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 CCDC-AC
Geoff Draper – L3Harris Technologies

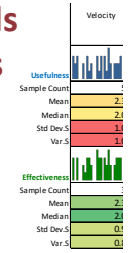
Workshop Background

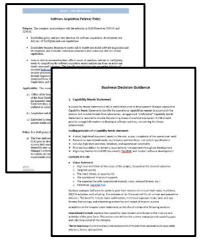
DSB SW

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 868 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.

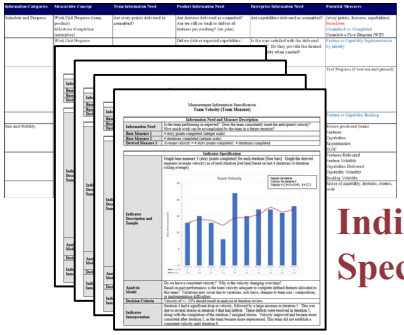
NDIA, INCOSE and PSM support the DSB and DIB concepts and the opportunities they offer to DoD and the defense industry. In 2018 the NDIA Systems Engineering Division commissioned a working group to study the implementation of the DIB and SWAP recommendations. NDIA offers the working group's recommendations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

Executive Summary Brief (pdf)

Executive Summary Brief (ppt with links) view in presentation mode

Full Study Brief

Transmittal Letter to Dr. Boleng, OUSD (A&S)



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 II: Evaluating Candidate Measures for Continuous Iterative Development

PSM **Draft**
Burndown (sprint/release)
Velocity
Acceleration
Cycle time
Lead time
Release frequency
Defect containment
Defect escapes
Defect resolution
Automated test coverage
Core PSM framework: • Cost (est. vs. actual) • Schedule (est. vs. actual) • Staffing • …etc.

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) Burndown Committed vs. Completed 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?	Feature or Capability Implementation by priority
	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?		Feature or Capability Backlog
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 Capabilites 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 desirements?	How volatile are capabilities or requirements? What is the ability to accommodate changes in customer desirements?	Features Delivered Feature Volatility Capabilites 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 Rework hours Rework stories Change Failure Rate or Defect Density
	Functional Correctness		How many defects escape the increment?		Defects Found in Pipeline (saves)
	Functional Correctness		What is the quality of code deployed to the field?	What is the quality of code deployed to the field?	Defect Escapes to field Defect Escape Ratio
	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) MTTR (Repair or Restore) 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)

Evaluations of Measures for Continuous Iterative Development

Two separate evaluations are desired:

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

Example:

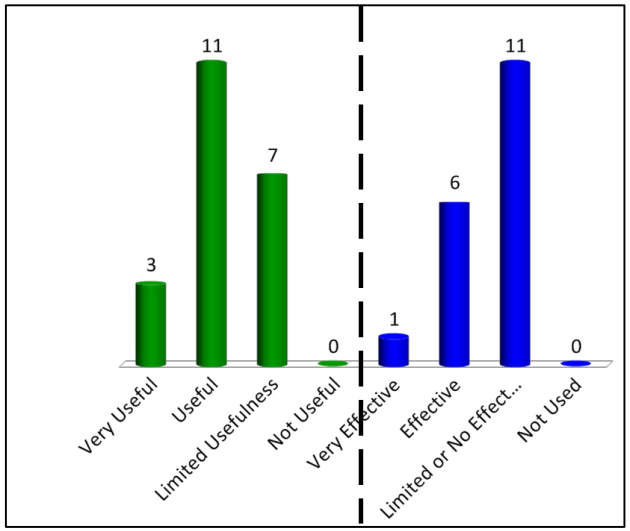
Is [measure name] a useful measure, and how effectively is it used to provide insight and impactful action in your organization?

Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

Select 1 from here

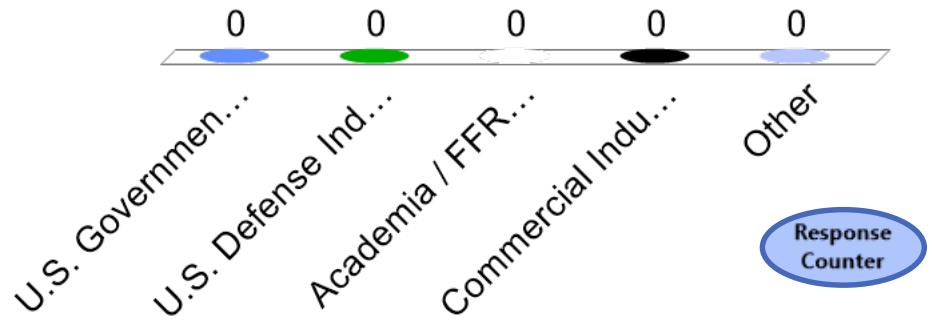
Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

and 1 from here

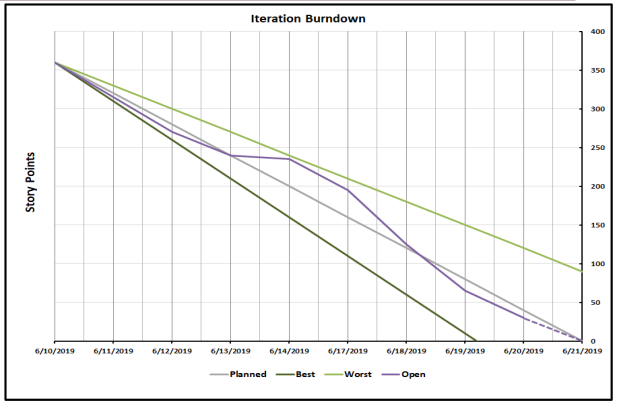


How would you best characterize your organization?

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



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

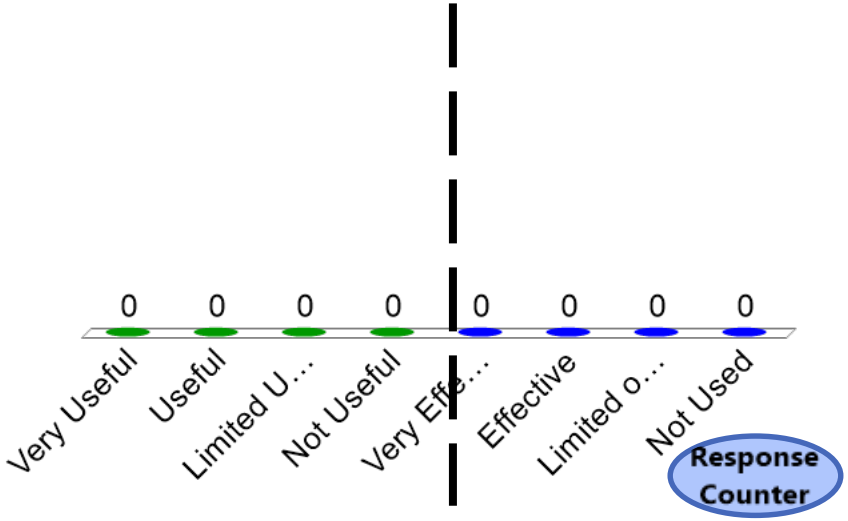


Information Need and Measure Description	
Information Need	What is the status of the iteration? Will all the remaining committed work be completed by the end of the iteration? Will the team deliver the committed story points?
Base Measure 1	Planned story points (integer scale)
Base Measure 2	Completed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

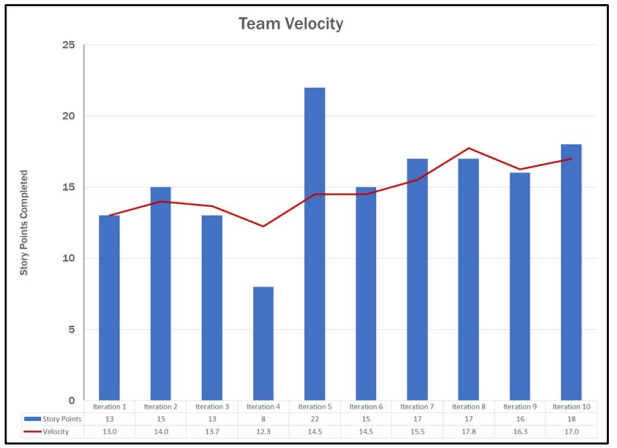
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

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?

