless of what type of operation or pull system you run the EPE interval for the pacemaker process needs to be calculated and determined. Some parts might be run on a daily basis, some on a weekly basis, others on a monthly basis and others as needed. However the more flexible the pacemaker processes (as well as batch machines in the value stream) the easier it is to deliver the right mix of parts on time. The price you have to pay to earn this benefit however is the challenge of quick changeover.

Question 6 – How to convey demand and what type of kanban to use?

It is important at all operations to be very clear about how demand information is conveyed and specifically what type of information signal is used. The word kanban is nothing more than the Japanese work for “sign board” or “signal”. Restaurants in Japan put out their “kanban” every day to indicate when they are open for business. Production processes often suffer from the problem of too many schedules floating around and thus no one puts much faith in one being accurate once the ink has dried. For reasons discussed earlier it helps at the value stream level if a single pacemaker exists for the area. The situation to strive to avoid is that of multiple dispatch lists from a central location that are issued and updated throughout the day. The unintended consequence of this system is often disharmony and less coordination than intended by well meaning parties.

Specifically pull systems address this problem by designating a pacemaker process for the value stream. All other processes subjugate themselves to this process. While local efficiency is minimized total system efficiency is optimized instead. In practice there are

different types of kanban that can be used for scheduling purposes. The major categories are as follows:



Instruction kanban are the category of kanban that tell a process “what to make”. There are in-process kanban that are geared towards making small lot or one piece flow situations and signal kanban which are for inherently large lot or batch processes.

Alternatively, there are also withdrawal kanban that instruct “what to take” or in other words what to convey between a process and a market or a market and a supplier for example.

A kanban signal can be as simple as an empty space on the floor, a piece of paper, or a more complex electronic signal – it is merely a method to make a clear unambiguous signal to make or bring something to the next process. Depending upon the circumstances the right type of kanban needs to be designed and used accordingly. Using the wrong type of kanban or implementing it in the wrong way will not deliver results.

Key Points: It is not the intent of this document to describe the types and uses of kanban in detail. The real issue is how are you scheduling the parts of your entire value stream and how is material handling signaled to move items in a timely manner? If this flow of information does not occur in a stable and consistent manner then part shortages and delays will inevitably occur. This aspect of scheduling requires significant time and expertise to design a system that will reliably work and be trusted by all parties involved.

Question 7 – How will you control production between processes or departments?

In addition to the correct use of kanban proper system design must also be placed upon the governing mechanisms between processes in order to avoid over production and better synchronize production. In general there are two ways; adhering to a market or adhering to some form of first in first out (FIFO).

In replenishment systems or at least where there are many fairly standard items produced a type of market mechanism can be used control flow between two points in a value

stream. In this type of system instructions to make and importantly instructions to stop making are governed by a supermarket of inventory that regulates flow. When some inventory is decreased and coupled with a kanban it acts as a signal to make some more of that item. Conversely, when the market area is full of a certain type it means to stop making that item. Although highly simplistic in concept a well designed market in conjunction with kanban is important to establish and can deliver significant results.

In other instances between processes a market mechanism and kanban may not be the right answer. After all markets hold inventory and when possible it is desirable to produce with less inventory instead of more. Where possible a simple system of FIFO lanes can often be established as a way of regulating production between two points. This ensures that product is processed in the order it is received and instructs the following process what to make in a clear unambiguous fashion. In job shop environments where no supermarkets are feasible FIFO lines are often the only way to clearly organize material flow between processes. In these environments FIFO lanes with designated spots of the floor at each process can be an effective technique for organizing what to make.

Key Points: FIFO lanes and market mechanisms often have to both be used in a value stream in order for it to flow as efficiently as possible. In order to make the lanes and markets work properly they have to be combined with the right amount of inventory (in the case of a market) and the right type of kanban in order to work correctly. 5S and visual control also play a role in making these simple items work since they do not work in isolation by themselves to govern production. Instead the FIFO lanes and markets are merely elements of a larger system that must be established (see next section on material handling as well).

Question 8 – How will you convey material between processes?

Specifically there are two simple ways to think about material handling. The normal case is a fixed time unfixed quantity route and an alternative method is the fixed quantity unfixed time route. The easiest way to consider the differences between the two is by an analogy. A fixed time unfixed quantity route is like a city bus driver on a standard route. The routes and pick up times for the route are known but exactly how many people will get on the bus is not known. Frequently this type of system works well in manufacturing for material handling especially in the case of purchased components being delivered throughout the shift to an assembly area. The material handler for a designated area can be assigned a specific route that forms a loop. The loop can be designed so that the material handler repeats it a fixed number of times per day (for example a one hour loop would usually entail 8 trips during the shift).

Importantly establishing the standard route with a parts withdrawal kanban and the regular delivery times leads to defining the amount of material required at the point of use location. Since any withdrawal kanban triggered on the last route should be delivered on the next route there is little reason to hold more inventory line side at the point of use than the hourly equivalent of the material handling cycle. If the material handling cycle is a fixed time route of one hour then one to two hours is all that needs to be held line side for safety. The rest should be stored in a market location and subjected to rules

regarding withdrawal and signals to either suppliers or internal processes as required. Frequently this type of fixed time unfixed quantity system is used with a dedicated material handling person utilizing tugger style vehicles or other mechanized devices.

