Table of Contents

"This is Lean: Resolving the Efficiency Paradox" - Niklas Modig

Reference

[This is Lean: Resolving the Efficiency Paradox](#)

Review and Notes

I thought this was a great book, a fabulous introduction to help you understand what lean is all about and, just as importantly, what it is not. It helps in understanding of Little's Law, Theory of Constraints, and why Flow Efficiency (and therefore "lean") is changing how we think about things.

The base premise is that lean is about "flow efficiency" as opposed to resources efficiency in how work gets done. Flow efficiency focuses on the amount of time it takes from identifying a need to satisfying that need, whereas resource efficiency focuses on efficiently using the resources that add value within an organization. What is interesting is that when you focus on flow efficiency you also improve resource efficiency in many cases because work which results from being resource efficient (and not delivering on customer needs - extra work that is created) means we can free up some people to do work that is more aimed at what the customer needs.

Why is this important? For more than two hundred years, industrial development has been built around increasing the utilization of resources. Efficient use of resources has long been the most common way of looking at efficiency. It continues to dominate the way in which organizations in different industries and sectors are organized, controlled, and managed. From an economic perspective, it makes sense to strive for the most efficient possible use of resources. The reason for this is the opportunity cost.

Flow efficiency is defined as a new form of efficiency. Flow efficiency is new in that it breaks with the historical and natural focus on the efficient utilization of resources. Flow efficiency is important because the focus is on the unit that "flows" through the organization: the flow unit. Flow efficiency is a measurement of how much a flow unit is processed during a specific time period. The time period is defined from the time a need is identified to the time it is satisfied.

In a process, something is moved forward; we call these flow units. A flow unit can be material, information, or people. The difference in the dependence is the key factor that differentiates the two forms of efficiency. In resource efficiency, it is more important to "attach work to people" to ensure that each resource always has a flow unit to process. In flow efficiency, however, it is more important to "attach people to work," that is, to ensure that each flow unit is always being processed by a resource.

The discussion then moves to a a look at Little's Law, which states that:

```
Throughput time = flow units in process × cycle time
```

A couple of definitions:

- "Flow units in process" are all the flow units within the chosen system boundaries: all the flow units that have begun the process but have not yet exited it
- Cycle time is the average time between two flow units' completing the process and refers to the pace at which flow units move through the process.

The book then looks at the theory of constraints - the law of bottlenecks states that throughput time in a process is primarily affected by the stage of the process that has the longest cycle time

Processes with bottlenecks have two key characteristics: Immediately prior to a bottleneck, there is always a line, regardless of whether it is material, information, or people flowing through the process. It is often clear which stage in the process is the bottleneck, particularly when the flow units are material or people. It can be more difficult to see the line at the bottleneck when the flow unit is information, but the line is there. The stages of activity after the bottleneck must wait to be activated, which means they will not be fully utilized. Because the bottleneck is the stage of activity that has the slowest throughput, the stages after the bottleneck will work at a slower pace than they could have.

Bottlenecks lengthen the throughput time as a line of flow units forms and waits to be processed. This can be understood using Little's Law. As there is a line, there are flow units in process. Given that we do not change the cycle time (by adding extra resources or working faster), adding flow units in process will increase throughput time.

There are two reasons bottlenecks appear in processes. The first condition for bottlenecks is fulfilled if the stages in the process must be performed in a certain order. The second reason bottlenecks exist is variation. There needs to be variation in the process. In principle, it is impossible to eliminate variation, and it has been shown to have a very negative effect on processes and flow efficiency.

This relationship was formalized in the 1960s by Sir John Kingman in his famous Kingman's Formula (see What Is Wrong With 100% Utilization Thinking? for more information). The figure shows how throughput time (on the vertical axis) is dependent on utilization (on the horizontal axis): Throughput time increases the higher we move up the vertical axis. Utilization (which we call resource efficiency) on the horizontal axis is a measure of how efficiently the resources are utilized. The closer to one hundred percent, the higher the resource efficiency.

Of course this is a particular problem for software systems - there is a lot of variation.

In general if we want to improve flow efficiency, the law gives us some ways to do this:

- Reduce the total number of flow units in process by eliminating the causes for the lines (of material, information, and people). (reduce WIP)
- Work faster, which reduces cycle time.
- Add more resources, which increases capacity and reduces cycle time.
- Eliminate, reduce, and manage the different forms of variation in the process.

The book then looks at the problem with focussing on resource efficiency at the the expense of flow

efficiency. While resource efficency may be beneficial from the organization's point of view, it can represent a problem from a customer perspective - long throughput times.

These negative effects result in three "sources of inefficiency." The first of these sources is related to people's ability to deal with long waiting times. Secondly the non-fulfillment of a need can create new types of needs, which in turn create new needs. In other words, there is a chain reaction. Thirdly waiting time closes important windows of opportunity. Fourthly you create inventory and inventory requires additional resources.