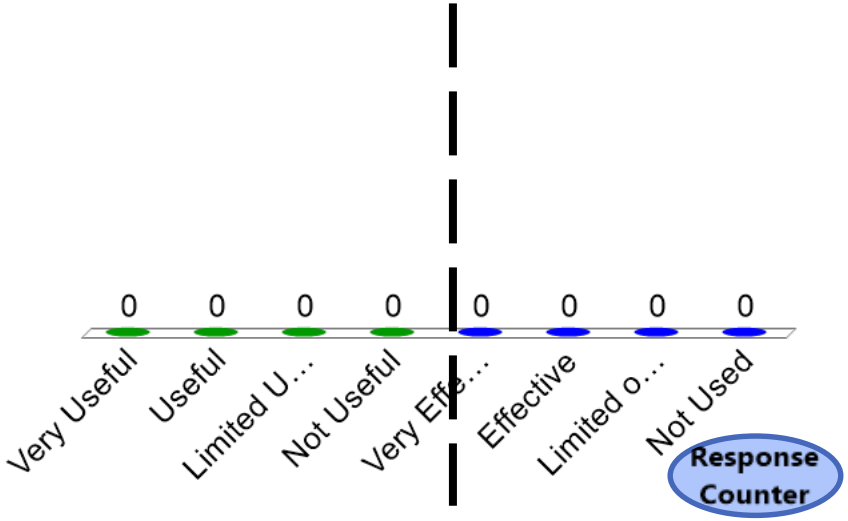


Information Need and Measure Description	
Information Need	Is the team performing as expected? Does the team consistently meet the anticipated velocity? How much work can be accomplished by the team in a future iteration?
Base Measure 1	# story points completed (integer scale)
Base Measure 2	# iterations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

Insight
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

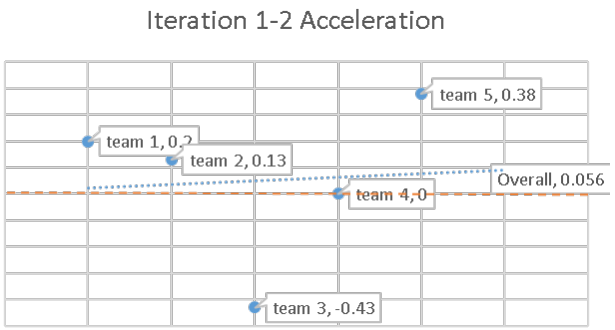
(select 1 from each)



Is 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)

