

# ***Practical Software and Systems Measurement***

***Objective Information for Decision Makers***



**Workshop:**  
***A Path Toward Consensus Measures for Iterative Software Development***

***17 September 2019***

***Cheryl Jones – US Army CCD-C-AC***  
***Geoff Draper – L3Harris Technologies***

# **Workshop Summary Results**

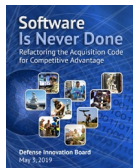
- **Intro – see presentation slides**
- **Interactive Evaluation Feedback of PCM Agile Framework**
  - **Part I: Information Needs**
  - **Part II: Measures**
  - **Part III: Preliminary Analysis Results**
  - **Next Steps**

# Workshop Background

## DSB SW



## DIB SWAP

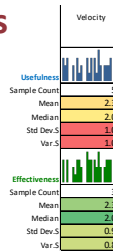


<https://innovation.defense.gov/software/>

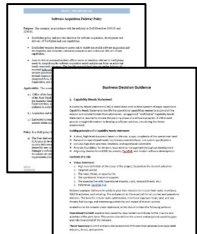
## Info Needs Measures

### Surveys

- PSM
- NDIA
- INCOSE
- SERC



## Draft DoD SW policy



<https://www.ndia.org/divisions/systems-engineering/studies-and-publications>

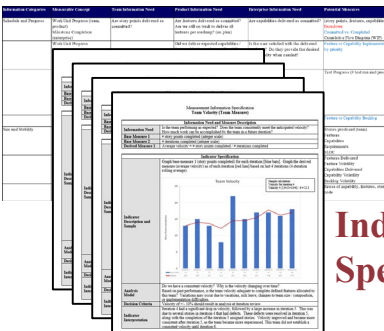
## Studies and Publications

### 2019 Reports

- Implementing Continuous Iterative Development and Acquisition

Defense Science Board (DSB) released a report in Feb-2018 containing seven recommendations regarding software design and acquisition. Section 808 of NDAA 2019 mandates implementation of these recommendations within 18 months. The Defense Innovation Board (DIB) Software Acquisition and Practices (SWAP) study group has also provided many insightful and largely compatible recommendations.

NDIA, INCOSE and PSM support the DSB and DIB concepts and the opportunities they offer to DoD and the defense industry. In 2018 the NDIA Systems Engineering Division commissioned a working group to study the implementation of the DIB and SWAP recommendations. NDIA offers the working group's recommendations below to OUSD(A&S) and OUSD(R&E) representing a broad "industry perspective" on a path forward.



## ICM Table

## Indicator Specs

Acceleration	Automated Test Coverage	Burndown (Sprint/Release)	Defect Containment
Defect Escapes	Defect Resolution	Cycle Time	Lead Time
Release Frequency	Velocity		

**This workshop advances development of a proposed consensus industry framework for continuous iterative development**

**Building on DoD and industry initiatives to improve the acquisition and execution of defense software programs**

**Follow on implementation from recommendations of joint NDIA/INCOSE/PSM WG**

## ***Objectives of the Workshop***

- ***Evaluate candidate measures for continuous iterative development to advance the draft PSM framework***

## ***Workshop Format***

- ***Interactive evaluation of candidate measures (usefulness, effectiveness)***
- ***Review, evaluate, and improve draft measurement indicator specs***

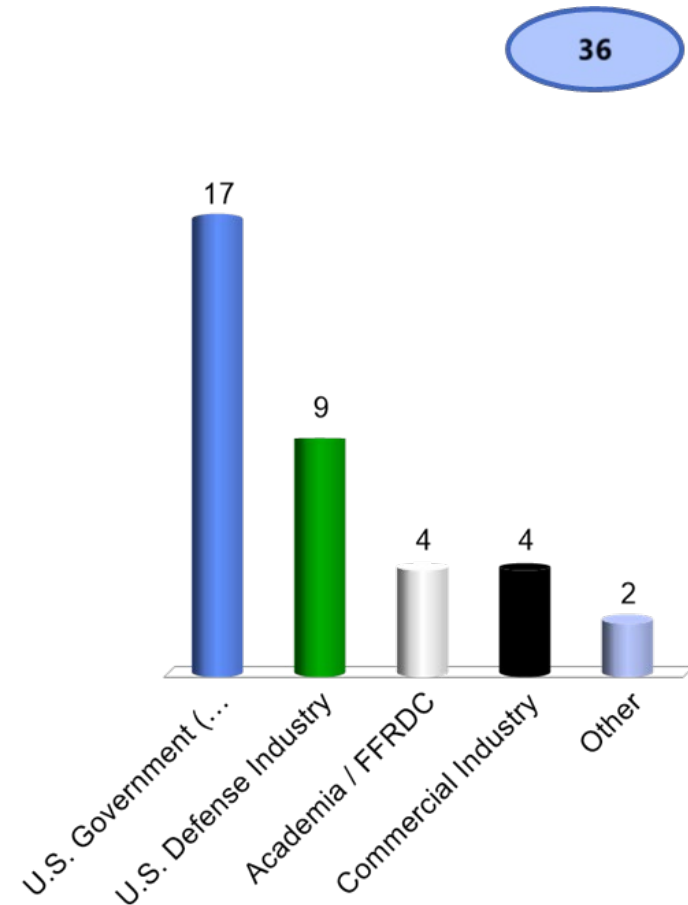
## ***Intended Output***

- ***ICM Table and measurement specs for agile measurement that are ready for use***
- ***Plan for white paper on measurement for continuous iterative development including an outline and writing assignments***

# **Part I: Evaluation of Information Needs**

***How would you best characterize your organization?***

- 1. U.S. Government (DoD, agency)**
- 2. U.S. Defense Industry**
- 3. Academia / FFRDC**
- 4. Commercial Industry**
- 5. Other**

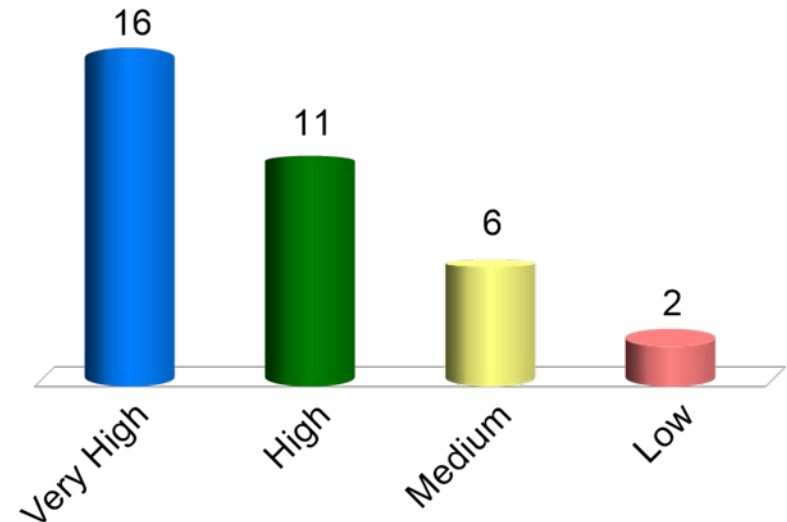




## Are new story points, features, or capabilities being delivered as committed?

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

1. **Very High**
2. **High**
3. **Medium**
4. **Low**



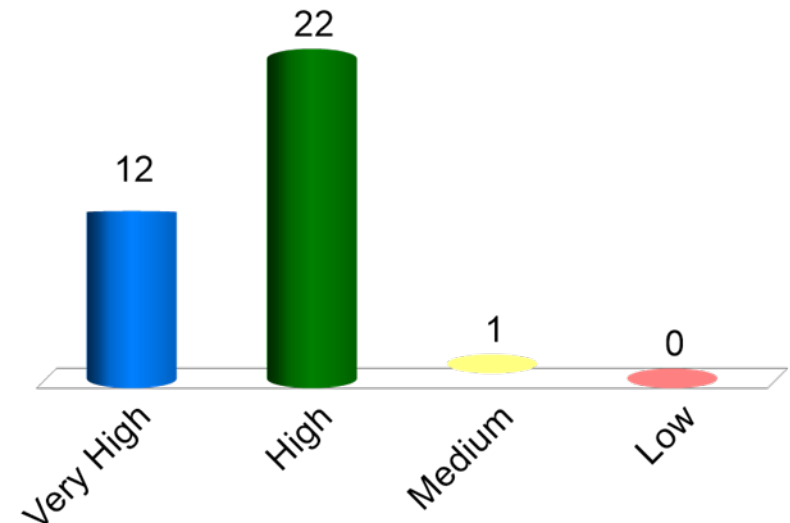
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Schedule and Progress	Work Unit Progress (team, product) Milestone Completion (enterprise)	Are story points delivered as committed?	Are features delivered as committed? Are we still on track to deliver all features per roadmap? (on plan)	Are capabilities delivered as committed?

35

## ***Do delivered products provide the expected functionality to users when needed?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



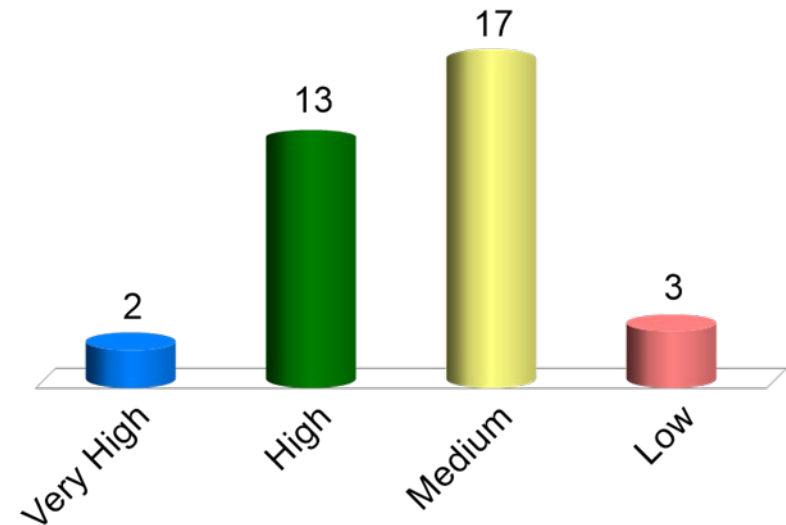
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Schedule and Progress	Work Unit Progress		Did we deliver expected capabilities / features? Is the roadmap still valid?	Is the user satisfied with the delivered products? Do they provide the desired functionality when needed?

35

## ***How much technical or mission debt exists in the backlog?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. *Very High***
- 2. *High***
- 3. *Medium***
- 4. *Low***



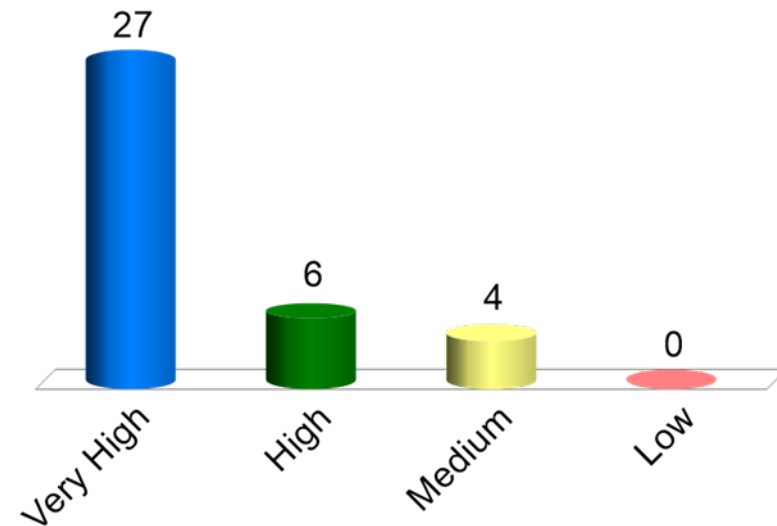
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Schedule and Progress	Work Backlog		How much outstanding technical or mission debt exists?	

35

## ***Is the product correct? Does new code functionality work as expected?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



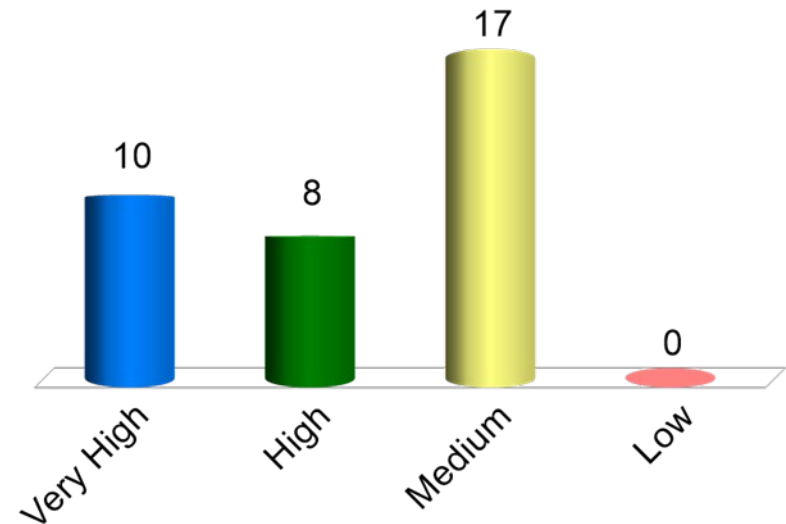
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Product Quality	Functional Correctness	Does new code functionality work as expected?	Does new code functionality work as expected?	Is rework identified and managed?

