Julia Wester & Cheryl Hammond

Data Driven Coaching
Who are we?

Julia Wester  
Co-Founder & Principal Consultant  
Lagom Solutions  
https://lagom.solutions  
@everydaykanban

Cheryl Hammond  
Hire Me!  
Seattle, Washington, USA  
https://linkedin.com/in/bsktcase  
@bsktcase
Development and Test Metrics 101

- To define the types of metrics used in Agile development and testing
- To show how these metrics are measured and interpreted
- To give the top three ways to begin capturing and using these metrics

WEDNESDAY

DevOps Metrics 101

- To define the types of metrics used for DevOps transformations
- To show how these metrics are measured and interpreted
- To give the top three ways to begin capturing and using DevOps metrics

Agile Quantified - Measuring the Impact of Agility

- To explain how this analysis was performed and why it can be trusted
- To identify the Agile practices that lead to measurable results
- To highlight that different practices affect different dimensions of performance

Cat Swetel
Julia Wester
Dominica DeGrandis
Larry Maccherone
Data Driven Coaching

- To show how team metrics are used during retrospectives
- To show how to avoid over-emphasis on any one metric
- To give immediately usable steps to use metrics and data to drive agile improvement

THURSDAY

Forecasting and Estimation Hands-On

- Discuss the top five reasons forecasts often fail
- Have a go with a very simple forecasting method
- Practice different ways to communicate your forecasts

Quantifying Value and Prioritization
Decision Making Frameworks

- Better understand why talking about value is hard
- Learn how to facilitate discussions about value and urgency of opportunities
- Learn how to prioritize and order work based upon value and the impact of delay
Planning

- WiP Levels
- WiP Age
- Escaped Defects

Stand-Up / Scrum

Time in Process
Throughput

Net Flow
Flow Efficiency
Defects per Environment
and more...

Retrospective

@everydaykanban | @bsktcase

#agile2018
Meaning is in the eye of the beholder

```
<table>
<thead>
<tr>
<th>10 bugs or more assigned to Aliases and Dev Team Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aliases</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Alias 1</td>
</tr>
<tr>
<td>Alias 2</td>
</tr>
<tr>
<td>Team member 1</td>
</tr>
<tr>
<td>Team member 2</td>
</tr>
<tr>
<td>Team member 3</td>
</tr>
<tr>
<td>Team member 4</td>
</tr>
<tr>
<td>Team member 5</td>
</tr>
<tr>
<td>Team member 6</td>
</tr>
<tr>
<td>Team member 7</td>
</tr>
<tr>
<td>Team member 8</td>
</tr>
<tr>
<td>Team member 9</td>
</tr>
<tr>
<td>Team member 10</td>
</tr>
<tr>
<td>Team member 11</td>
</tr>
<tr>
<td>Unassigned</td>
</tr>
</tbody>
</table>
```

“What a slacker!”

“Wow, I bet they’re working on other priorities right now.”

“Makes sense. That person’s code was part of a deep exploratory testing session.”

Courtesy of Troy Magennis (@t_magennis)
The two main goals of data driven coaching:

1. Help teams identify areas of improvement
2. Avoid over-emphasizing a single measure of success
“Tell me how you’ll measure me and I’ll tell you how I’ll behave.

If you measure me in illogical ways, do not complain about illogical behavior.”

—Eli Goldratt
Best bets for measuring logically

Troy Magennis
@t_magennis

Metric advice: Trends NOT numbers, Team NOT individual, Multiple NOT single metric, Similar context NOT everything, Show me when to worry.

35 RETWEETS 38 LIKES
1:06 PM - 17 Feb 2017
Highlight trends, not data points

to avoid distraction from insignificant variation
Trends show us:
How to put a focus on trends

No Y-Axis

Trend Line

Tiny #’s

When applicable, connect data point with lines

NET FLOW PER WEEK (ITEMS COMPLETED - ITEMS STARTED)

More completed than started

More started than completed

@everydaykanban | @bsktcase

#agile2018
Compare trends across teams

Throughput All - Close rate of work items / Devs

Predictability All - Consistency of delivery pace

Better and with company trend

Oops. Still good, but trending adversely

How many work items have been closed: Higher is better.

How variable is work throughput: Lower is better.

Courtesy of Troy Magennis (@t_magennis)
Measure teams, not individuals

if you want people to act like a team
What you measure tells the story of what you value...
“Fear invites wrong figures. Bearers of bad news fare badly. To keep his job, anyone may present to his boss only good news.”

—W. Edwards Deming

“Today we are going to decide who to blame.”
Unexpected responses

you are measuring safety whether you intend to or not

#agile2018

@everydaykanban | @bsktcase
Use multiple competing metrics to create and maintain a stable system
The danger of over-optimizing for one thing...