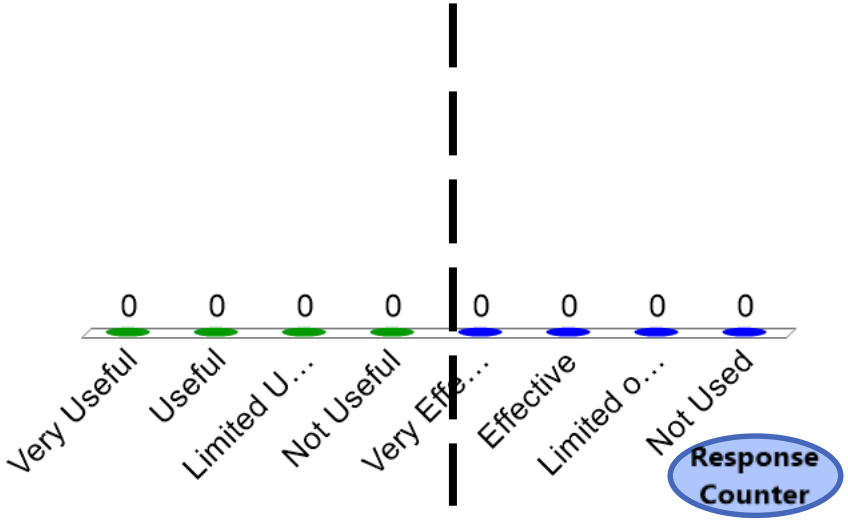


Information Need and Measure Description	
Information Need	Is the team's productivity increasing, decreasing or holding steady?
Base Measure 1	# story points completed (velocity) this increment (integer scale)
Base Measure 2	# story points completed (velocity) in previous comparison increment (integer 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

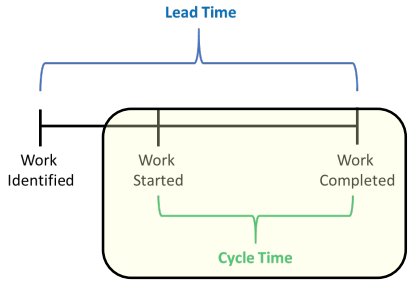
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

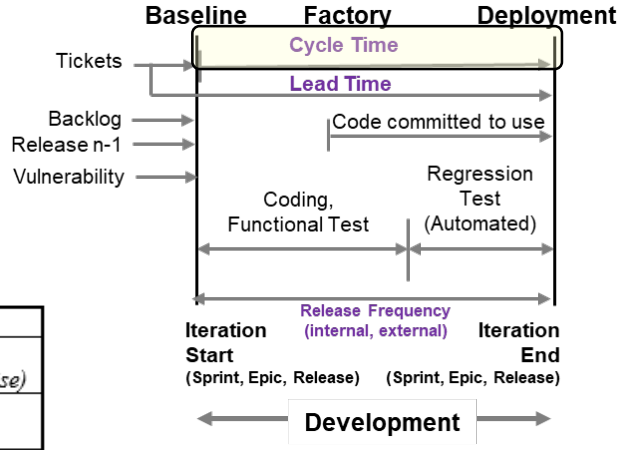
Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

(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?



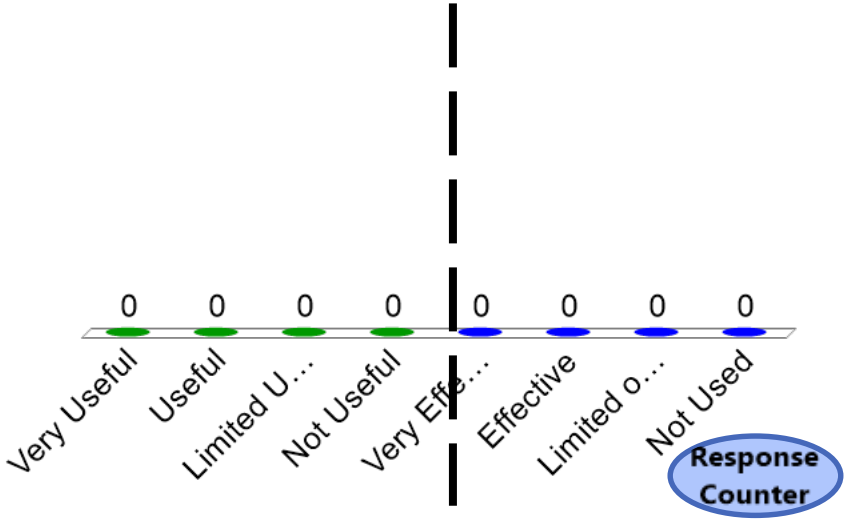


Information Need and Measure Description	
Information Need (Cycle Time)	How long does it take to complete a process activity? *(team)* How long does it take to develop an identified feature/capability or release? *(product or enterprise)*
Information Need (Lead Time)	How long does it take to get a viable 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 = ceiling(End Time – 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.)*