To understand the first and second order effects, take your email example. Too many emails trigger stress. Email is [art of what we do, but an inbox with 200 unsorted emails can be somewhat overwhelming. Where do you start? The primary need is to answer important emails. However, the large number of emails creates a secondary need for a strategy to sort emails. One strategy could be to address the emails in date order, while another could be to start with the most important people first. Or perhaps you could look for "flagged" messages or discard those emails on which you have only been copied. Whatever method you use, sorting, structuring, and searching are activities that fulfill the secondary need, which is to handle the large amount of email. The primary need is to read, answer, and store email, but the large number of messages necessitates certain activities in order to gain some kind of overview. Apart from creating unnecessary work, handling a large number of emails at once can also be stressful.

But there is more to it than that - another source of inefficiency: many restarts per flow unit. Starting over on the same task generates mental set-up time. Restarts are created when you have to start over on the same task. Continuing with the management of a large email inbox as the example. There is a high risk that you will need to read messages more than once. Some messages are simply too complex to deal with at that time, so you read them, file them, and return to them later. Sometimes you have to come back to them more than once, perhaps because you need further information. When work to be done just sits in a pile waiting, it is easy to lose sight of the big picture. The time and energy we spend on categorizing and structuring the work creates delays. The delays and the different type of activities (such as searching, identifying, categorizing, and structuring) mean that we have to re-familiarize ourselves repeatedly with the same information. The challenges of starting over on the same task are also driven by mental set-up times. A person needs time to focus his or her mind on a task, and it is mentally challenging to deal with several tasks at the same time. It is particularly challenging when we must repeatedly shift our focus from one task to another. The fewer tasks we have to deal with at the same time, the easier it is to focus. The more frequently we have to switch between tasks, the longer the mental set-up time becomes in relation to the total time. Therefore, the limitations of the human mind mean that a high number of restarts will generate new secondary needs that would not have been created if the work had been finalized the first time.

This helps us understand why we need to focus on flow efficiency and, when we do, we often also improve resource efficiency. Resource efficiency improves because we don't work on things that are caused by long processing teams - these secondary effects.

Simply ask yourself - "what is the root cause of secondary needs? Essentially, an over-focus on resource efficiency creates low flow efficiency. This creates "efficient islands" in which fulfillment of customer needs is split up into several, smaller steps that are performed by various individuals or parts of an organization. No one island has a full overview of the entire process; each island sees only its own part. How much of the time that you spend at work is spent on fulfilling secondary needs? In other words, how

much of your total working time is dedicated to superfluous work? The efficiency paradox is explained by superfluous work. By over-focusing on resource efficiency, process laws guarantee that flow efficiency will suffer. If flow efficiency suffers, then several secondary needs will be generated. Activities to meet these secondary needs may seem like value-adding activities, but they would not be necessary if the primary need were already fulfilled. The paradox is that we believe we are utilizing our resources efficiently, but we are actually being inefficient, since much of that utilization comes from superfluous work and non-value-adding activities."

The book then goes into a discussion about the origins of lean from the Toyota Production System and differences between this and what we now call lean. In particular it looks at the world of "lean" ideas that exists out there and tries to make sense of what lean is, and what it isn't.

There are number of problems with current definitions of lean. Firstly, the definitions are made at different levels of abstraction. Secondly, lean has become a means instead of an end. Thirdly, lean seems to have become all that is good, and all that is good is lean (and everything has become lean).

The authors then build a model to explain what it means to be lean - called the "Efficiency Matrix". The efficiency matrix builds on the two forms of efficiency that were presented in the first part of this book and illustrates how an organization can be classified based on (a) low respective high resource efficiency and (b) low respective high flow efficiency.

Efficient islands In the top left-hand corner of the matrix is a state we call efficient islands. In this state, resource efficiency is high and flow efficiency is low. The organization consists of sub-optimized parts that operate in isolation, where each part works towards maximizing its resource utilization. Through the efficient use of its own resources, each part contributes by lowering the costs for the goods or services being produced. However, efficient utilization of resources comes at the expense of efficient flow. Flow efficiency for every individual flow unit is low. In manufacturing, this is represented by each component/product's spending most of its time as inventory. In services, this is often represented in the form of unwanted waiting time during which the customer does not receive any value.

The efficient ocean In the lower right-hand corner of the matrix is a state we call the efficient ocean, where flow efficiency is high but resource efficiency is low. The focus is on the customer and meeting their needs as efficiently as possible. In order to maximize flow efficiency, there needs to be free capacity in the organization's resources. Flow is efficient at the expense of an efficient use of resources. Resources are only used when there is an actual need to satisfy. Creating an efficient ocean and creating flow require a good understanding of the big picture, not just independent and efficient islands.

Wasteland In the lower left-hand corner of the matrix, the organization is unable to use its resources efficiently or create an efficient flow. Obviously, this is not a desirable state to be in because it wastes resources and creates less value for the customer. In this state, there are neither efficient islands nor an efficient ocean. It is wasteland, that is, poor utilization of resources and poor flow.

The perfect state In the top right-hand corner is the perfect state. Organizations that achieve this state have both high resource efficiency and high flow efficiency. It should be clear by now that it is difficult to reach the perfect state. But this state is impossible as you need two prerequisites:

- The first prerequisite for being able to reach the star is perfect predictability of demand. The organization must be able to predict perfectly: What is demanded - When it is demanded - Which amount is demanded.
- Secondly means perfect predictability of supply with similar problems to the above.