**37**

## ***Does new code break previous functionality? (change failure rate, rollback)***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



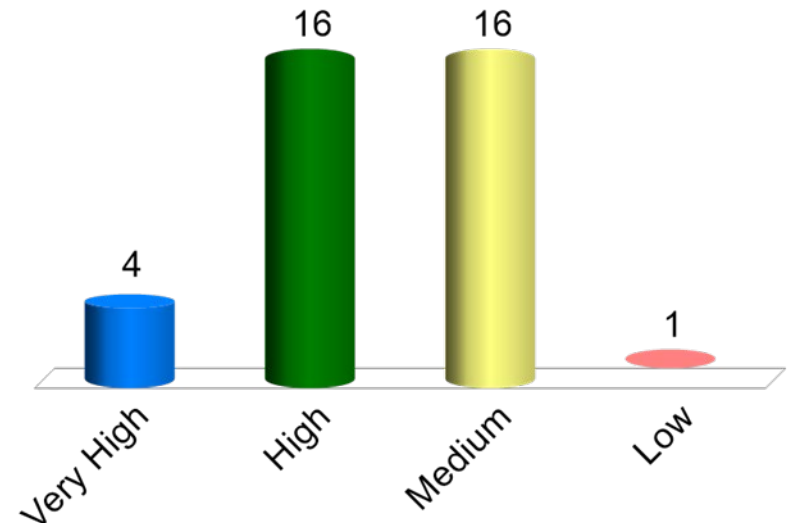
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Product Quality	Functional Correctness	Does new code break previous functionality?	Does new code break previous functionality? (change failure rate, rollback)	

35

## ***How many defects escape the increment?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



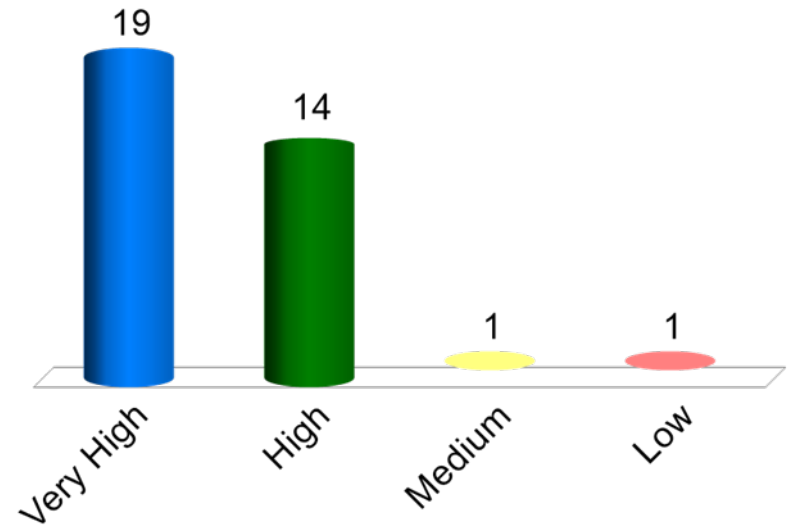
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Product Quality	Functional Correctness		How many defects escape the increment?	

37

## ***What is the quality of code deployed to the field?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



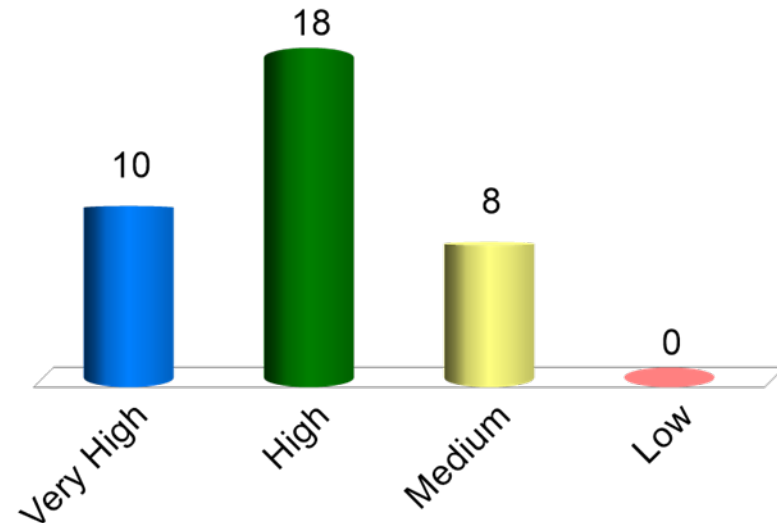
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Product Quality	Functional Correctness		What is the quality of code deployed to the field?	What is the quality of code deployed to the field?

35

## ***What is the reliability and availability of operational service capabilities?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Product Quality	Supportability - Maintainability Dependability - Reliability		What is the reliability and availability of operational service capabilities?	

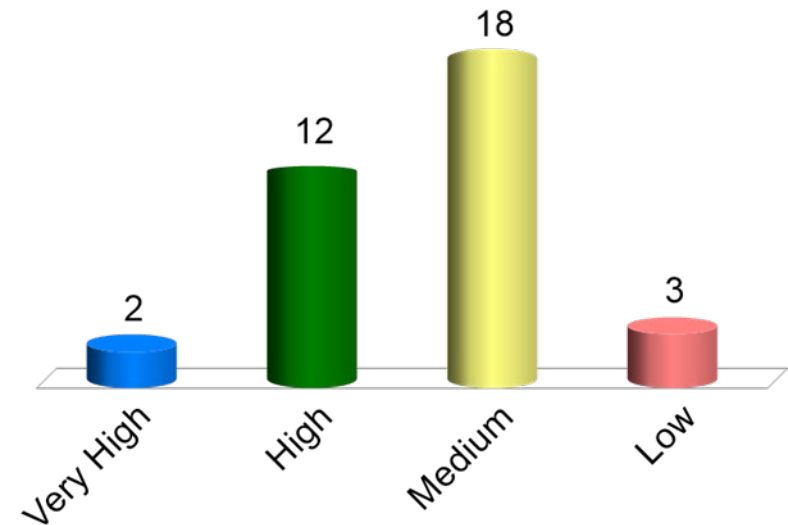
36



## ***How quickly can we address bug reports from the field?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



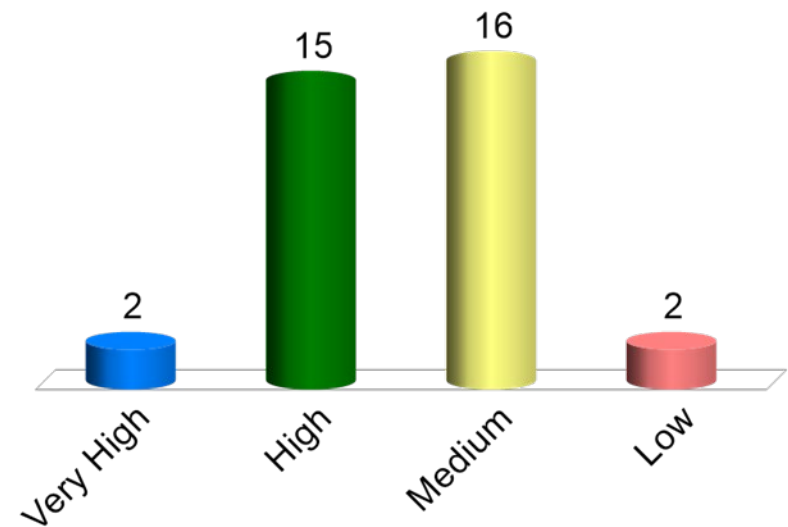
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Efficiency - Speed Supportability - Maintainability Dependability - Reliability		How quickly can we address bug reports from the field?	

35

## *Are teams performing as productively as expected?*

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

1. **Very High**
2. **High**
3. **Medium**
4. **Low**



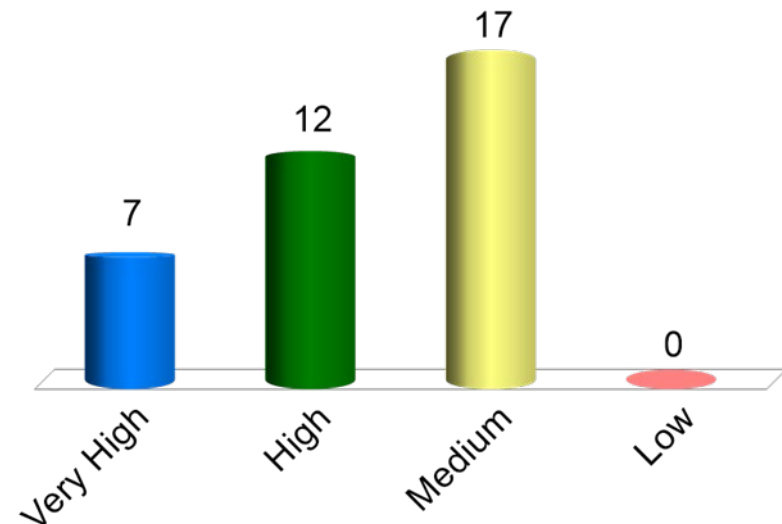
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Efficiency - Speed	Is the team performing as expected?	Are teams performing as expected?	

35

## *How long does it take to deploy an identified feature/capability?*

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

1. **Very High**
2. **High**
3. **Medium**
4. **Low**



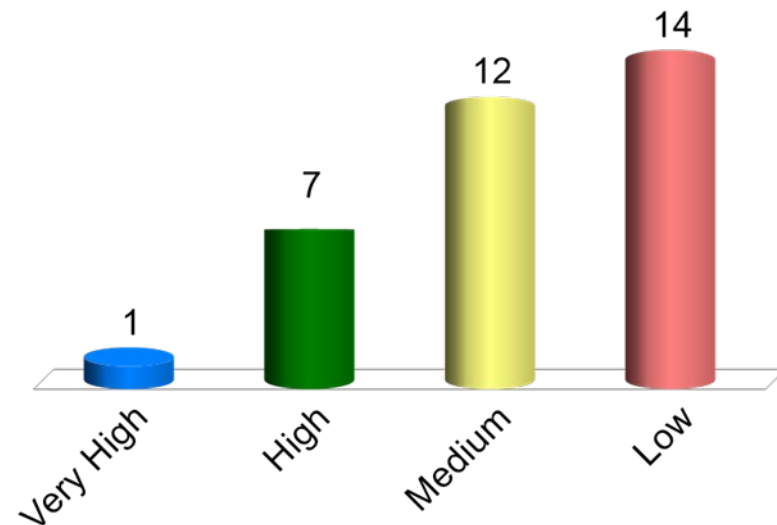
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Efficiency - Speed		How long does it take to deploy an identified feature/capability?	

36

## ***What is the frequency of product release or deployment?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. *Very High***
- 2. *High***
- 3. *Medium***
- 4. *Low***



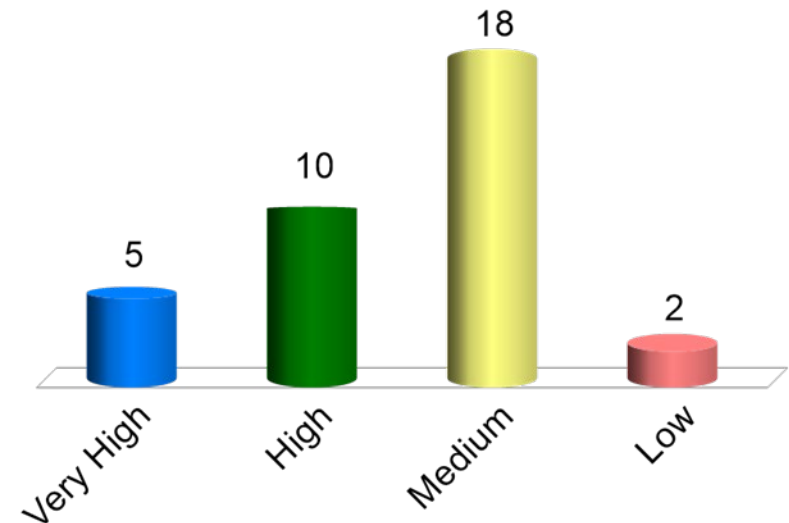
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Efficiency - Speed		What is the frequency of product release or deployment?	What is the frequency of product release or deployment?