Measuring this,

but not this.
1. Quality (how well)
   - Escaped defect counts
   - Forecast to complete defects
   - Measure of release “readiness”
   - Test count (passing)

2. Productivity (how much, delivery pace)
   - Throughput (/ team size?)
   - Velocity (/ team size?)
   - Releases per day

3. Responsiveness (how fast)
   - Lead time
   - Cycle time
   - Defect resolution time

4. Predictability (how repeatable)
   - Coefficient of variation (SD/Mean)
   - Standard deviation of the SD
   - “Stability” of team & process

Courtesy of Larry Maccherone (@LMaccherone)
QUALITY
How well?

RESPONSIVENESS
How fast?

PRODUCTIVITY
How much? Keeping pace?

PREDICTABILITY
How repeatable?

Competes with
Are certain quadrants more important for your team?
Define “health” for your context
Decide on **outcomes** THEN
Determine your **metrics**

- Reduced time / cost for rework
- % defects found per environment
- Are we finding defects early enough in process?
- Should we increase time spent on testing?
- % defects found per environment

**INSIGHT**

think of small, day-to-day decisions.

**MEASURE**

Question if the measure really provides expected insight!

**DECISION**

think of small, day-to-day decisions.

**OUTCOME**

think of small, day-to-day decisions.

**THINK**

think of small, day-to-day decisions.

**EFFECT**

think of small, day-to-day decisions.

Courtesy of Larry Maccherone (@LMaccherone)

@everydaykanban | @bsktcase

https://medium.com/@lmaccherone/odim-12d80823222 #agile2018
ESCAPED DEFECTS

defects found in production

Jan 4
Feb 5
Mar 8
Apr 8
May 10
Jun 6
Jul 12
Aug 8
Sep 11

50%
Defects By Environment

percents

KEY

DEV
INT
STAGE
PROD

% OF DEFECTS

Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep

@everydaykanban  |  @bsktcase
#agile2018
QUALITY
How well?

RESPONSIVENESS
How fast?

PRODUCTIVITY
How much? Keeping pace?

PREDICTABILITY
How repeatable?

pick best representative metric for each quadrant
A helpful template to guide you...

We will measure ___________ trended every ___________ (day/week/sprint) as our measure of QUADRANT. We will also measure ___________ to detect if we over-drive improving QUADRANT and suffer elsewhere.
A helpful template to guide you...

We will measure ___________ trended every ___________ (day/week/sprint) as our measure of ______________. We will also measure ___________ to detect if we over-drive improving ___________ and suffer elsewhere.
Behavioral Polarity Worksheet

Measurement Category: Quality, Responsiveness, Productivity, Predictability (circle one)

- Positive results from focusing more this polarity
- Positive results from focusing more this polarity
- Positive results from focusing more this polarity
- Positive results from focusing more this polarity

- Totally ignore this metric
- Totally ignore this metric

- Early warning signs that you are getting into the negative aspect of this side
- Early warning signs that you are getting into the negative aspect of this side

- Negative results from focusing more this polarity
- Negative results from focusing more this polarity
- Negative results from focusing more this polarity
- Negative results from focusing more this polarity

- Focused Objective

Courtesy of Troy Magennis (@t_magennis)
<table>
<thead>
<tr>
<th>Decision</th>
<th>Natural Tendency</th>
</tr>
</thead>
<tbody>
<tr>
<td>What decision are you trying to make?</td>
<td>Check the box in the section below that resembles your normal operating condition</td>
</tr>
</tbody>
</table>

**Too Little (Extreme)**

What consequences occur when you operate at this extreme... positive and negative!

**Lagom (Just Enough)**

What are the ideal outcomes when you do this just right?

**Too Much (Extreme)**

What consequences occur when you operate at this extreme... positive and negative!

**Hypothesis**

What steps do you think will get you closer to lagom?

**Plan / Do**

- How will you test your hypothesis?
- How will you know if you are closer to lagom?
- When will you review?

**Study**

What do your success metrics tell you? Are you closer to lagom, farther away or the same distance?

**Adjust**

Do you need to tweak the plan & continue or make a new hypothesis?
A helpful template to guide you...

We will measure \( \underline{\text{Time in Process}} \) trended every \( \underline{\text{Sprint}} \) (day/week/sprint) as our measure of \( \underline{\text{RESPONSIVENESS}} \). We will also measure \( \underline{\text{Escaped Defects}} \) to detect if we over-drive improving \( \underline{\text{RESPONSIVENESS}} \) and suffer elsewhere.

\( \underline{\text{QUALITY}} \)
Give a cheat sheet!

- What is the chart
- Intended behavior
- How to game it
- When overdriven, what moves?


---

**Team Dashboard Spreadsheet Cheat Sheet**

<table>
<thead>
<tr>
<th>When overdriven, what moves?</th>
<th>What is the intended behavior?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality metric goes up as more defects reported.</td>
<td>Help teams understand the time in-process (cycle time) for completing items, and to look for ways to reduce the time (and variability) by eliminating unnecessary steps and waste.</td>
</tr>
<tr>
<td>Predictability metric goes up as more defects reported.</td>
<td>How is it gamed?</td>
</tr>
<tr>
<td>Teams can prematurely sign-off work only to have defects reported later (seen by increase in quality defect ratio).</td>
<td>Only fast simple work is started (seen by initial increasing productivity only to regress later)</td>
</tr>
<tr>
<td>Teams can slow down starting new work (seen by a decrease in throughput productivity)</td>
<td>Premature sign-off of items as finished (seen by decreasing defect counts)</td>
</tr>
</tbody>
</table>

**Productivity - How Much**

- This chart measures productivity and shows the number of tasks completed in a given time frame.
- The chart illustrates the trend and peak values for productivity over time.