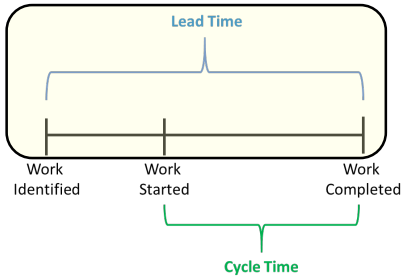
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

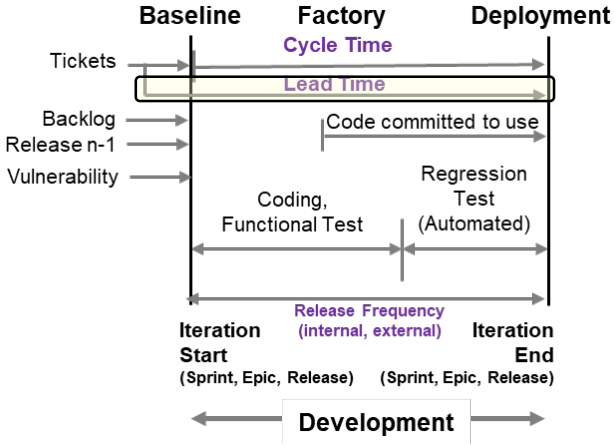
Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

(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?



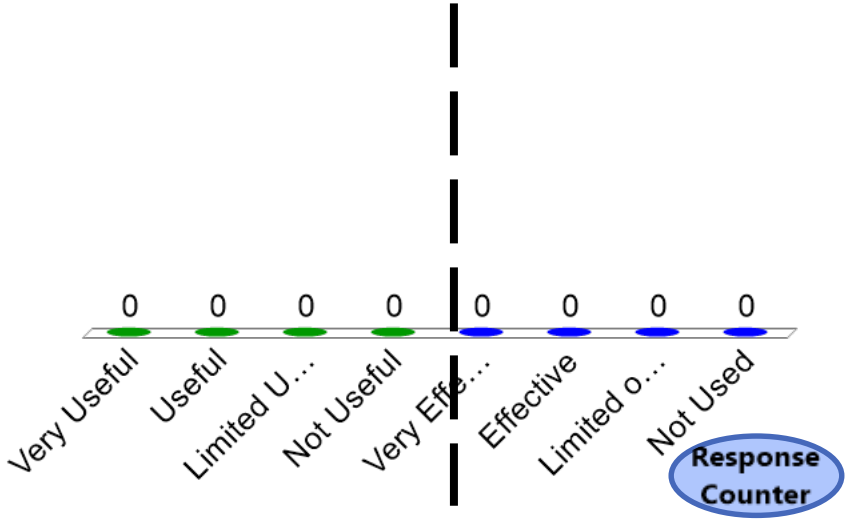


Information Need and Measure Description	
Information Need (Cycle Time)	How long does it take to complete a process activity? *(team)* How long does it take to develop an identified feature/capability or release? *(product or enterprise)*
Information Need (Lead Time)	How long does it take to get a viable 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 = ceiling(End Time – 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

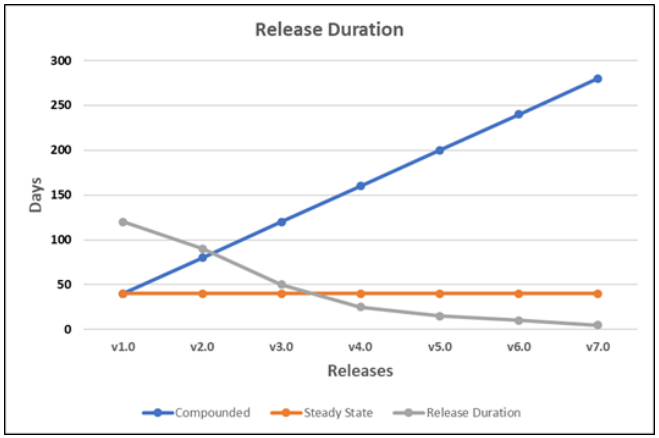
Impact
- 5. Very Effective
- 6. Effective
- 7. Limited or No Effectiveness
- 8. Not Used

(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?