34

## ***How long does it take to release a viable product?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



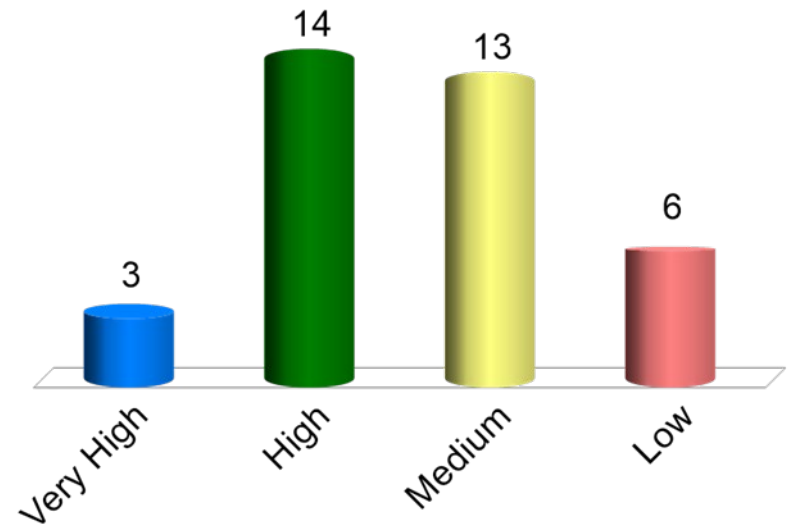
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Efficiency - Speed	How long does it take to release a viable product?	How long does it take to release a viable product?	How long does it take to release a viable product?

35

## ***How much of the testing is automated? How often do we perform automated testing?***

What importance would you place on this measurement information need for planning and managing continuous iterative SW development?

- 1. Very High**
- 2. High**
- 3. Medium**
- 4. Low**



36

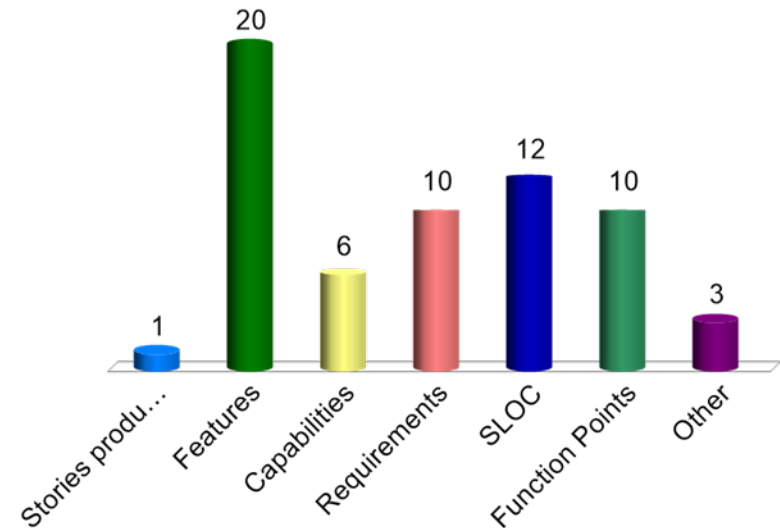
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need
Process Performance (Process Effectiveness)	Process Effectiveness		How much of the testing is automated? How often do we perform automated testing?	How much of the testing is automated? How often do we perform automated testing?

## How big is our system?

How should we count size for continuous iterative development programs (e.g., for estimating)?

*(pick up to 2 choices)*

1. **Stories produced (team)**
2. **Features**
3. **Capabilities**
4. **Requirements**
5. **SLOC**
6. **Function Points**
7. **Other**



Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need	Potential Measures
Size and Stability	Functional Size and Stability Physical Size and Stability	How big is our system?	How big is our system?	How big is our system?	Stories produced (team) Features Capabilities Requirements SLOC

## ***Part II: Evaluating Candidate Measures for Continuous Iterative Development***

<b>PSM <i>**Draft**</i></b>
<b>Burndown (sprint/release)</b>
<b>Velocity</b>
<b>Acceleration</b>
<b>Cycle time</b>
<b>Lead time</b>
<b>Release frequency</b>
<b>Defect containment</b>
<b>Defect escapes</b>
<b>Defect resolution</b>
<b>Automated test coverage</b>
Core PSM framework: <ul style="list-style-type: none"><li>• Cost (est. vs. actual)</li><li>• Schedule (est. vs. actual)</li><li>• Staffing</li><li>• ...etc.</li></ul>

See PSM framework for details.

- Information categories
- Measurable concepts
- Information needs
- Cross-reference mappings

Additional candidate measures are defined in draft ICM table but not implemented in first release.



## ICM Table (Draft) Excerpts most relevant to PSM CID workshop – 1 of 3

Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need	Potential Measures
Schedule and Progress	Work Unit Progress (team, product) Milestone Completion (enterprise)	Are story points delivered as committed?	Are features delivered as committed? Are we still on track to deliver all features per roadmap? (on plan)	Are capabilities delivered as committed?	(story points, features, capabilities) <b>Burndown</b> <b>Committed vs. Completed</b> Cumulative Flow Diagram (WIP)
	Work Unit Progress		Did we deliver expected capabilities / features? Is the roadmap still valid?	Is the user satisfied with the delivered products? Do they provide the desired functionality when needed?	<b>Feature or Capability Implementation by priority</b>
	Work Unit Progress		Is the integration and test progress proceeding as planned?		Test Progress (# test run and passed)
	Work Backlog		How much outstanding technical or mission debt exists?		<b>Feature or Capability Backlog</b>
Size and Stability	Functional Size and Stability Physical Size and Stability	How big is our system?	How big is our system?	How big is our system?	Stories produced (team) Features Capabilities Requirements SLOC
	Functional Size and Stability		How volatile are capabilities or features? Are we adding more features? What is the ability to accommodate changes in customer desiresments?	How volatile are capabilities or requirements? What is the ability to accommodate changes in customer desiresments?	Features Delivered Feature Volatility Capabilities Delivered Capability Volatility Backlog Volatility
	Functional Size and Stability	How much of the product is newly developed vs. reused from other sources?			Reuse of capability, features, stories, code

## ICM Table (Draft) *Excerpts most relevant to PSM CID workshop – 2 of 3*

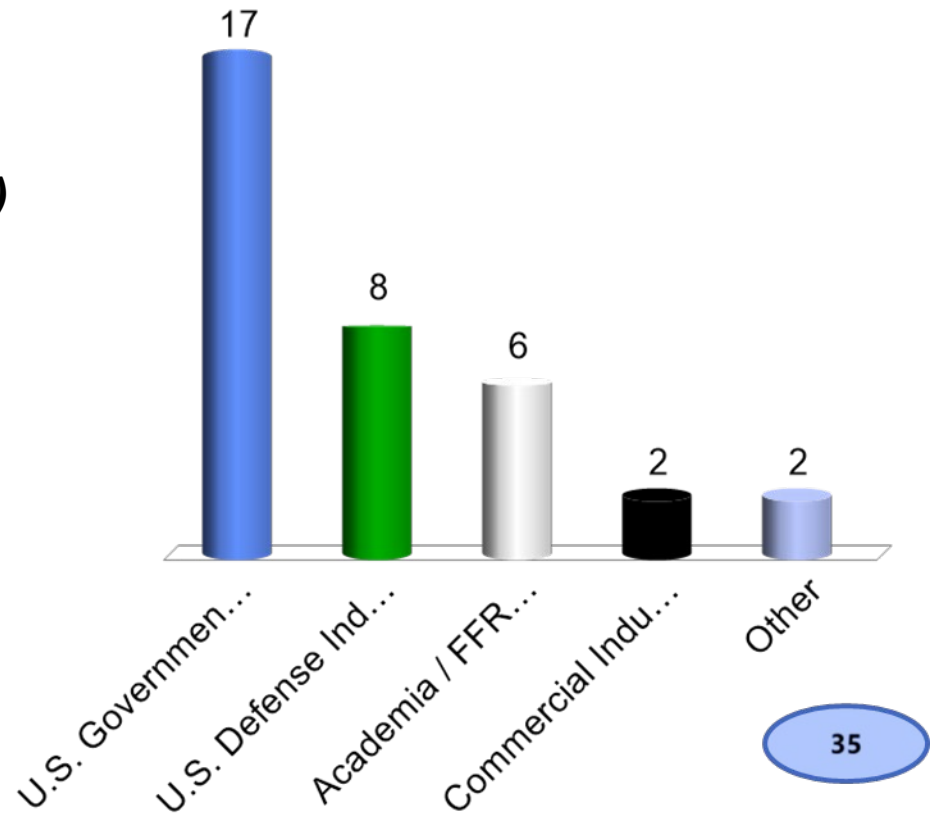
Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need	Potential Measures
Product Quality	Functional Correctness	Does new code functionality work as expected?	Does new code functionality work as expected?	Is rework identified and managed?	Stories Accepted (increment demo) Rework Stories Change Reports (defects) Written
	Functional Correctness	Does new code break previous functionality?	Does new code break previous functionality? (change failure rate, rollback)		Change reports (defects) written <a href="#">Rework hours</a> Rework stories <a href="#">Change Failure Rate</a> or Defect Density
	Functional Correctness		How many defects escape the increment?		<a href="#">Defects Found in Pipeline (saves)</a>
	Functional Correctness		What is the quality of code deployed to the field?	What is the quality of code deployed to the field?	<a href="#">Defect Escapes to field</a> <a href="#">Defect Escape Ratio</a>
	Security - Safety		How secure is the product		Vulnerabilities by severity
	Supportability - Maintainability Dependability - Reliability		What is the reliability and availability of operational service capabilities?		Mean-Time-To: MTTD (Detect) <a href="#">MTTR (Repair or Restore)</a> MTBF (Between Failure) MTTF (Failure) Ao (Operational Availability)

## ICM Table (Draft) *Excerpts most relevant to PSM CID workshop – 3 of 3*

Information Categories	Measurable Concept	Team Information Need	Product Information Need	Enterprise Information Need	Potential Measures
Process Performance (Process Effectiveness)	Process Efficiency - Speed Security - Safety		How quickly can new security vulnerabilities be patched and deployed to fielded products?		Security vulnerability lead time Mean Time to Restore
	Process Efficiency - Speed Supportability - Maintainability Dependability - Reliability		How quickly can we address bug reports from the field?		Mean Time to Restore MTTD
	Process Efficiency - Speed	Is the team performing as expected?	Are teams performing as expected?		Velocity (average story points per increment) Capacity (staffhours per increment) Story points delivered vs. committed (on average) Cumulative flow diagrams
	Process Efficiency - Speed		How long does it take to deploy an identified feature/capability?		Lead time
	Process Efficiency - Speed		What is the frequency of product release or deployment?	What is the frequency of product release or deployment?	Release or deployment frequency
	Process Efficiency - Speed	How long does it take to release a viable product?	How long does it take to release a viable product?	How long does it take to release a viable product?	Release frequency Cycle time (increment, release, mean/median) Time to Minimum Viable Product (MVP)
	Process Efficiency - Speed		How much time does it take to conduct a full regression test? How much time for the automated regression test?		Test duration Automated test duration
	Process Effectiveness		How much of the testing is automated? How often do we perform automated testing?	How much of the testing is automated? How often do we perform automated testing?	Automated test coverage Automated test frequency
	Process Effectiveness	Is the backlog being managed appropriately?	Is the backlog being managed appropriately?		Cumulative flow diagram Backlog readiness
Customer Satisfaction	Customer Support		How long does it take to get a viable product released? (specific)	How long does it take to get a viable product released? (multiple systems) - time to market	Time to Minimum Viable Product (MVP)

## ***How would you best characterize your organization?***

- 1. U.S. Government (DoD, agency)**
- 2. U.S. Defense Industry**
- 3. Academia / FFRDC**
- 4. Commercial Industry**
- 5. Other**



## Evaluations of Measures for Continuous Iterative Development

Two separate evaluations are desired:

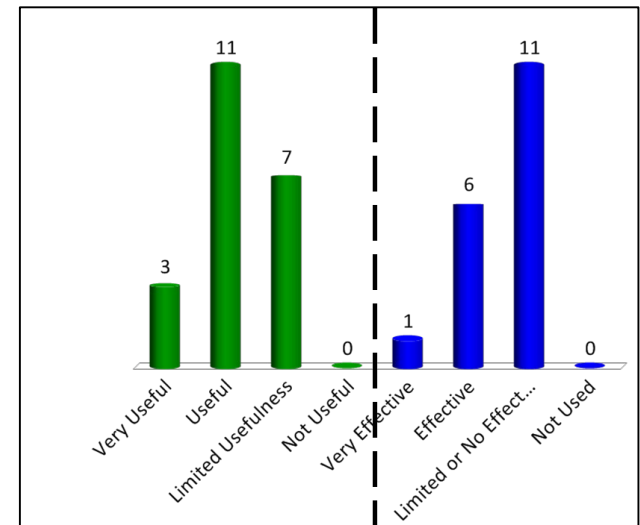
- **Usefulness:** Is the measure itself useful for providing insight?
- **Effectiveness:** How effectively does your organization use it?