**Responsiveness - How Well**

- Responsiveness is measured as the median (blue line) calendar time in-progress for all items completed in the same week. The minimum, 25th & 75th percentile and maximum values are shown to see the variability each week.

**What is the intended behavior?**

- Help teams continuously triage and complete defect backlog items at a consistent rate. Try and drive this percentage lower without deferring defects.

**How is it gamed?**

- Defects are entered as additional stories (seen by growing throughput, but lowering customer satisfaction and complaints).
- Responsiveness metric increases for defect types.
- The number of open defects grows if story work is done in preference.

**When overdriven, what moves?**

- Quality metric goes up as more defects reported.
- Predictability metric grows into higher peaks as more work and more defects get opened.

---

For more probability forecasting resources: focusedobjective.com Email: troy.magennis@focusedobjective.com

Work licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

#agile2018

@everydaykanban  |  @bsktcase

Courtesy of Troy Magennis (@t_magennis)
Sometimes it’s ok to sacrifice a bit in one area to bring the whole into harmony
Compare only in context to ensure you highlight meaningful differences
Bad Comparisons lead to Bad Insights which lead to Bad Decisions
Show me when to worry

don’t make me guess or waste my time off
“First, do no harm.”

“If anyone adjusts a stable process, the output that follows will be worse than if (s)he had left the process alone.”

–W. Edward Deming
Not all variation is cause for alarm

WiP Aging

Courtesy of Cat Swetel (@catswetel)

@everydaykanban | @bsktcase

#agile2018
Use markers to show what’s concerning

Courtesy of Cat Swetel (@catswetel)
Recap: The two main goals of data driven coaching:

1. help teams identify areas of improvement
2. avoid over-emphasizing a single measure of success
Recap: Key Tips to Remember

Metric advice: Trends NOT numbers, Team NOT individual, Multiple NOT single metric, Similar context NOT everything, Show me when to worry.
What steps will you take next?
Resources to keep learning

Balanced Metrics Dashboard
http://focusedobjective.com/team-metrics-right/
by Troy Magennis, Focused Objective

Lagom Discovery Canvas
by Julia Wester, Lagom Solutions

One Metric to Rule Them All
https://bit.ly/2M3gVCE
by Cheryl Hammond, Secret Garden
Data Driven Coaching

- To show how team metrics are used during retrospectives
- To show how to avoid over-emphasis on any one metric
- To give immediately usable steps to use metrics and data to drive agile improvement

THURSDAY

Forecasting and Estimation Hands-On

- Discuss the top five reasons forecasts often fail
- Have a go with a very simple forecasting method
- Practice different ways to communicate your forecasts

Quantifying Value and Prioritization

Decision Making Frameworks

- Better understand why talking about value is hard
- Learn how to facilitate discussions about value and urgency of opportunities
- Learn how to prioritize and order work based upon value and the impact of delay

@everydaykanban  |  @bsktcase  
#agile2018
How You Can Contact Us

Julia Wester  
Co-Founder & Principal Consultant  
Lagom Solutions  
https://lagom.solutions  
@everydaykanban

Cheryl Hammond  
Hire Me!  
Seattle, Washington, USA  
https://linkedin.com/in/bsktcase  
@bsktcase
Appendix (previous slides)
Pro Tip:
Pies are best for eating, less so for charting...

Poll: are doughnuts a gateway to bad charting decisions?

Question: Does consuming sugary foods like delicious doughnuts prime people to think pie charts are an appropriate way to encode complex data?

33.3% NO
Even better, because they’re so delicious.

33.3% YES
I mean, this chart is proof. But they’re so delicious.

33.3% MAYBE
But let me eat my delicious doughnut before answering.

Source: Golden Doughnut

Percentages do not equal 100 because of rounding

http://www.thefunctionalart.com/2012/06/fun-note-on-pie-charts.html

#agile2018
Counterpoint!
Why not both?
Don’t
Just don’t
Correlation isn’t causation
Avoid Confusing Activity with Progress

Activity metrics

- lines of code
- # of issues closed
- # of deploys per day

These aren’t “bad” metrics. We should:

1. Understand what they really show
2. Show them with competing metrics to reduce gaming
Know The Expiry Dates For Your Metrics
Goodhart’s Law

“When a measure becomes a target, it ceases to be a good measure.”
Exercise: Map outcomes to metrics with ODIM

1. List one **outcome** you are trying to achieve.

2. Determine a **decision** you need to make to achieve the outcome.

3. List an **insight** that can help you make the decision.

4. List one or more **metrics** that can provide the needed insight.

5. **Repeat** until time runs out.