Herbie – An Introduction to WiP Limits

Speaker: John Yorke
Agile Coach at WWT Asynchrony Labs

Email: John.Yorke@asynchrony.com
Twitter: @JohnyVindex
Blog: www.YorkeSoftware.com
Game: MotorCityGame.com
Method formally devised by Taiichi Ohno

1. Visualize Your Work
   - Make ALL work visible
   - Make Policies explicit
   - Identify Workflows

2. Limit Work in Progress

3. Improve Flow
   - Inspect and Adapt
Motor Manufacturing

Motor Car invented by...
• Karl Benz in 1886
Assembly Line

Assembly line invented by...
• Marc Isambard Brunel (1803)

Motor Car Assembly line by...
• Ransom E. Olds

What about Henry Ford?
• He invented the Moving Assembly Line
Why was the Model T only available in Black?

Model T came off the line every 3 minutes, taking just 93 minutes to build one. Only the black paint would dry fast enough until a new paint was developed 12 years later in 1926.
Why does that matter?

An assembly line/replaceable parts is the first time people were not working towards a specific and visible end product.

Prior to that over production or under production was not an issue in the same way.
Terms

Work In Progress
Work in Progress Limits

Why might we use WiP Limits?
1. Help Identify bottlenecks
2. Improve cycle time
3. Reduce Inventory Costs
4. Ability to respond to change more quickly
5. Encourage T-shaped people
6. Lack of explicit policy creates confusion
7. Increase Throughput
Work in Progress Limits

Why might WiP Limits be undesirable?
1. Too tight – can starve your flow and reduce throughput
2. Too loose and you don’t see the problems
3. Not always easy to understand or apply
4. Encourages T-shaped people*
More Terms

- Cycle Time
- Wait Time
- Lead/Delivery Time
Forecasting – Little’s Law

- A Customer has asked when his car will be complete, he is not asking to expedite merely to forecast.

- Little’s Law – Mathematical theory on queues

- Average Cycle time = Average WiP/Average Throughput

BORING!
Little’s Law – in Starbucks

Minutes until I get coffee = \[
\frac{\text{Number of people in line}}{\text{people served per minute}}
\]

11 in line
1 per minute

= 11 minutes to get coffee
Little’s Law – in Starbucks

• Minutes until I get coffee = \frac{\text{Number of people in line}}{\text{people served per minute}}

How do I get served more quickly?

Option 1 – Reduce the number of people in the queue

Option 2 – Increase the number of people served each minute
How might you limit work in progress?

What can we learn from the past?

• Henry Ford, observed excess WiP and reduced space

• Taiichi Ohno, used cards to limit items worked on

• Eli Goldratt, observed that WiP is constrained only by your bottleneck

• Balance production line, all workstations to have same average throughput/cycle time.*
There are 4 broad types of WiP Limits that I have observed in Kanban.

1. Column WiP Limits (Henry Ford)
2. ConWiP Limits (Taiichi Ohno)
3. DBR - Drum, Buffer, Rope (Eli Goldratt)
4. Balanced Production to Maximize Efficiency (USA Motor Industry)*
WiP Limits

If a WiP limit exists it cannot be exceeded at any point in the turn.
Drum-Buffer-Rope - Herbie

Diagram showing the flow of production points in a manufacturing process, with annotations for 'Selected', 'Manufacture', 'Assembly', 'Quality', 'Paint', and 'Done'. The diagram highlights the 'Drum', 'Buffer', and 'Rope' stages.
The Mythical Balanced System

In a balanced system the desire is to attempt to have all workstation’s average production to be equal.

Mathematically this gives the superficial impression that you can maximize utilization without hurting throughput.

However it ignores reality,
• All workstations have statistical variations
• Work (especially in software) is not consistent
• The focus is on utilization rather than throughput.
Motor City Experiment

- Teams will run for 8 iterations (hours) (dice rolled by facilitator)
  - Compare throughput – How many completed cars
  - Teams will review progress, What would you change?
  - Can you see a bottleneck?
Motor City Experiment

- How many cars are still in flight?
- In manufacturing the blocks represent effort and materials, in software it is the same, code is our inventory and it costs real money.
- Our client/Employer pays for our time and so their expense reflects work in progress AND completed items
- Whose cost is highest?
Do WIP limits help us?

• What are the consequences of too tight or too loose limits

• Why should WiP limits be explicit?

• Which WiP limit style is best?
  • In the game, when used correctly Throughput should be the same, only amount of WiP is impacted.
  • Understanding your workflow and your limitations is key. E.g. Only have a designer 1 day a week then ensure there is a buffer of work for them.
  • If Devs regularly get ahead of testing or deployment, then limit Dev column to encourage leaving comfort zone.
  • Simple is usually easier – system wide limit focuses on flow, but be aware of buffering work for constraints
Why might we use WiP Limits?
1. Help Identify bottlenecks
2. Improve cycle time
3. Reduce Inventory Costs
4. Ability to respond to change more quickly
5. Encourage T-shaped people
6. Lack of explicit policy creates confusion

JohnsYorke@gmail.com
www.YorkeSoftware.com
www.MotorCityGame.com