**THEORY**  
**PRACTICE**

### Example:

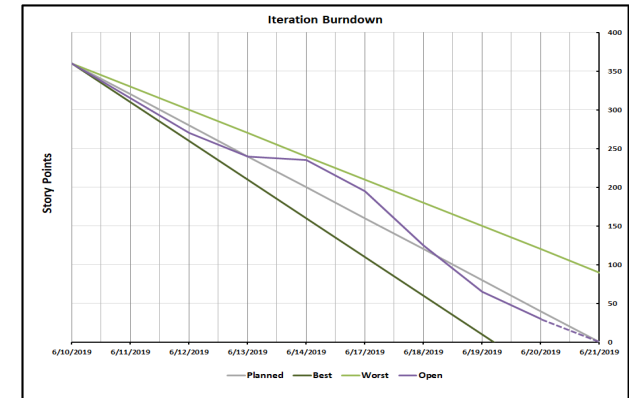
Is [measure name] a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

↑ Insight ↓	1.	Very Useful	Select 1 from here
	2.	Useful	
	3.	Limited Usefulness	
	4.	Not Useful	
-----			
↑ Impact ↓	5.	Very Effective	and 1 from here
	6.	Effective	
	7.	Limited or No Effectiveness	
	8.	Not Used	



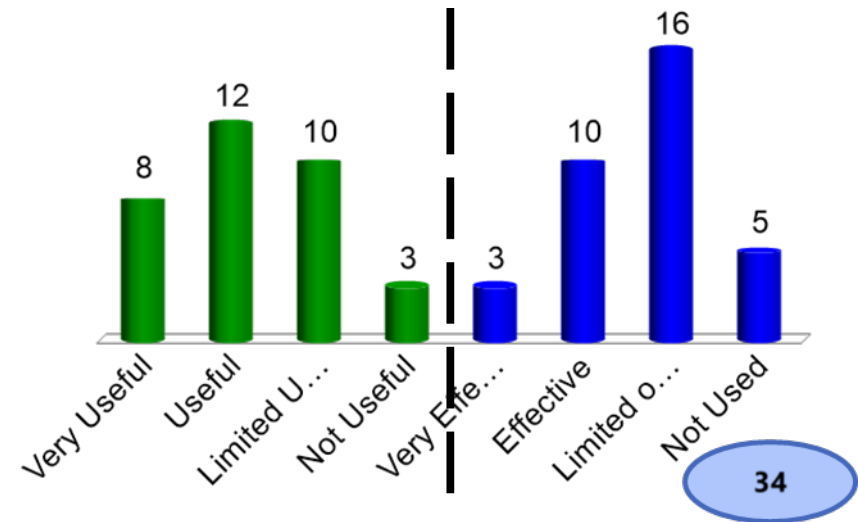
Is **Sprint Burndown** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Information Need and Measure Description	
Information Need	What is the status of the iteration? Will all the remaining committed work be completed by the end of the iteration? Will the team deliver the committed story points?
Base Measure 1	Planned story points (integer scale)
Base Measure 2	Completed story points (integer scale)
Derived Measure 1	Open story points = planned story points – completed story points



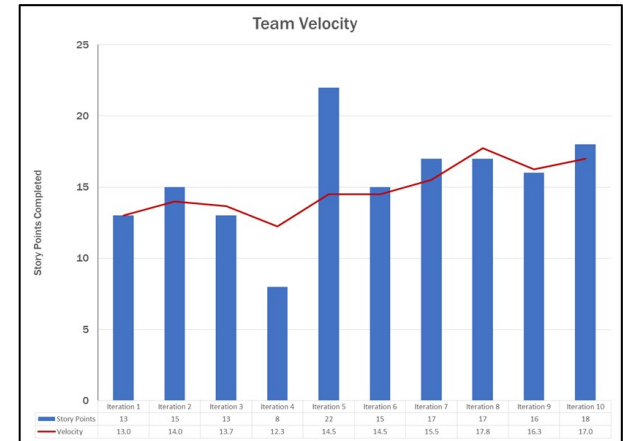
- ↑ **Insight**
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓
- ↑ **Impact**
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**
- ↓

(select 1 from each)

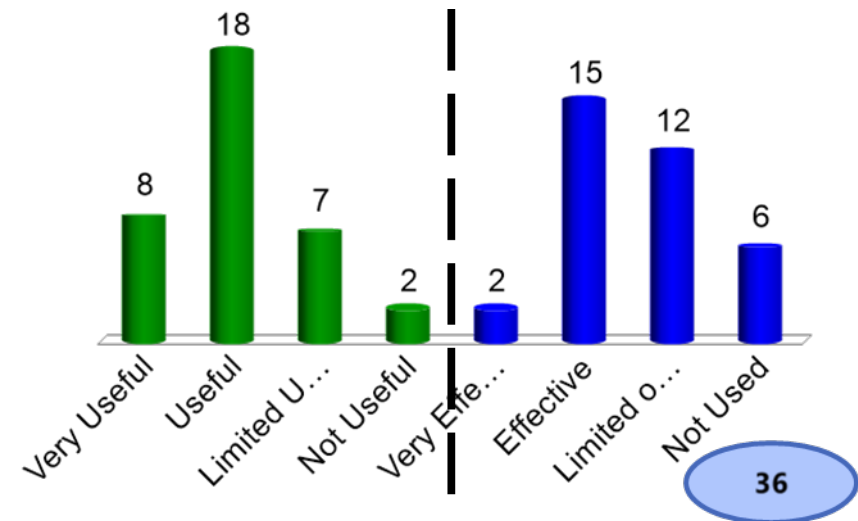


Is **Velocity** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Information Need and Measure Description	
Information Need	Is the team performing as expected? Does the team consistently meet the anticipated velocity? How much work can be accomplished by the team in a future iteration?
Base Measure 1	# story points completed (integer scale)
Base Measure 2	# iterations completed (integer scale)
Derived Measure 1	Average velocity = # story points completed / # iterations completed



- ↑ Insight
- ↓ Impact
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
  5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**



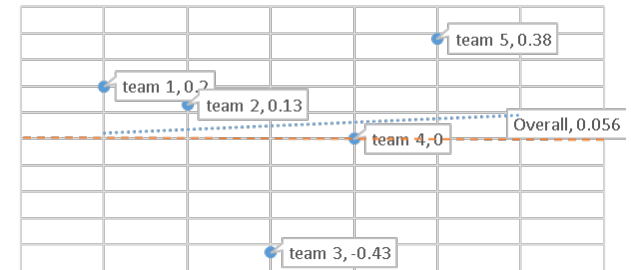
(select 1 from each)

Is **Acceleration** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

	Iter 1 pts	Iter 2 pts	Acceleration
team 1	10	12	0.2
team 2	8	9	0.13
team 3	14	8	-0.43
team 4	12	12	0
team 5	8	11	0.38
Overall			0.056

Sample calculation:  
Team 1 acceleration =  $12 - 10 / 10 = .2$   
(20% positive acceleration)

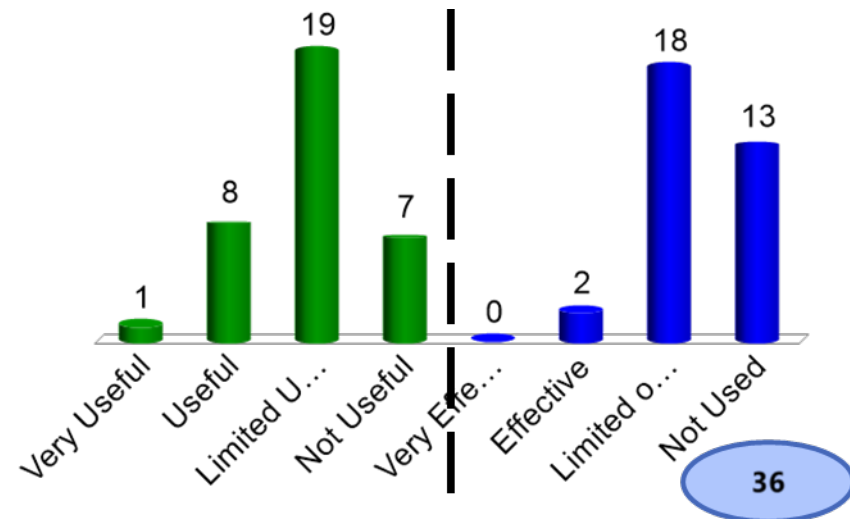
Iteration 1-2 Acceleration



Information Need and Measure Description	
Information Need	Is the team's productivity increasing, decreasing or holding steady?
Base Measure 1	# story points completed (velocity) this increment (integer scale)
Base Measure 2	# story points completed (velocity) in previous comparison increment (integer scale)
Derived Measure 1	Team Acceleration = (Current increment velocity - comparison increment velocity) / comparison increment velocity
Derived Measure 2	Overall Acceleration = Team Acceleration 1 .... Team Acceleration N / N

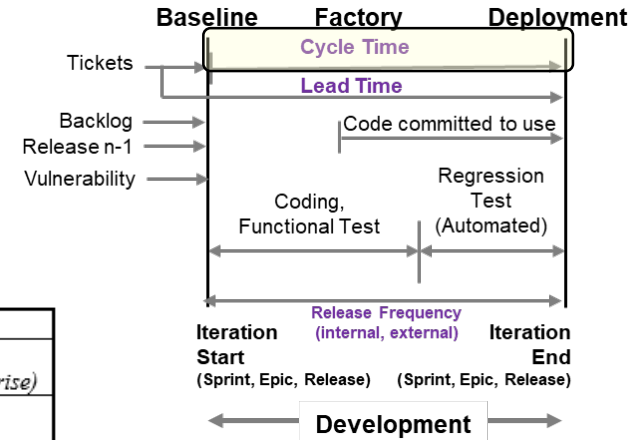
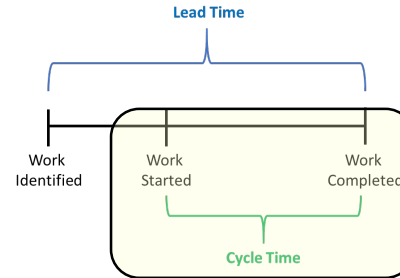
- ↑ Insight
1. Very Useful
  2. Useful
  3. Limited Usefulness
  4. Not Useful
- ↓
- ↑ Impact
5. Very Effective
  6. Effective
  7. Limited or No Effectiveness
  8. Not Used
- ↓

(select 1 from each)





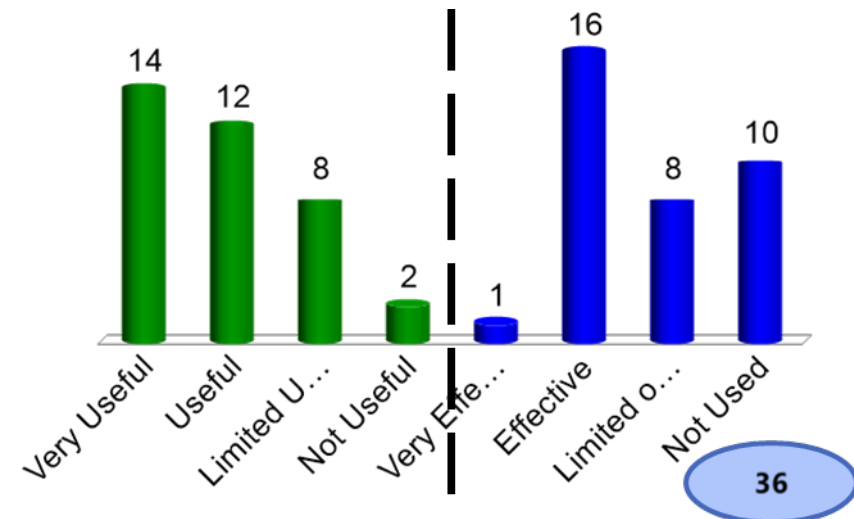
Is **Cycle Time** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?



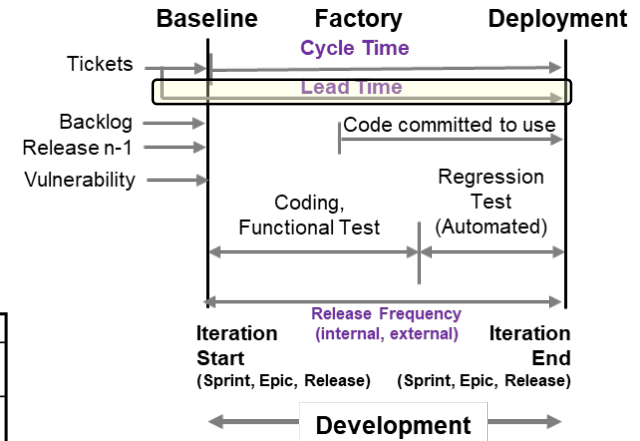
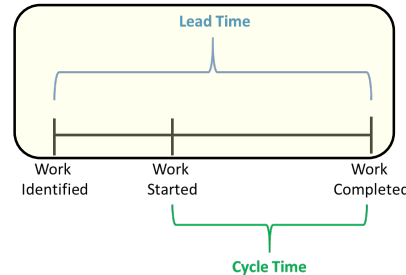
Information Need and Measure Description	
<b>Information Need (Cycle Time)</b>	How long does it take to complete a process activity? ( <i>team</i> ) How long does it take to develop an identified feature/capability or release? ( <i>product or enterprise</i> )
<b>Information Need (Lead Time)</b>	How long does it take to get a viable product released (time to market)?
<b>Base Measure 1</b>	Start time for a process activity ( <i>date and time</i> )
<b>Base Measure 2</b>	End time for a process activity ( <i>date and time</i> )
<b>Derived Measure 1</b>	Elapsed Time = $\text{ceiling}(\text{End Time} - \text{Start Time})$ (Units may vary based on team context, capability, cadence; hours, days, weeks, months. May also vary based on calendar time vs. work days. Results with fractional values are rounded up to the next unit.)