However this application does not work perfectly in every case. Frequently parts are too large for example to be easily conveyed and lifter by a person. Some machines do not run every day due to low demand and are not staffed daily. Also regular and predictable frequency of need is not always well known in advance. In this instance a fixed quantity but unfixed time based system works better. Again an analogy is useful. In this case a city bus driver is not the right method for delivery. Instead a taxi driver “on call” is the right solution. The taxi driver waits until a specific call is made requiring his attention. In other words a standard looping route is not necessary. The passenger calls and orders the pick up and instructs the driver regarding time and where to go. Until that point in time the driver did not have any advance knowledge of the situation. In some cases in manufacturing this application works quite well. Rather than standardizing the route a defined number of material handlers (on fork lifts for example) are available “on call” for a signal to move material between locations instead.

Key points: Most companies pay too little attention to material handling when attempting pull or synchronized production. Material handling is like blood flowing through the human body paced by the heart. If the flow of blood in the body stops for any extended period of time the body will die. Manufacturing processes also starve in the same way for lack of material and poor pacing. For this reason material handling needs to be designated as either a fixed time unfixed quantity or fixed quantity unfixed time based system (each method has pro’s and con’s). However the key is to standardize the approach taken and ensure that a clear unambiguous signal to move the right part at the right time exists. Otherwise downstream processes will grind to a stop multiples times during the shift and destroy productivity with minor stoppages. For further detail on this topic you should refer to the LEI workbook by Rick Harris entitled “Making Materials Flow”.

Question 9 - How to schedule batch processes?

Inability to properly schedule and get material through a batch processes is also often a primary cause for part shortages. This is particularly true when there are certain constraint processes in a value stream that require production on some type of lot size for reasons of either quality or efficiency. The goal in these instances is to figure out the right lot size (ideally fairly small) to run, continually reduce the changeover time at the process and enact the right type of scheduling method with the machines.

In general there are three different ways to conceptually deal with a batch style production process. The machine may operate on a fixed sequence basis (but unfixed quantity of work), or it may operate on an unfixed sequence with an established lot size. With regard to the latter case there is the special case of signal or triangle kanban that can be established if applicable.

In order to ensure on-time delivery downstream from a batch process it is critical to understand the lead-time the process requires to deliver an item once it receives that signal. There are then specific logical assumptions that can be made to release production in time to arrive at the down stream process. Alternatively in replenishment environments specific re-order point and re-order quantities can be established as well. A specific example is provided in the CLP workbook.

Key Point: It is critical to first designate the type of scheduling method that will be used at a batch machine. Then accurate measurements and assumptions must be established for the lead-time to respond for each process in the value stream. Running batch machines on EOQ logic and not having balanced this with the right EPE interval for the value stream can lead to massive delays and forced schedule disruptions due to part shortages. It takes some expertise to schedule these machines and synchronize production accordingly – merely using rules of thumb will typically either result in delays or excess inventory.

Question 10 - How to implement and expand pull systems?

Typically the causes for problems in internal plant scheduling and logistics are the issues highlighted in the previous nine questions. Addressing them one by one can often lead to improved internal on-time delivery performance, reductions in inventory, reductions in lead-time, and improvements in both direct and indirect worker productivity. In most instances it is worth planning carefully how you will implement and expand level pull systems with some precise detail and planning. Specifically a detailed value stream map combined with further specific demand and inventory analysis needs to be executed. Each of the previous nine questions needs to be considered in light of the existing environment and a proper future state scheduling system should be designed in concept.

There is only so much “book knowledge” that one can accumulate however on this topic and at some point “learning by doing” is the only way to master the topic and obtain results. The best way to learn is by implementation in a controlled pilot area that is representative enough in scale to encompass all the elements discussed in the previous questions. Inventory levels need to be defined and located, the right type of pull system needs to be designed, the mix and quantity produced at the pacemaker needs to be established, the right type of kanban must be implemented, and the right flow methods and material handling need to be created.

Key Points: Once the above items are implemented on a value stream there is still the need however to roll out the system more broadly across the entire plant. Here again two different methods can be employed. One can either follow a value stream by value stream approach (customer specific approach) or roll things out department by department (process approach). There are valid cases for both arguments. The value stream approach works well when there are relatively few value streams in question and processes can be easily designated to a value stream (i.e. there are not many shared processes). Conversely when there are many value streams or many shared processes

exist it is often easier to fix problematic areas department by department. The key in both cases is to begin where the greatest need factually exists.

Question 11 – How to sustain level pull systems?

Unfortunately level pull systems are fragile devices – by design they function to surface problems for organizations to observe and repair. If the organization does not react to the problems being surfaced then the system will not function in an optimal manner. There are several stereotypical reasons why level pull systems tend to bread down.