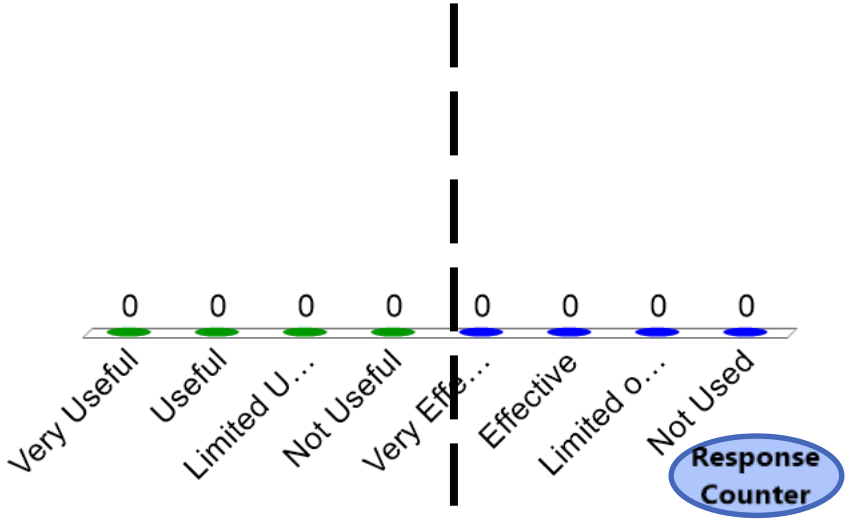
Information Need and Measure Description	
Information Need	How long does it take to develop and release viable products? *(duration/time to release new capability)* Are release candidates being produced at the cadence needed? *(frequency of releases)* How long (duration/time) and how much effort/cost?) does it take to transition candidate products to a completed product baseline release? *(duration and effort/cost to deploy release candidates)*
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_1 release) – (start date for MVP_1 release) *(initial release of useful capability)* • Time to Next Viable Product (NVP_n) = (end date for NVP_n release) – (end date for prior NVP_{n-1} release) *(subsequent deployments of incremental capability)* • 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 = ∑ (release duration) / (# of releases) *Note: weighting can be used to emphasize the most recent releases.*
Derived Measure 3	Average release transition time = ∑ (release transition time) / (# of releases)



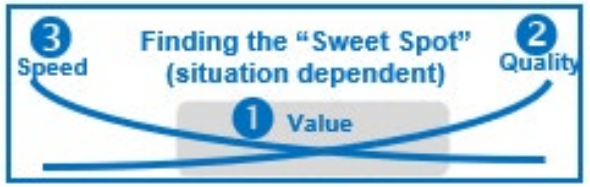
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

(select 1 from each)