- ↑ **Insight**
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓
- ↑ **Impact**
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**
- ↓

(select 1 from each)



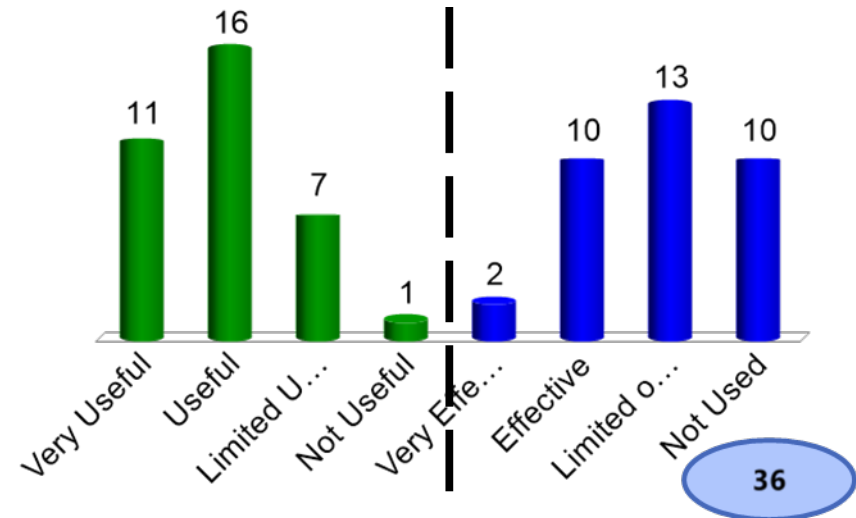
Is **Lead Time** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?



Information Need and Measure Description	
Information Need (Cycle Time)	How long does it take to complete a process activity? (team) How long does it take to develop an identified feature/capability or release? (product or enterprise)
Information Need (Lead Time)	How long does it take to get a viable product released (time to market)?
Base Measure 1	Start time for a process activity (date and time)
Base Measure 2	End time for a process activity (date and time)
Derived Measure 1	Elapsed Time = $\text{ceiling}(\text{End Time} - \text{Start Time})$ (Units may vary based on team context, capability, cadence; hours, days, weeks, months. May also vary based on calendar time vs. work days. Results with fractional values are rounded up to the next unit.)

- ↑ Insight
1. Very Useful
  2. Useful
  3. Limited Usefulness
  4. Not Useful
- ↓ Impact
5. Very Effective
  6. Effective
  7. Limited or No Effectiveness
  8. Not Used

(select 1 from each)

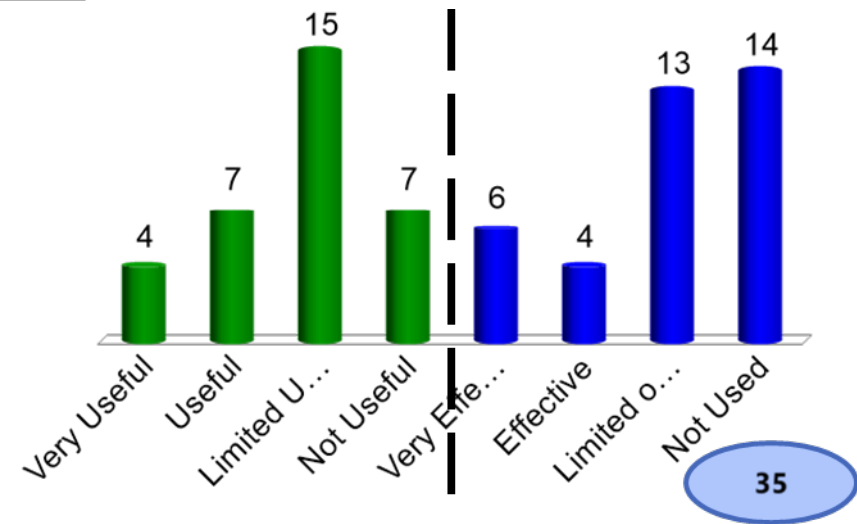
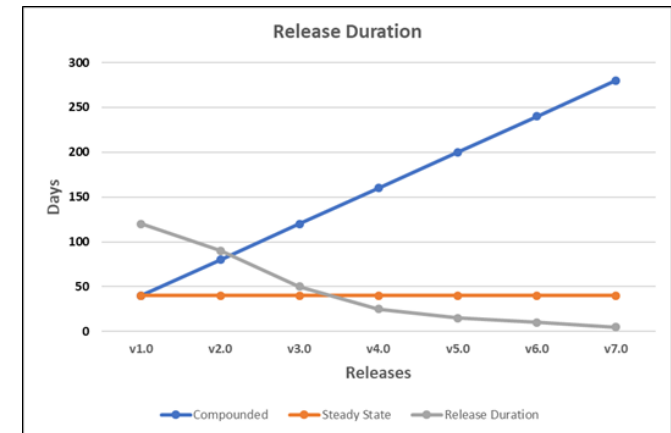


Is **Release Frequency** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?

Information Need and Measure Description	
Information Need	How long does it take to develop and release viable products? ( <i>duration/time to release new capability</i> )
	Are release candidates being produced at the cadence needed? ( <i>frequency of releases</i> )
	How long (duration/time) and how much effort/cost(?) does it take to transition candidate products to a completed product baseline release? ( <i>duration and effort/cost to deploy release candidates</i> )
Base Measure 1	Start and end dates for a product baseline release (date)
Base Measure 2	Effort hours to transition candidate products to a completed product baseline release
Derived Measure 1	Release duration = (release end date) – (release start date)
	• Time to Minimal Viable Product (MVP) = (end date for MVP <sub>1</sub> release) – (start date for MVP <sub>1</sub> release) ( <i>initial release of useful capability</i> )
	• Time to Next Viable Product (NVP <sub>n</sub> ) = (end date for NVP <sub>n</sub> release) – (end date for prior NVP <sub>n-1</sub> release) ( <i>subsequent deployments of incremental capability</i> )
	• Time to release a Candidate Product (internal) (e.g., nightly, sprint, increment, other)
	• Time to release an operational product
Derived Measure 2	Average release duration = $\sum (\text{release duration}) / (\# \text{ of releases})$ <i>Note: weighting can be used to emphasize the most recent releases.</i>
Derived Measure 3	Average release transition time = $\sum (\text{release transition time}) / (\# \text{ of releases})$

- |         |   |                                |
|---------|---|--------------------------------|
| Insight | ↑ | 1. Very Useful                 |
|         |   | 2. Useful                      |
|         |   | 3. Limited Usefulness          |
|         | ↓ | 4. Not Useful                  |
| Impact  | ↑ | 5. Very Effective              |
|         |   | 6. Effective                   |
|         |   | 7. Limited or No Effectiveness |
|         | ↓ | 8. Not Used                    |

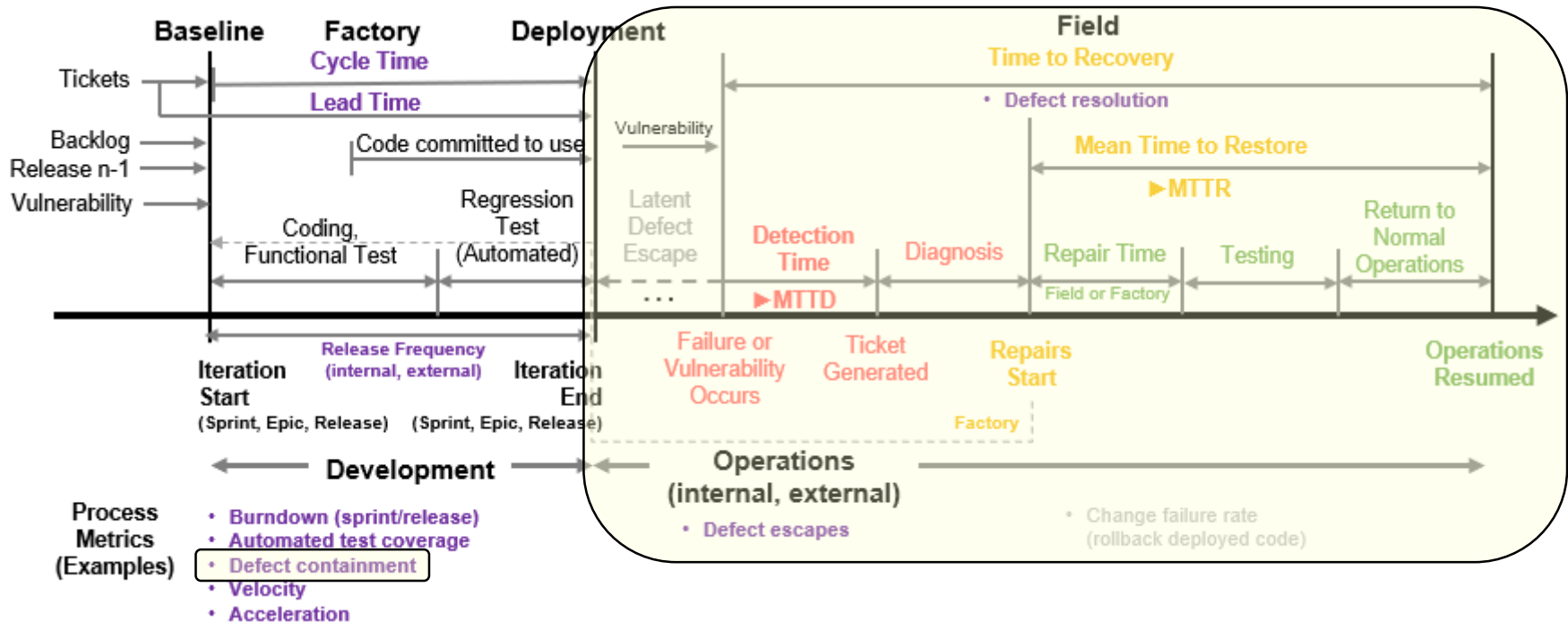
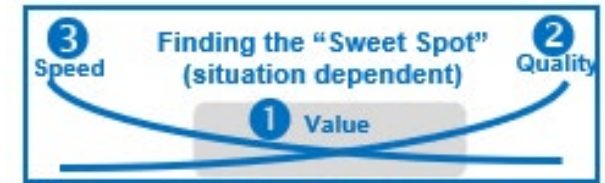
(select 1 from each)



# Quality Measures

## Concepts:

- Speed can not be optimized without also managing quality
- Quality objectives will vary according to context and domain
- Code quality is integrated into the factory pipeline processes
- Automated verification to the extent practical
- Defect measures are based primarily on escapes from development to operations (internal, external)



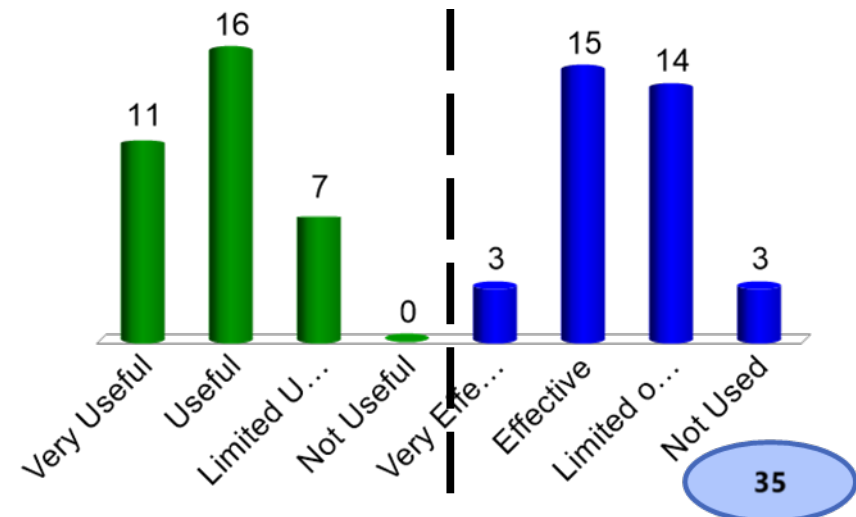
Is **Defect Containment** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Information Need and Measure Description	
Information Need	<ul style="list-style-type: none"> <li>When were defects introduced into the system (originated) and when were these defects discovered?</li> <li>How effective was the defect discovery process? What was the relative difference between Defect Originated and Defect Discovered?</li> <li>How many defects were not caught until later iterations? Why did these defects escape detection?</li> <li>How can we discover defects earlier in the product lifecycle?</li> </ul>
Base Measure 1	Number of defects originated and discovered in same iteration (integer scale)
Base Measure 2	Number of defects discovered one iteration past the iteration originated (integer scale)
Base Measure 3	Number of defects discovered two or more iterations past the iteration originated (integer scale)
Base Measure 4	Number of defects that originated as Legacy defects (e.g., previous projects)
Base Measure 5	Number of defects that have an unknown origin