First, there is often insufficient process stability in place in order to support the system. If there is massive downtime or quality problems in an operation a pull system will not work well. It is worth time stabilizing these problems in order to achieve more consistent and predictable levels before attempting pull production. In other words stability is required before agility.

Secondly, there is a failure to establish and monitor critical process metrics after the system is established. If the system was designed with certain lead-time, quality, or downtime assumptions then these items need to be monitored and tracked over time. If process stability deteriorates then the logistical system built around it will break down as well. Metrics for quality, downtime, changeover time, delivery, productivity, and other critical parameters need to be created and tracked over time for improvement.

Third, customer demand changes! Once a scheduling system based upon level pull production is set up it needs to be monitored and reviewed over time. Customer demand quantity and mix changes over time frames associated with the characteristics of the business environment. When a change in demand occurs a corresponding change needs to take place with regards to staffing levels, equipment capacity needs, inventory levels, and delivery from suppliers. A proper review interval and the key items to check needs to be established and ownership clarified. Otherwise the system will stop working and someone will inevitably say “Well we tried pull production here once and it stopped working after a while”.

Fourth, daily supervision is required to make the system run on a consistent basis. The internal workings of a level pull system are only as strong as the daily supervision governing the environment. On a daily basis problems will be surfaced that will require action. The ability in the supervisor ranks to spot problems and take action is critical to ensure success.

Question 12 – How to improve the level pull system?

The question of how to improve the system that has been discussed is difficult to address without specific data and situational information. The actual levers to improve differ case by case. In general however the following check points can always be evaluated as starting points for improvement opportunity.

- What is on time delivery to the customer (internal and external)?
- What stops us from being at a higher on-time level than today? (facts not opinions)
- What is our inventory level? (Raw, WIP, FG's)
- How can we reduce the inventory level and maintain on-time delivery?
- What is the lead-time through from order to delivery?
- Where is the majority of the product lead-time spent and why? (facts not opinions)
- How much time is lost looking for parts and why?
- How much downtime in final assembly is due to upstream part shortages?
- What is supplier on time delivery performance?
- What aspects of process stability (scrap, rework, downtime, etc.) can be improved?
- What is direct and indirect labor productivity? How can better delivery of materials impact this metric?

SUMMARY

The points contained in this document are intended as an aid for individuals attempting to improve internal aspects of manufacturing logistics specifically related to shop floor scheduling, basic inventory control, and material handling in a value stream as starting points. This is not an exhaustive guide for all aspects of scheduling or logistics by any means. It is intended as a practical starting point for questioning some basic shop floor practices and for identifying areas that typically cause problems in both low volume high mix and high volume low mix production settings. The material summarized here is a concise summary of the elements described in greater detail with a practical example in the author's workbook "Creating Level Pull" published by the Lean Enterprise Institute in Brookline, Massachusetts.

APPENDIX

Further Points for Low Volume High Mix

Point #1: Assembly schedule stability is critical

The nature of most low volume high mix operations is a final assembly department supplied by many component shops and purchased parts. The stereotypical problem is the inability to consistently get the right parts to show up in the right place at the right time. This symptom then manifests itself in poor on-time delivery performance to the end customer, plant inefficiency, and substantial build up of work in process inventory internally. The majority of the time the problem typically resides in the upstream component shops or suppliers and their inability to sequence product on time to final assembly. However, often final assembly is a contributing factor to the problem by changing the final assembly schedule on very short notice. For any scheduling, inventory control, and material handling system to work there needs to be some basic level of assembly schedule stability in place. Even in Toyota Motor Corporation making vehicles the final assembly schedule is fixed over a 2-3 day period – otherwise it sets off too many ripple effects upstream that disrupts manufacturing and worsens productivity and on-time delivery performance. This destructive phenomenon is known as the “bull whip effect” as demand amplification transmission occurs upstream. For an internal logistical system to work best this effect needs to be analyzed and minimized.

Point #2: Traditional scheduling practices often only makes things worse

Traditional scheduling practices often make things worse due to inherent problems in MRP systems. MRP programs often assume infinite capacity, and only function as well as the data and assumptions provided as inputs to the program. Systems are not necessarily updated real time throughout the day with accurate information regarding progress and changes in demand. Compounding this problem is that the internal logic of the system for inventory and lot sizing is based upon a modified form of economic order quantity (EOQ). Additionally some type of schedule optimization algorithm often exists that seeks to maximize work center productivity often at the expense of on-time delivery. Factor in traditional operational metrics that reward efficiency (actual hours compared to standard hours) and productivity (efficiency times utilization) and you have a system design to over produce and contribute to making the wrong parts at the wrong time. Inevitably “hot” lists, “red hot” lists, and the eventual “screaming pull system – make me this now or you are out of a job!” results. Supervisors and shop floor material coordinators react by circumventing the published schedule and working by “rules of thumb” in order to survive. Once the different component shops supplying final assembly start behaving in this manner it is virtually impossible to get the right parts to assembly at the right time except by sheer management force, excessive communication, and internal expediting.