Ship Building: A workshop for the WIP curious

Length	60 - 90 min
Topics	Systems Thinking, Little's law, Different policies (Push vs Pull), Flow of work
Number of attendees	6 - 20
Description	In the workshop, the audience will actively participate in a flow system running under two sets of processing policies, and draw conclusions from what they experience. The idea is that lead time and WIP can vary under the different policies, even though throughput remains the same.

Introduction

This is a workshop that was originally run by Dr Klaus Leopold - we have taken the format, tweaked and extended it. Please see the links at the bottom for the video of the workshop that inspired us.

The workshop works best with a minimum of 6 attendees with 2 facilitators - during the experiments this translates as 6 worker stations, with 1 facilitator recording timings, and the other ... well, facilitating. It is possible to run this with a single facilitator, but we have found in practice that we benefited from having 2 people running things, especially when you get larger numbers of attendees.

Due to the hands-on nature of the workshop, it is unsuitable for remote attendees.

The ideal audience for the workshop are people who are new to the ideas of flow efficiency, skeptical of the use of WIP limits, or interested in studying the systems in which we work.

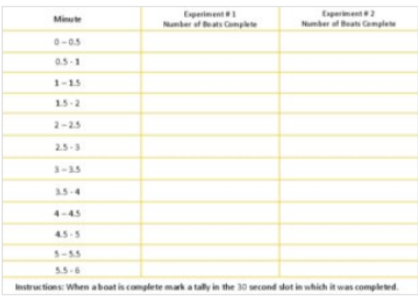
Prerequisites

Equipment

The facilitator(s) will need the following equipment:

<table>
<tr><th>Resource</th><th>Description</th></tr>
<tr><td>OrderGrid.pdf
</td><td>A stack of laminated A4 paper showing a 3 X 3 grid (this forms a stack of orders for boats, and also forms the raw material for each boat).

Around 25 sheets per team of 10 should be sufficient. The sheets will require cleaning between experiments to be re-used.</td></tr>
<tr><td>Water based felt-tipped pens</td><td>Using a water based pen allows you to clean the ink off easily between experiments, or at the end.</td></tr>
<tr><td>Baby Wipes (or similar)</td><td>Baby wipes are also water based, meaning the pen comes off very easily. (Don't use whiteboard pens, we've found felt-tips to be far better!)</td></tr>
<tr><td>Kitchen Roll</td><td>To dry the laminated sheets after the baby wipes.</td></tr>
<tr><td>
</td><td>One A4 sheet (per team) with a grid and a boat drawn on it. This should be cut up into the squares, and given to each station.

The picture can be anything you like, but ensure that some rectangles contain more detail than others, this will make the bottlenecks in the system obvious.</td></tr>
<tr><td>ScoreCard.pdf

<table>
<tr><th>Minute</th><th>Experiment # 1
Number of Boats Complete</th><th>Experiment # 2
Number of Boats Complete</th></tr>
<tr><td>0 – 0.5</td><td></td><td></td></tr>
<tr><td>0.5 - 1</td><td></td><td></td></tr>
<tr><td>1 – 1.5</td><td></td><td></td></tr>
<tr><td>1.5 - 2</td><td></td><td></td></tr>
<tr><td>2 – 2.5</td><td></td><td></td></tr>
<tr><td>2.5 - 3</td><td></td><td></td></tr>
<tr><td>3 – 3.5</td><td></td><td></td></tr>
<tr><td>3.5 - 4</td><td></td><td></td></tr>
<tr><td>4 – 4.5</td><td></td><td></td></tr>
<tr><td>4.5 - 5</td><td></td><td></td></tr>
<tr><td>5 – 5.5</td><td></td><td></td></tr>
<tr><td>5.5 - 6</td><td></td><td></td></tr>
<tr><td colspan="3">Instructions: When a boat is complete mark a tally in the 30 second slot in which it was completed.</td></tr>
</table></td><td>A time score card is used by one person in each team to record the throughput.

Each row represents a 30-second period. The score keeper marks a tally in the 30-second interval in which each boat is completed.</td></tr>
<tr><td>Stopwatch</td><td>One per group of 10 - this is used by the timekeeper to help track the timings of completed boats.</td></tr>
<tr><td>Flip-chart or whiteboard</td><td>To show overall metrics and calculations from the experiments, and help explain any ideas that come up. Note you will need one flip-chart table for each group of 10, or space for both on a whiteboard.</td></tr>
</table>

Facilities

The facilitators will need to organise the attendees around a single table, near enough to each other that a piece of paper can be passed from one to the next without any trouble.

If you are running this for more than 10 people then it is easier to split into two groups, on two separate tables.