Defect Containment		Defect Discovered (Iteration)							
As of 11 Jun 19									
Defects		1	2	3	4	5	6		
Unknown	0								
Legacy	0								
1	82	29	2	19	17	4	11		
2	123		27	71	6	7	12		
3	282			122	60	29	71	Blank	0 %
4	112				16	2	94	Threshold	41 %
5	7					5	2	Goal	21 %
6	54						54	Expected	38 %
Total		29	29	212	99	47	244		
	660								

- ↑ **Insight**
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓
- ↑ **Impact**
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**
- ↓

(select 1 from each)



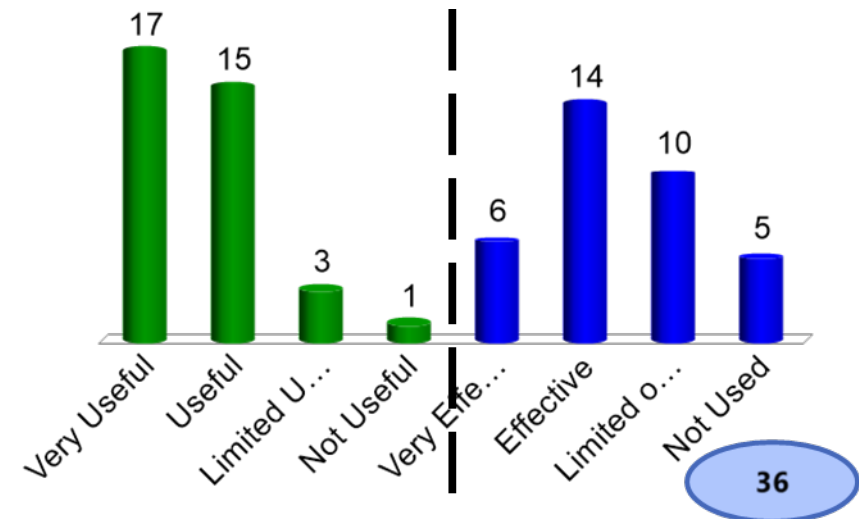
Is **Defect Escapes** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Information Need and Measure Description	
Information Need	For each release (iteration), how many defects were found in internal testing? How many defects were fielded? What is the ratio of fielded defects to all defects? How many defects were found before release to the customer? How many defects were found after release to the customer?
Base Measure 1	Internal Defects (integer scale). Defects found by the development team before release to the internal or external customer. The customer is the team that receives the delivered product. Include priorities 1-3.
Base Measure 2	Fielded Defects (integer scale). Defects found after release to the internal or external customer. Include priorities 1-3.
Derived Measure 1	Defect Escape Ratio = Fielded Defects / (Internal Defects + Fielded Defects)

Defect Escapes				
Release	Defects		Total Defects	Ratio
	Internal	Delivered		
Release 1.0	48	11	59	19%
Release 1.1	55	6	61	10%
Release 1.2	31	4	35	11%
Release 2.0	64	6	70	9%
Release 2.1	55	8	63	13%
Release 2.2	48	4	52	8%
Release 2.3	31	3	34	9%
Release 3.0	20	1	21	5%

- ↑ **1. Very Useful**
- ↑ **2. Useful**
- ↑ **3. Limited Usefulness**
- ↓ **4. Not Useful**
- ↑ **5. Very Effective**
- ↑ **6. Effective**
- ↑ **7. Limited or No Effectiveness**
- ↓ **8. Not Used**

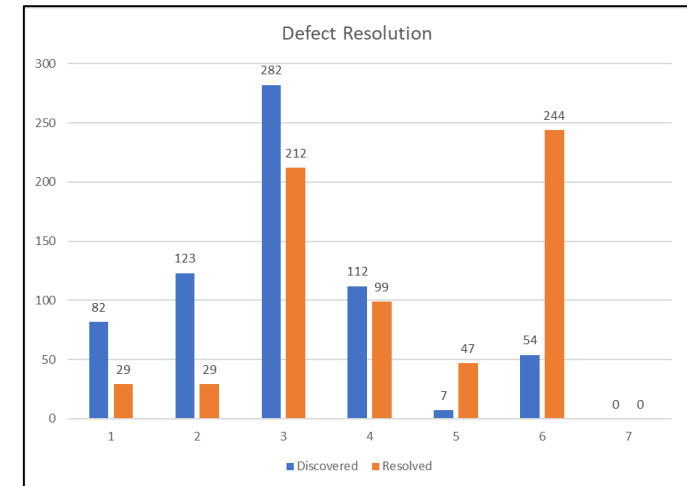
(select 1 from each)





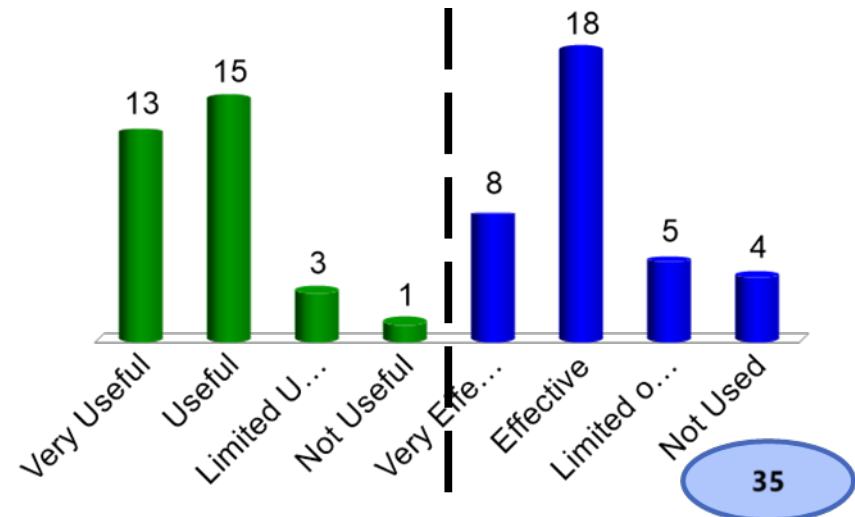
Is **Defect Resolution** a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Information Need and Measure Description	
Information Need	<ul style="list-style-type: none"> <li>When were discovered defects resolved? How effective was the defect resolution process?</li> <li>How can we resolve defects earlier in the product lifecycle?</li> </ul>
Base Measure 1	Number of defects discovered per iteration (integer scale)
Base Measure 2	Number of defects resolved per iteration (integer scale)
Derived Measure 1	Total number of defects discovered in each iteration
Derived Measure 2	Total number of defects resolved in each iteration (integer scale)
Derived Measure 3	Expected Percentage (Resolved) = the number of defects that are resolved in the same iteration they were discovered in (Defect Resolved is the same as Defect Discovered) / the total number of defects
Derived Measure 4	Goal Percentage (Resolved) = the number of defects that are resolved 1 iteration after being discovered / the total number of defects
Derived Measure 5	Threshold Percentage (Removed) = the number of defects that are resolved more than 1 iteration after being discovered / the total number of defects
Derived Measure 6	Defect Age (for active/non-closed defects) = Current Increment – Increment Discovered
Derived Measure 7	Defect Cycle Time (for closed defects) = Increment Resolved – Increment Discovered

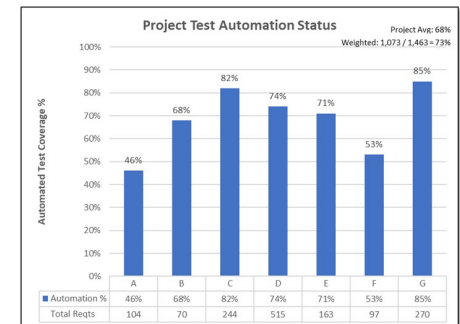
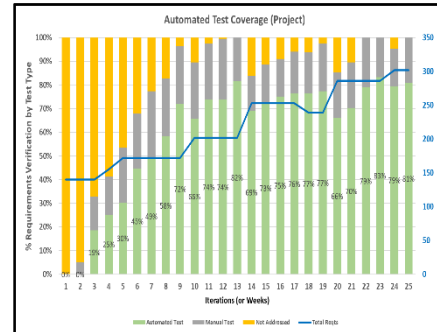


- ↑ Insight
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓ Impact
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**

(select 1 from each)



Is **Automated Test Coverage** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?



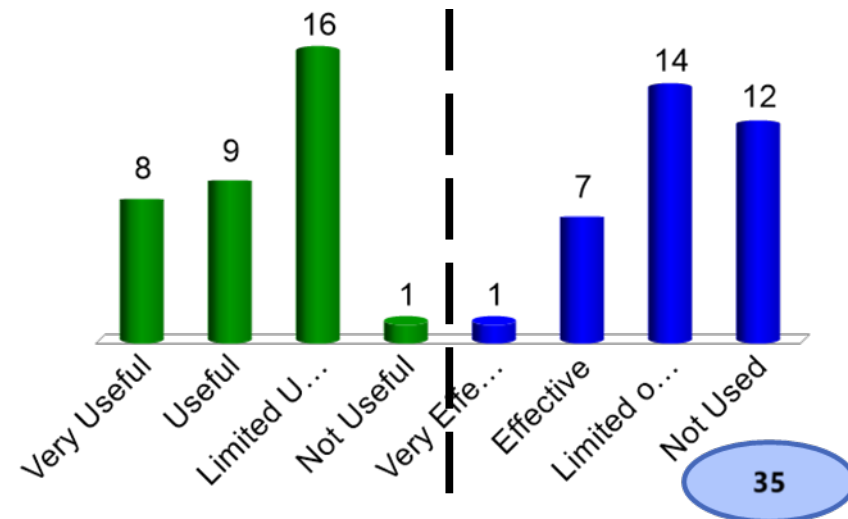
Information Need and Measure Description	
Information Need	How much of the testing is automated? Based on automated test results, what is the quality of the product baseline?
Base Measure 1	Requirements coverage from automated testing (counts, %) Requirements tested by automated test Requirements tested
Base Measure 2	Automated test coverage of code constructs.
Derived Measure 1	% of automated testing coverage for functional requirements
Derived Measure 2	% of automated testing coverage for code constructs (e.g., classes, conditionals, files, lines, packages)

Project

Enterprise

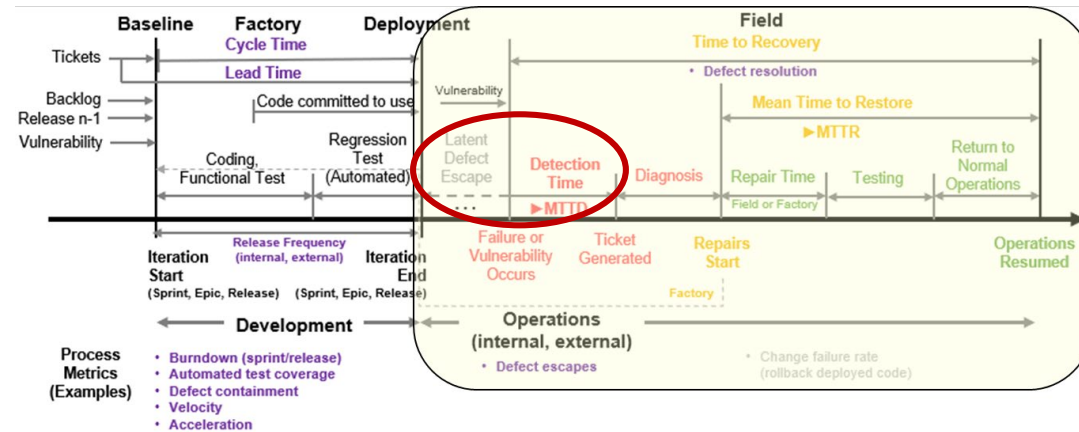
- ↑ **Insight**
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓ **Impact**
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**

(select 1 from each)



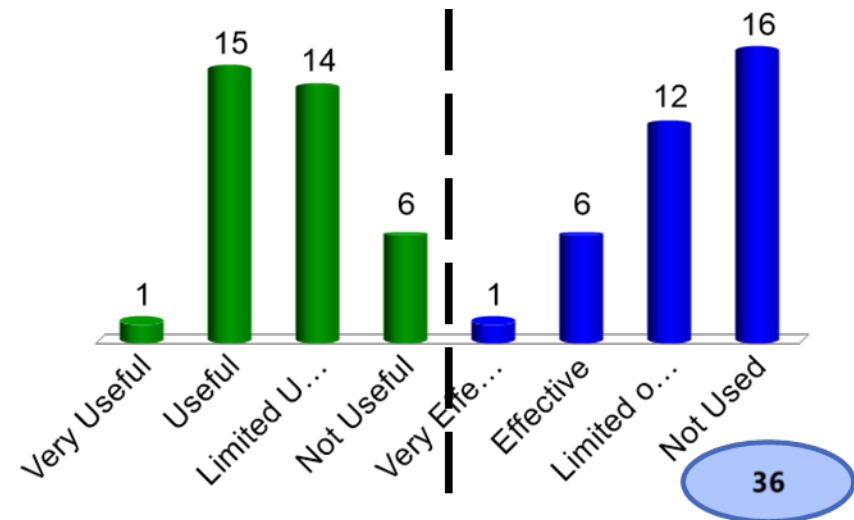


Is **Mean Time to Detect (MTTD)** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?