Level of variation establishes the efficiency frontier Therefore, it is the level of variation in demand and supply that determines which operational states an organization can achieve. Variation limits the possibilities of reaching for the star. Variation creates "an efficiency frontier." You can only get so close because of variation.

Examples of organizations facing high variation are those in which the main flow unit is people. Many service organizations will fall into this category. People introduce an element of variation that is very hard, if not impossible, to avoid. We cannot standardize or control people in the same way we can material or information. However, regardless of the type of organization, it is often possible to become better at eliminating, reducing, and managing variation.

Business need to make choices. It might make sense to focus on resource efficency if that is what your business model supports. So the business needs to determine a business strategy and develop an operations strategy to help decide what to do:

- A business strategy (what) defines what The business strategy defines the value the company will offer the customer; that is, the value the customer will experience when a good or service is consumed. At the highest level of abstraction (the "fruit" level), an organization can focus on either differentiation or cost. In this context, differentiation includes a range of things, such as delivering a better experience, better food, faster service, or a wide range of products from which to choose. In other words, differentiation is anything that a customer considers valuable. The cost is the sacrifice, in terms of money, time, or energy, that the customer must make in order to satisfy his or her need.
- An operations strategy (how) defines how An operations strategy helps realize a business strategy and defines how value is to be produced. All organizations have an operations strategy, whether it is explicit or not. The operations strategy answers the question "How shall we produce value?" We are assuming here that we have already defined the type of need the organization is trying to satisfy and the target market. There should be a clear link between the business strategy and the operations strategy. Given that we have defined the organization's business strategy, we can now develop an operations strategy.

Once we have our strategies we can now go back to our definition of lean. Movement in the matrix can occur in two dimensions: Resource efficiency can be increased or decreased. Flow efficiency can be increased or decreased. In brief, lean is an operations strategy that prioritizes flow efficiency over resource efficiency. In other words, lean is a strategy for moving "to the right and up" in the efficiency matrix. This is what we mean by "continuous improvement" - it is crucial to understand that the realization of a lean operations strategy is a journey that never ends.

One side effect of this thinking is that, based on an observation of a single day, you cannot determine whether an organization is truly lean. You can only determine if you are lean by comparing yourself to what you were doing yesterday - "Is there any little adjustment that can make us a little bit more beautiful than we were yesterday?"

Other Ideas

What does "focus on the baton not the runner" mean?

The flow-efficient organization is analogous to a relay race. In a good 4 x 100 meters relay team, the handovers are smooth and all four runners can see what is happening all the time. By the time the first runner has nearly completed the first hundred meters, the second runner has already started running in order to simplify the handover and speed up the race. When the baton is handed over, no time is lost as both runners are at top speed. A case in point was Yohan Blake handing over the baton to Usain Bolt in the final of the 4 x 100 meters at the 2012 London Olympics. The Jamaican team ran the 400 meters in 36.84 seconds, a world record for baton flow efficiency! In a resource-efficient organization, however, the first "runner" is carrying many batons at once. In fact, the more batons the better. But after he has run the first hundred meters, there is no one to meet him. A phone call reveals that the second runner is in Thailand for a meeting. It takes several more calls to find someone who is free to run the second leg. By the time the batons are handed over, nine days later, two have been lost and one has been forgotten. This is not a gold medal-winning formula, but it is, unfortunately, how many organizations behave."

How Does Toyota's Pull System Work?

The key to the pull system was that Toyota saw the whole production process as one flow made up of different production steps. Every step had two roles: internal supplier and internal customer. The figure shows a simplified version of the production process comprising four steps, where the fourth step is closest to the customer. In step four, the customer order is taken, and the need is identified: what, when, and how many. The customer's need is then broken down by asking the following questions: What (components/materials) do I (step four) need to satisfy the needs of the external customer? When do I (step four) need these (components/materials) to be able to produce and deliver the finished product to the external customer at the promised time? How many (components/materials) do I (step four) need to be able to produce the product?

In order to achieve the pull system, Toyota mapped out the entire production process. The needs of external customers were the trigger in a long chain of value-adding activities. With this customer-orientated view, Toyota's goal was to maximize flow through the process: a fast information flow in one direction and a fast product flow in the other. Toyota wanted to avoid having work-in-progress between the steps in the production process and strived to eliminate anything that could inhibit the flow through the process. All forms of inefficiency or waste that did not add value to the product were eliminated to improve flow.

Book, Learning, Improvement, System, Transformation, Lean, TPS, Toyota, Review

~~LINKBACK~~ ~~DISCUSSION~~

From:
https://www.hanssamios.com/dokuwiki/ - **Hans Samios' Personal Lean-Agile Knowledge Base**

Permanent link:
https://www.hanssamios.com/dokuwiki/this_is_lean_-_resolving_the_efficiency_paradox_-_niklas_modig?rev=1478106989

Last update: **2020/06/02 14:27**