The first 9 stations on the table should have:

- A piece of the picture for them to draw - these should be in a sensible order to avoid confusion (we suggest starting top left and working left-to-right across each row in the grid).
- Felt-tipped pen.

The tenth station should have:

- Score card.
- Felt-tipped pen.
- Stopwatch.

Attendees

Minimum: 6

Maximum: For a single group, 10 people is the maximum. You can theoretically have as many groups as you like though. We've proven this will work with 2 groups (i.e. a maximum of 20 attendees).

Instructor Notes

Workshop Introduction

During this workshop, you will actively participate in a flow system running under two sets of processing policies, and draw conclusions from what you experience. This will allow us to explore several key topics, and "what-if" scenarios.

The key areas covered are:

- Flow of work - what is a flow system?
- Different policies
    - Push vs Pull
    - Unlimited queuing vs WIP limiting
    - Stable vs unstable systems
    - Resource vs flow efficiency
    - Effort in both cases
        - Slack time
    - Throughput in both cases
        - Bottlenecks and the Theory of Constraints
    - Value delivered in both cases
    - Lead time
- Little's law
- Systems thinking

Experiment Introduction

The facilitators introduce the simulation as follows:

"We are an awesome boat-building company, and we're going to make a ton of money - just look at this stack of orders! We work in a production-line system to build up each boat over a number of steps. I'm going to show you each a step, so that you can work as a team to build these boats, and help me to become a very rich man."

- Run through each person **teaching / showing them their role in the system**. Point out who is next in the system.
- Allow each group to practice at least one boat, and show the group a **completed boat**.
- **Appoint the time keeper** and explain how to record the throughput.

Experiment 1

"You are now all specialists in your particular area of the build. We will now introduce a couple of policies.

*Policy 1: this is a **push** system. This means when you finish your step, you push your work from your station to the queue ahead of the next station. Policy 2: there are **no limits** on the number of items in these queues. As soon as you are free from one piece of work, you start the next. Take work from the queue in a **FIFO** manner.*

***Quality is good**; but we want to make money! So **speed is better**. Remember, we only make a profit on finished boats, so get your work done and push it on as soon as you can.*

*In order to understand how well our system is performing, we're going to run the system for **two minutes**, and record some timings as we go."*

- Run the system for **2 minutes and then pause (do not reset) the stopwatch**. Ask everyone to stop what they are doing.
- Keep all work in the system and **record the work in progress** (the amount of items currently unfinished in the system).

"We are now going to introduce a new 'Special' order. Once this order exits the system, we will stop the production line. Remember that this is a FIFO system, so you must complete all queued work in order."

- Now **add a new boat** with a special marking.
- Restart the timer, reminding the **time keeper to note throughput as before**.

"In a normal system, orders would continue to flow, so the system would continue to work at 100% utilisation."

- When the **'Special' order reaches the end of the system** then **pause the stopwatch**.
- Now record the following **metrics**:

- Throughput per 30 seconds as recorded by the time keeper.
- Lead time for the special order (this will be the final time on the stopwatch minus 2 minutes).

"We will come back to these numbers after the next experiment."

Experiment 2

*"We are now going to change our policies, and see what effect this has. New policy 1: this is a **pull** system. This means that you cannot push your finished work to the next worker; **they must pull it from you**. There are **no more queues** - you will take your work directly from the previous worker. New policy 2: we are going to impose a **WIP limit** on each worker. The WIP limit will be **1**. This means you may not take new work while you have any work in process. So if the worker ahead of you is at their limit, simply **put up your hands** to indicate you have finished your piece, and wait. When they become free, they will pull the work from you, and this is your signal to begin a new piece of work.*

Again, quality is good, speed is better. But for the purpose of science, consistency with the previous experiment is key. Do not try to make any optimisations - we want to run the experiment with as few variables changed as possible.

We will run the system for 2 minutes and analyse the results."

- People may find that the pull approach is tricky to grasp at first (the instinct is to push work once it is complete), so a **dry run of two boats** is useful at this point.
- **Remind the time keeper of role** to note down throughput as before.
- Run the system for **2 minutes and then pause**.

- Keep all work in the system and **record the work in progress** (the amount of items currently unfinished in the system).

"As before we are now going to introduce a new 'Special' order."

- Now **add a new boat** with a special marking.
- Restart the timer, remind the **time keeper to note throughput as before**.

"In a normal system orders would continue to flow."

- When the **'Special' order reaches the end of the system** then **pause the stopwatch**.
- Now record the following **metrics**:

- Throughput per 30 seconds as recorded by the time keeper.
- Lead time for the special order.