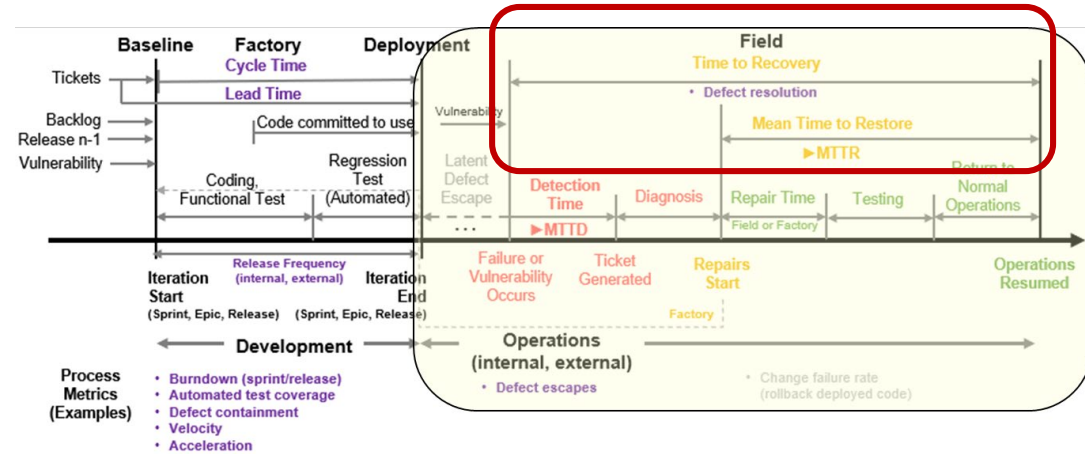


- ↑ **Insight**
1. Very Useful
  2. Useful
  3. Limited Usefulness
  4. Not Useful
- ↓
- ↑ **Impact**
5. Very Effective
  6. Effective
  7. Limited or No Effectiveness
  8. Not Used
- ↓

(select 1 from each)

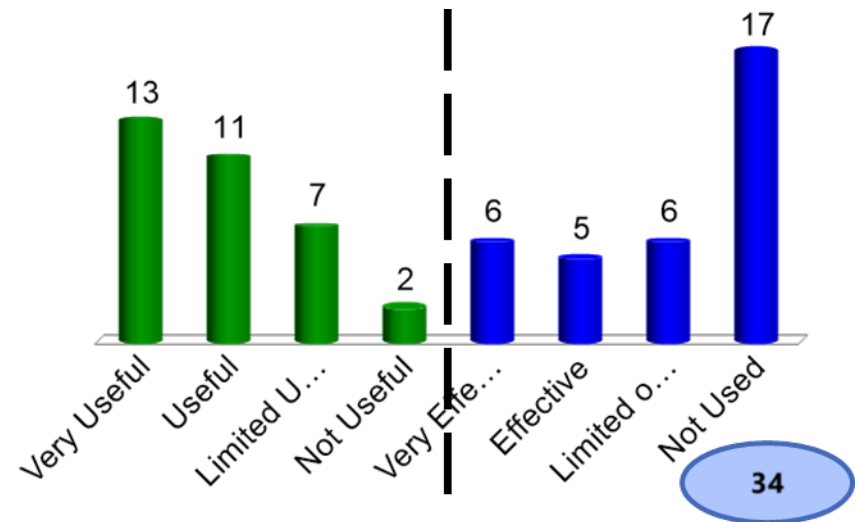


Is **Mean Time to Restore (or Recover) (MTTR)** a **useful** measure, and how **effectively** is it **used** to provide insight and impactful action in **your** organization?



- ↑ **Insight**
1. **Very Useful**
  2. **Useful**
  3. **Limited Usefulness**
  4. **Not Useful**
- ↓
- ↑ **Impact**
5. **Very Effective**
  6. **Effective**
  7. **Limited or No Effectiveness**
  8. **Not Used**
- ↓

(select 1 from each)



***Thank you for your feedback on the proposed CID measures!***

***Next steps:***

- ***Finalize info needs and measures for ICM table and CID measurement framework***
- ***Refine measurement specs***
- ***Publish consensus industry framework for community review***
- ***Use measurement framework to inform DoD acquisition policy and guidance***

**Volunteers are needed to help bring the project to completion**



**WE WANT YOU!**

[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)

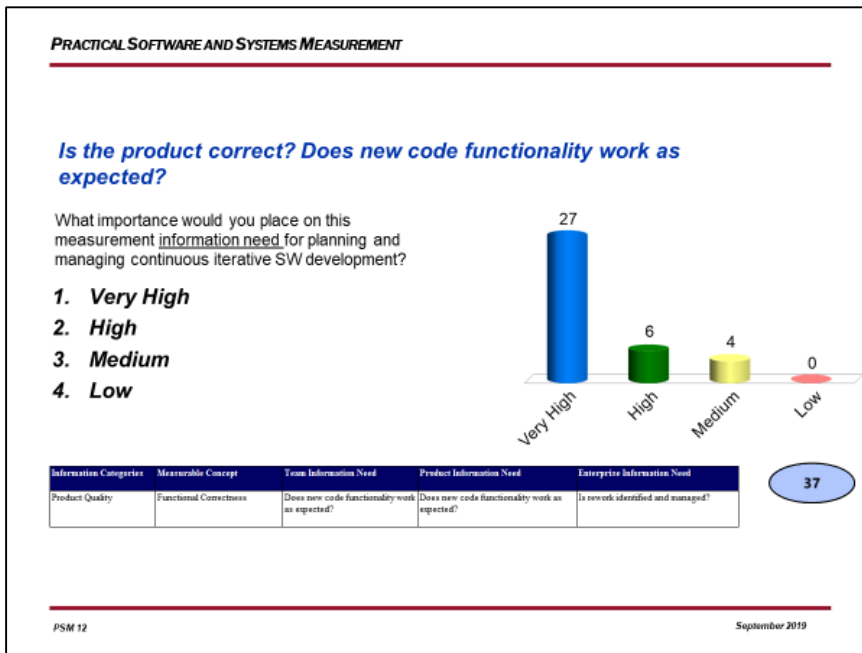
## **Outbrief:**

# **Part III: Preliminary Analysis and Results**

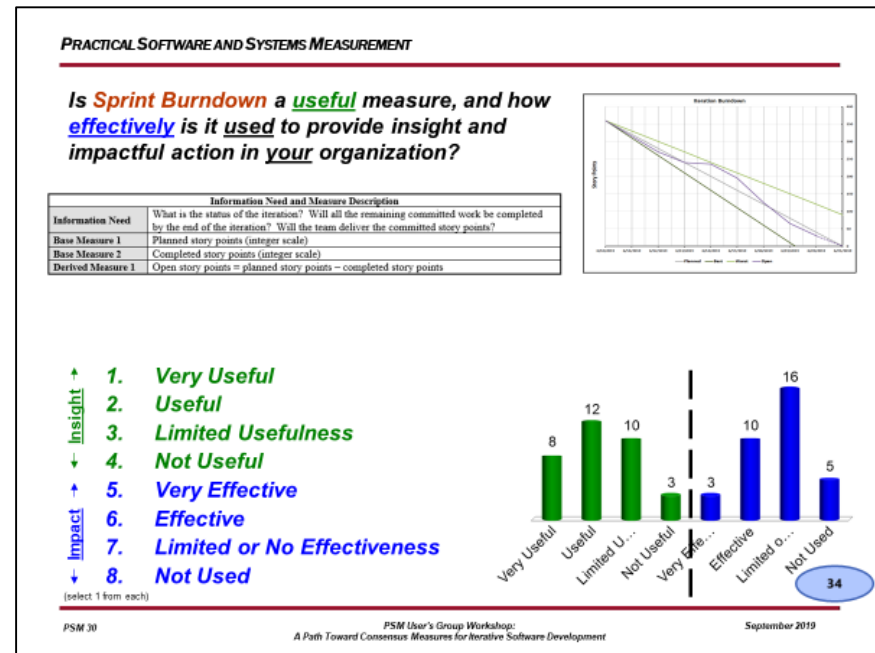
# Workshop Outbrief....

- Interactive evaluation of draft PSM Agile framework

## Information Needs

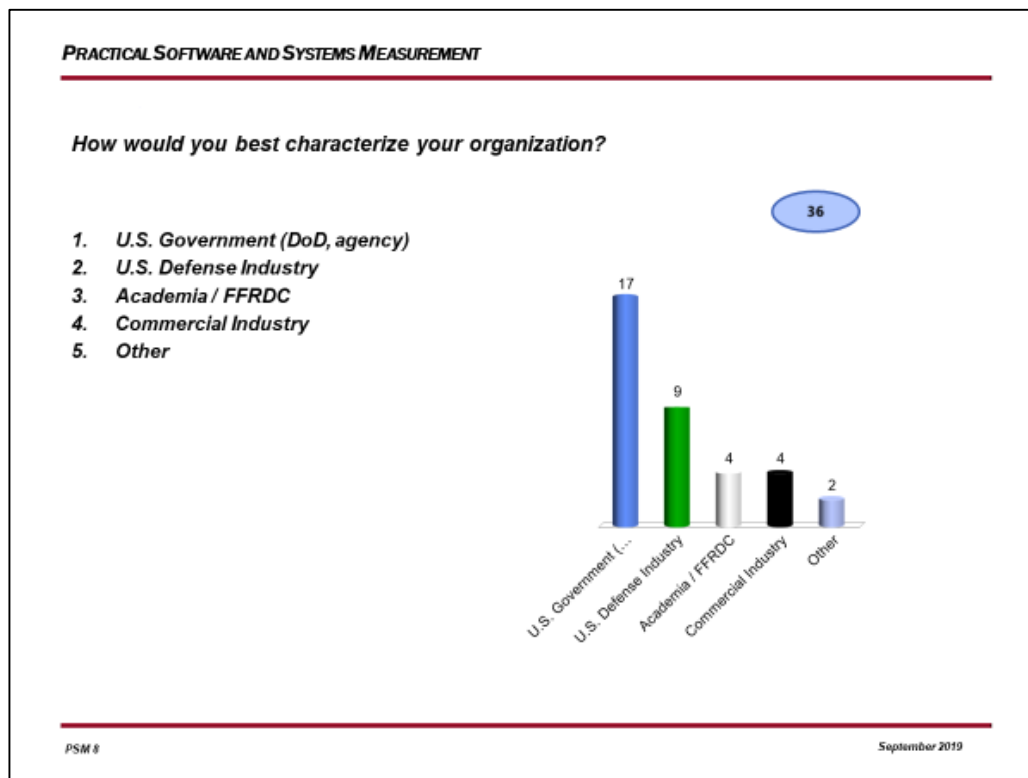


## Potential Measures



# Workshop Participants

- **Presentation survey: 33-36 participants**
- **Workshop participants TBD – from attendance list**



## Evaluation of PSM Draft Agile Framework: Info Needs

Evaluation and Ranking of (Draft) PSM Agile Framework - Information Needs Interactive Survey of SMEs at PSM User's Group, Sep 2019 (VH=1; H=2; M=3; L=4)														
Scoring: 4: Least favorable ... 1: Most favorable	Work Unit Progress		Work Backlog	Functional Correctness			Reliability	Process Efficiency	Process Efficiency - Speed				Process Effectiveness	
	Delivery vs. Committed?	Provide Expected Functionality?	Technical Debt in Backlog?	Is Product Correct?	Change failure rate?	How Many Defect Escapes?	Fielded Code Quality?	Operational Reliability?	Field Bug Fix Speed?	Team Performance as Expected?	How Long to Deploy a Capability?	Frequency of Releases?	Time to Release a Viable Product?	How Much Testing is Automated?
1	16	12	2	27	10	4	19	10	2	2	7	1	5	3
2	11	22	13	6	8	16	14	18	12	15	12	7	10	14
3	6	1	17	0	17	16	1	8	18	16	17	12	18	13
4	2	0	3	0	0	1	1	0	3	2	0	14	2	6
Rank	5	3	11	1	7	6	2	4	12	9	8	15	14	13
Sample Count	35	35	35	33	35	37	35	36	35	35	36	34	35	36
Mean	1.83	1.69	2.60	1.18	2.20	2.38	1.54	1.94	2.63	2.51	2.28	3.15	2.49	2.61
Median	2.00	2.00	3.00	1.00	2.00	2.00	1.00	2.00	3.00	3.00	2.00	3.00	3.00	3.00
Std Dev.S	0.92	0.53	0.74	0.39	0.87	0.72	0.70	0.71	0.73	0.70	0.78	0.86	0.82	0.87
Var.S	0.85	0.28	0.54	0.15	0.75	0.52	0.49	0.51	0.53	0.49	0.61	0.74	0.67	0.76