Debrief

- Ensure you have all the **following metrics noted.**

Boats completed per interval	Experiment 1	Experiment 2
0 - 0.5	~~1~~	~~1~~
0.5 - 1	4	5
1 - 1.5	4	6
1.5 - 2	5	4
2.5 - 3	6	~~3~~
3 - 3.5	6	
3.5 - 4	~~3~~	
WIP @ 2 minute mark	22	8
Throughput (average #boats completed per min)	10 bpm	10 bpm
Lead time for special order boat	1 min 57 sec	0 min 40 sec

*"Lets now **calculate the throughput** per minute. We **disregard** the figures for **first slot** and the **last slot**. Why do you think we do this?"*

- Numbers in the first slot are disregarded as the system was not fully primed
- Numbers in the last slot are disregarded as this is usually incomplete data

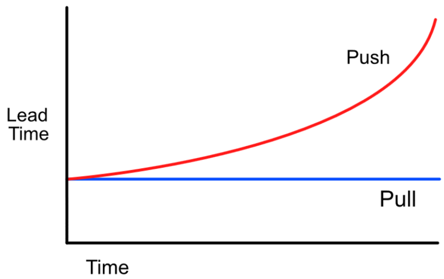
*"Lets now **add up the remaining figures, divide by the number of slots, and multiply by 2** to give us the average throughput per minute."*

> *i.e. for Experiment 1, Throughput is ((4+4+5+6+6) / 5) * 2*

Introducing the workshop, running the experiments and calculating the results generally tends to take 35-45 minutes. From this point forward the remainder of the workshop is based around questions, observations and discussion. It is likely that the attendees will self discover a number of the points covered below. Your job is to facilitate and guide them through the 3 key areas; Metrics, Observations and Thought Experiments. Asking the questions below to ensure the majority of subjects are covered.

Metrics

- The following **Metrics questions** should be discussed:

Question	Notes
What's the difference in the amount of money earned per minute? (Explain how this is a direct result of throughput.)	• There should be little or no difference – the throughput should be the same. • Throughput is constrained by the slowest part of the system (the bottleneck) - even if there are multiple bottlenecks.
How did the lead time for the red ship differ between policies? Why?	• Lead time is longer under the push system (and ever increasing). This is an unstable system – meaning we are unable to effectively predict lead time • Lead time is shorter under the pull system (and consistent). This is a stable system, and lead times are very predictable and consistent. • There is Work in Progress in each case – but with a WIP limit of 1, there are no queues (new work is queued outside the system, so commitment is deferred).
What about our ability to forecast lead time? **What would be the lead time if we captured snapshots of WIP and lead time under each policy at the 5, 10 30 minute mark?**	• With unlimited queues, we are unable to effectively predict lead time - might as well roll a dice. • With a WIP limit, the stability allows a predictable lead time. • The following graph can be used to explain the stable vs unstable systems, and their impact on forecasting:  The red line represents the first experiment, and an unstable system. The line will rise exponentially, the longer the system is running. The blue line represents the second experiment, where the system is stable and the lead time is predictable.

Use Little's law on the figures gathered from the experiments to demonstrate the relationship between lead time, WIP and throughput (LT = WIP / TP).
 	
Use the relationship to show how varying WIP and throughput would affect Lead Time.
	• There are two ways to improve lead time - lower the WIP or increase the throughput.
	• Lowering the WIP is easy to do, and cheap.
	• Increasing the throughput could be done, but is invariably harder and more expensive - you could:
	  ◦ Applying the Theory of Constraints to remove bottlenecks (but this has variable cost, and is only viable up to a point)
	  ◦ Increasing resource (but this is expensive - e.g. it might involve recruitment and/or training of personnel)

What would happen if we continued to lower the WIP? As WIP tends towards 0, does lead time also tend towards 0?	• There is a minimum possible value for lead time. Past this point, it would not matter if you lowered WIP - you would not see a resulting reduction in lead time. • In fact, when you go below this level, throughput will begin to suffer instead. So there is a sweet spot with WIP limits. • A rule of thumb is often used is that 'you can lower your WIP to the number of workstations in your production line'. • Note that in software development, "a workstation" may not be equal to "a person" (pair programming for example).

Observations

- The facilitator can now ask a few **questions about what the attendees observed**.

Question	Notes
How did the attendees feel working under the different policies?	• The pull policy does not always feel natural at first, but is more effective at achieving consistent flow. • Some of the workers should notice slack time as they wait for their downstream partners to pull work from them. • Workers should generally feel less busy, and less stressed as a result - although the throughput remains the same. • Two workers will not see any difference - the worker at the front of the line, and the worker at the worst bottleneck.
How does your relationship with your coworkers change under the two policies?	• With a limited WIP and a pull policy, upstream workers pull work only as it is needed - information flows from the end to the front of the system. • With no WIP and a push policy, the downstream worker pushes work in to the system with no signal to stop when there is too much. There is less information available to the workers in the system about the busy-ness of their co-workers.
Considering the policy on quality did not change (i.e. faster = gooder), was quality affected under the pull policy?	• With the WIP limit, there’s more slack time to spend on the drawing – so quality can naturally improve. • Alternatively, people's slack time could be used to improve their skills, or the process, or to reallocate workers to reduce bottlenecks.
What were your utilisation levels in each system? How much time did you (personally) spend actively working under each policy?	• 100% in Experiment 1 and varied in Experiment 2. • Often companies feel they should optimise for resource utilisation, which is not actually ideal - this would lead to a system like Experiment 1. Instead, the system in Experiment 2 optimises for flow efficiency - ensuring that work flows through the system efficiently, even if this means resources sometime stand idle.

How does this relate to working in real life?	• Knowledge work is different - work items are not uniform, and effort on each one is therefore variable...
	    • However, controlling WIP still allows for more consistency even under a variable system.
	    • Inventory is still dangerous and wasteful (think about unpicking part-finished code changes, or needing to support code changes that were committed but never went live).

Thought experiments

- We can now **explore some theoretical scenarios** - what would happen if we changed the system, or different demands were placed.

Question	Notes

How could delivery of the red boat be sped up?	• Under the push policy, the red boat could be expedited (FIFO -> LIFO) . With the push policy that ship would go through faster - the lead time would be shorter.
	    • But production on the other boats would slow down – so their lead time would increase.
	    • And what happens when other customers find out about the expedite policy? Without WIP limits for urgent work - nothing is urgent!
	• Under the pull policy, the red ship could be prioritised for construction (options).
	• There are many ways to reduce or relieve the bottleneck to speed up delivery of the red boat, of course. But this comes at a cost, as discussed earlier.

What would happen if the red boat order was cancelled?	• With the push policy, the work is wasted - so having inventory is dangerous and wasteful. • Under the pull policy, the work has not yet started, so there is less chance of waste. Late commitment is possible under pull – commit as late as possible, but once a commitment is made, work on it fast and deliver it!
What would happen if one part of the system (not the bottleneck) was optimised (to go faster)?	• Under either policy, that part of the system would go faster. • But under the push system, the overall lead time would go up, as more queues formed. • "Almost always, a local optimisation leads to a global sub-optimisation!" (It's either pointless or harmful - unless we optimise the bottleneck then we won't increase throughput, but if we build up more WIP we increase the Lead Time).
What would happen if a WIP was introduced in only some parts of the system?	• Same thing - local optimisation, global sub-optimisation (as the areas of the system not controlled by a WIP become black holes for inventory).
What would happen if a bonus was introduced in only some parts of the system?	• Local optimisation, global sub-optimisation - PLUS less motivation to help outside the team!

Glossary

Throughput - the rate at which products are delivered (i.e. out of the system). In the case of the workshop, this is the average number of ships finished per minute.

WIP - the work in progress. This can usually be taken to mean one of two distinct things: the total amount of work in a system at a point in time, or the amount of work in a particular workstation within the system. Both these metrics are used in the workshop, so it's important to differentiate between the two. The terms "WIP" vs "WIP limit" are usually used to do this, but this is not always very clear.

Lead time - the total elapsed time from the moment a product enters the system, to the moment it exits the other end. We also refer to this as cycle time.

Little's Law - in a system, the average lead time at a particular point in time can be calculated as the average number of items in progress (WIP) divided by the average rate of items delivered (throughput).

Useful Links

Ship building simulation: https://www.youtube.com/watch?v=iIc9ttGurUo

Klaus Leopold - Flow Exercise: https://www.leanability.com/en/blog-en/2015/09/flow-exercise-building-paper-boats/

Little's Law: https://en.wikipedia.org/wiki/Little%27s_law

A good blog post on the Theory of Constraints: https://blogs.msdn.microsoft.com/ericgu/2016/04/25/agile-and-the-theory-of-constraints-part-1/

Minute	Experiment # 1 Number of Boats Complete	Experiment # 2 Number of Boats Complete
0 – 0.5		
0.5 - 1		
1 – 1.5		
1.5 - 2		
2 – 2.5		
2.5 - 3		
3 – 3.5		
3.5 - 4		
4 – 4.5		
4.5 - 5		
5 – 5.5		
5.5 - 6		

Instructions: When a boat is complete mark a tally in the 30 second slot in which it was completed.