**Top ranking information needs:** based on PSM User Group input (34-36 participants)

- **Is the product correct? Does new code functionality work as expected?**
- **What is the quality of code deployed to the field?**
- **Do delivered products provide the expected functionality to users when needed?**
- **What is the reliability and availability of operational service capabilities?**
- **Are new story points, features, or capabilities being delivered as committed?**

## ***Workshop participants provided suggestions for refinement of information needs***

#	Feedback / Comments	Team Action
1	Terminology: avoid typing or implying Scrum (iterations, etc.). Consider other approaches (e.g., Kanban).	
2	Terminology: keep MVCR, MVP/NVP?	Review latest OSD policy.
3	Add to principles = relationships among concepts, fact-based decision making.	
4	Ideally, measures should aggregate across teams/projects.	
5	Are sprints too level to measure? Is there enough measurable value in story points? Suggestion (Will): consider “ideal days” (8 hrs) as basis for story points.	
6	Value depends on stakeholder perspective – not size, cost, etc.	
7	What are the assumptions and constraints acquirers are trying to measure against? How do we measure value to the warfighter?	
8	Proxies for the amount of work always end up aligning with hours, and acceptance criteria.	
9	Size measures and traditional measures may not align well with agile.	
10	Customer satisfaction and acceptance should include measures (e.g, scope, value, time bound, criteria met to agreed-upon level of maturity	
11	Debt ≠ backlog. Debt = things we must do, must be specific. Mission debt.	Info need: technical debt
12	Rework should also apply to teams.	Add rework to ICM team info need
13	Expectations must be explicit – not implied	?
14	New code break functionality: enterprise info needs = rework, COPQ	Add enterprise info needs for code breaking prior functionality
15	Defect escapes > add to Team Info Need. Change “Iteration” to “Release”. Focus on handoffs?	Defect escapes: add to Team info need
16	Consider adding: “first pass fix rate” (Recidivism?)	
17	“Bug reports” > restoration of service (and priority). Fix time vs. Deployment lead time.	
18	“Are teams performing as <del>productively</del> predictably as expected”	
19	Story points (and derivatives) are useful only at the team level.	
20	Release frequency: importance depends on domain (e.g., Amazon, Netflix). Also depends on Product Owner inputs.	
21	Maintaining automated testes should consider refactoring as a new necessary skill set.	Consider adding to Automated Test Coverage info need
22	Does testing address high risk areas?	Ensure addressed in measurement spec
23	Size: “how big” > how can we quantify the scope for estimating?	
24	EVM concepts can be applied for agile	(Side discussion – N/A for ICM or specs)



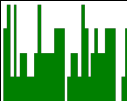
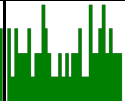

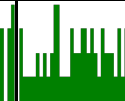
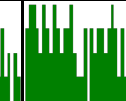
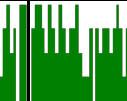
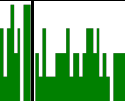
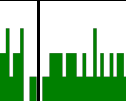
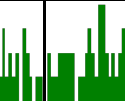
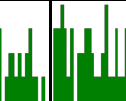
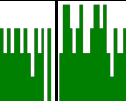
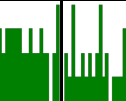



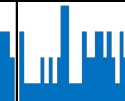
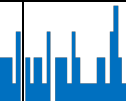
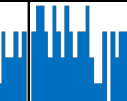
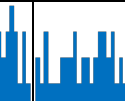
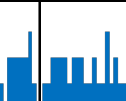
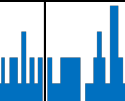
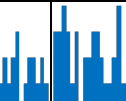
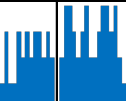
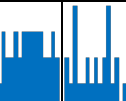
## Evaluation of PSM Draft Agile Framework: Potential Measures

Evaluation and Ranking of (Draft) PSM Agile Framework Measures Interactive Survey of SMEs at PSM User's Group, Sep 2019 (Usefulness: 1-4; Effectiveness in Use: 1-4)																								
Measurable Concept	Work Unit Progress		Process Efficiency - Speed										Funtional Correctness						Process Effectiveness		Supportability - Maintainability - Reliability - Dependability - Process Effectiveness			
Measure	Sprint Burndown		Velocity		Acceleration		Cycle Time		Lead Time		Release Frequency		Defect Containment		Defect Escapes		Defect Resolution		Automated Test Coverage		MTTD		MTTR	
	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective	Useful	Effective
4: Least favorable																								
1: Most favorable																								
4	3	5	2	6	7	13	2	10	1	10	6	12	0	3	0	5	1	4	1	11	6	15	2	16
3	10	16	7	12	19	18	8	8	7	12	15	12	7	13	3	10	3	4	15	14	13	12	7	6
2	12	9	18	14	8	2	12	16	16	10	7	4	16	15	15	14	15	18	9	7	15	6	10	5
1	8	3	8	2	1	0	14	1	11	2	4	6	11	3	17	5	13	7	8	1	1	1	13	6
Rank: Usefulness	8		6		10		4		3		12		2		1		5		9		11		7	
Rank: Effectiveness		5		4		12		6		7		8		2		3		1		10		11		9
Sample Count	34	33	36	34	35	33	36	35	35	34	35	34	34	34	35	34	33	33	34	33	36	34	33	33
Mean	2.35	2.70	2.19	2.65	2.91	3.33	1.94	2.77	1.94	2.88	3.14	2.88	1.88	2.47	1.60	2.44	1.91	2.15	2.44	3.06	2.83	3.21	2.12	2.97
Median	2.00	3.00	2.00	3.00	3.00	3.00	2.00	3.00	2.00	3.00	3.00	3.00	2.00	2.00	2.00	2.00	2.00	2.00	2.50	3.00	3.00	3.00	2.00	3.00
Std Dev.S	1.12	0.85	1.04	0.85	0.74	0.60	0.92	0.91	0.80	0.91	1.67	1.09	0.73	0.79	0.65	0.93	1.18	0.91	1.31	0.83	1.18	0.84	1.41	1.19
Var.S	1.27	0.72	1.08	0.72	0.55	0.35	0.85	0.83	0.64	0.83	2.77	1.20	0.53	0.62	0.42	0.86	1.40	0.82	1.71	0.68	1.40	0.71	1.98	1.41

**Top ranking agile measures:** based on PSM User Group inputs (33-36 participants)

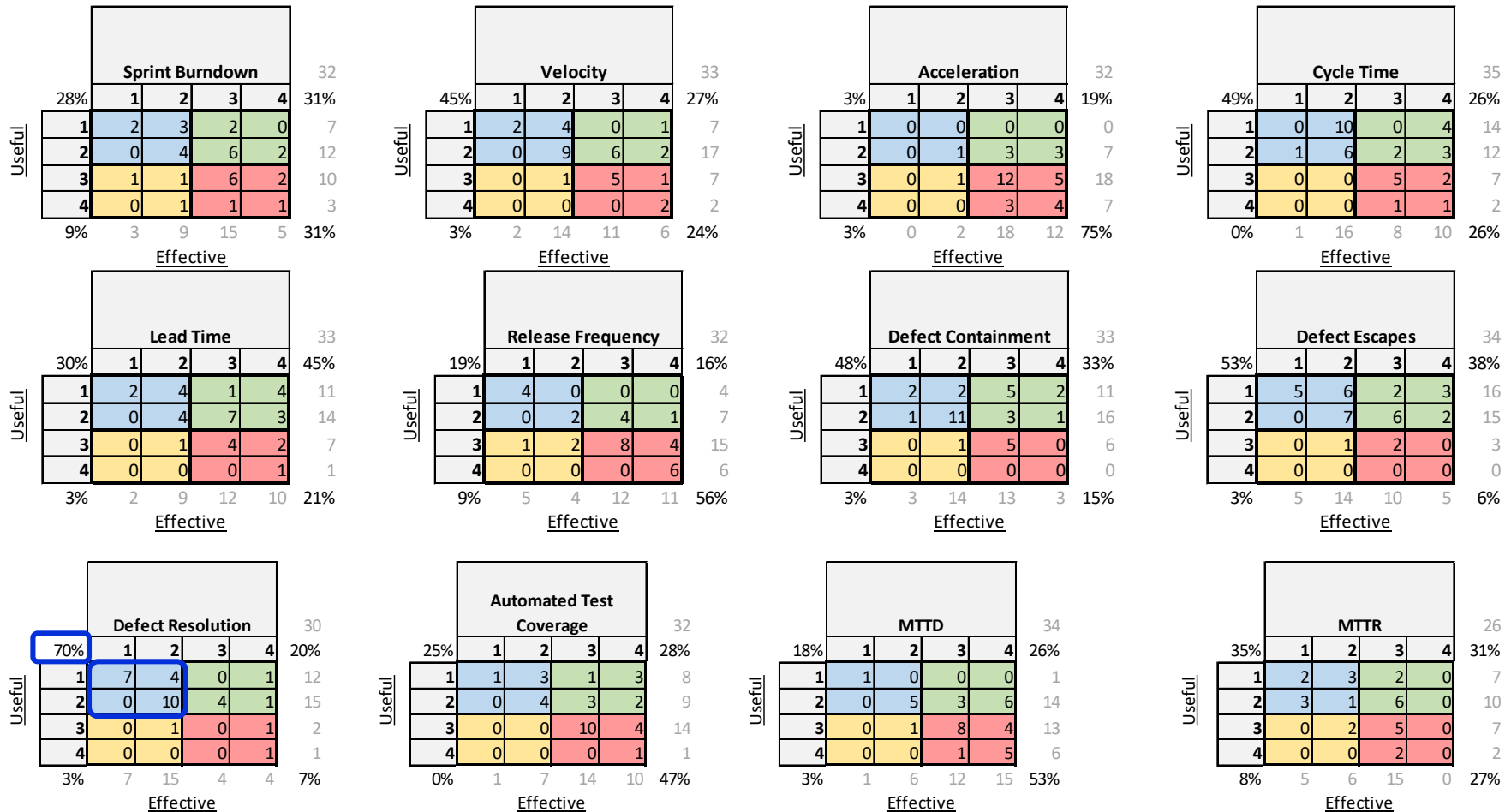
- **Usefulness**
  - Defect Escapes; Defect Containment; Lead Time; Cycle Time; Defect Resolution
- **Effectiveness**
  - Defect Resolution; Defect Containment; Defect Escapes;

## Evaluation of PSM Draft Agile Framework: Potential Measures

Evaluation and Ranking of (Draft) PSM Agile Framework Measures Interactive Survey of SMEs at PSM User's Group, Sep 2019 (Usefulness: 1-4; Effectiveness in Use: 1-4)												
Measurable Concept	Work Unit Progress	Process Efficiency - Speed					Functional Correctness			Process Effectiveness	Supportability - Maintainability - Reliability - Dependability - Process Effectiveness	
Measure	Sprint Burndown	Velocity	Acceleration	Cycle Time	Lead Time	Release Frequency	Defect Containment	Defect Escapes	Defect Resolution	Automated Test Coverage	MTTD	MTTR
<b>Usefulness</b>												
Least favorable												
Most favorable												
Rank	8	6	10	4	3	12	2	1	5	9	11	7
Sample Count	34	36	35	36	35	35	34	35	33	34	36	33
Mean	2.35	2.19	2.91	1.94	1.94	3.14	1.88	1.60	1.91	2.44	2.83	2.12
Median	2.00	2.00	3.00	2.00	2.00	3.00	2.00	2.00	2.00	2.50	3.00	2.00
Std Dev.S	1.12	1.04	0.74	0.92	0.80	1.67	0.73	0.65	1.18	1.31	1.18	1.41
Var.S	1.27	1.08	0.55	0.85	0.64	2.77	0.53	0.42	1.40	1.71	1.40	1.98
<b>Effectiveness</b>												
Least favorable												
Most favorable												
Rank	5	4	12	6	7	8	2	3	1	10	11	9
Sample Count	33	34	33	35	34	34	34	34	33	33	34	33
Mean	2.70	2.65	3.33	2.77	2.88	2.88	2.47	2.44	2.15	3.06	3.21	2.97
Median	3.00	3.00	3.00	3.00	3.00	3.00	2.00	2.00	2.00	3.00	3.00	3.00
Std Dev.S	0.85	0.85	0.60	0.91	0.91	1.09	0.79	0.93	0.91	0.83	0.84	1.19
Var.S	0.72	0.72	0.35	0.83	0.83	1.20	0.62	0.86	0.82	0.68	0.71	1.41

- Comparison of Usefulness (stronger) vs. Effectiveness (weaker) scores suggests good measures are not yet fully implemented or leveraged fully in practice

*The best measures for software acquisition may be those that rate highest in both Usefulness (insight) and Effectiveness (impact/adoption)*



***Thank you for your feedback on the proposed CID measures!***

***Next steps:***

- ***Finalize info needs and measures for ICM table and CID measurement framework***
- ***Refine measurement specs***
- ***Publish consensus industry framework for community review***
- ***Use measurement framework to inform DoD acquisition policy and guidance***

**Volunteers are needed to help bring the project to completion**



**WE WANT YOU!**

[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)