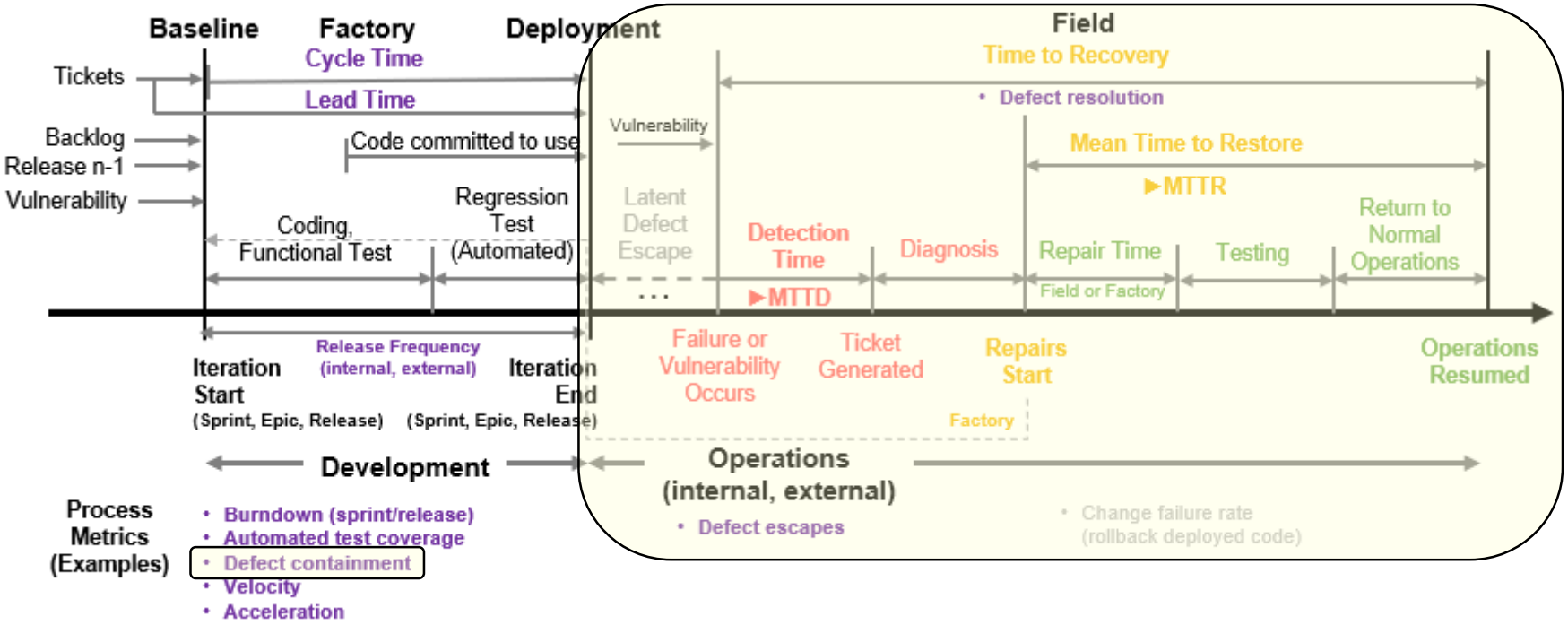


Quality Measures



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	• When were defects introduced into the system (originated) and when were these defects discovered? • How effective was the defect discovery process? What was the relative difference between Defect Originated and Defect Discovered? • How many defects were not caught until later iterations? Why did these defects escape detection? • How can we discover defects earlier in the product lifecycle?
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

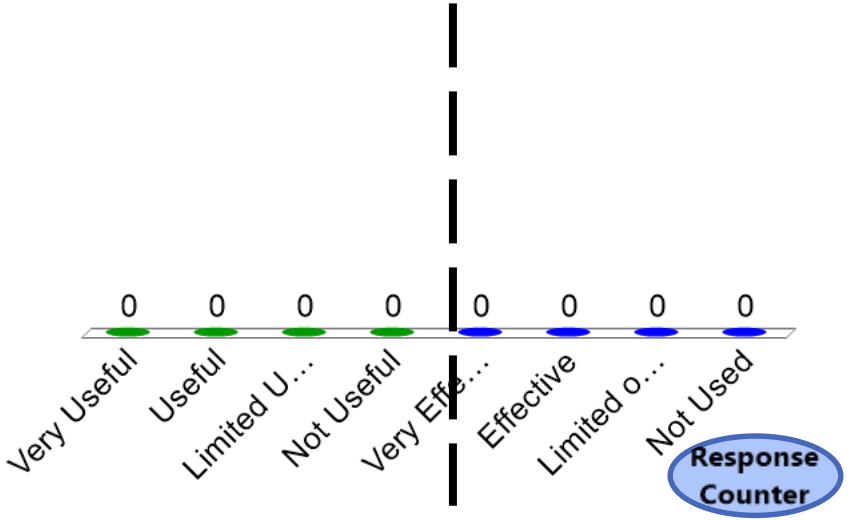
Defect Containment
As of 11 Jun 19

Defects			Defect Discovered (Iteration)							
			1	2	3	4	5	6		
Defect Originated (Iteration)	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)

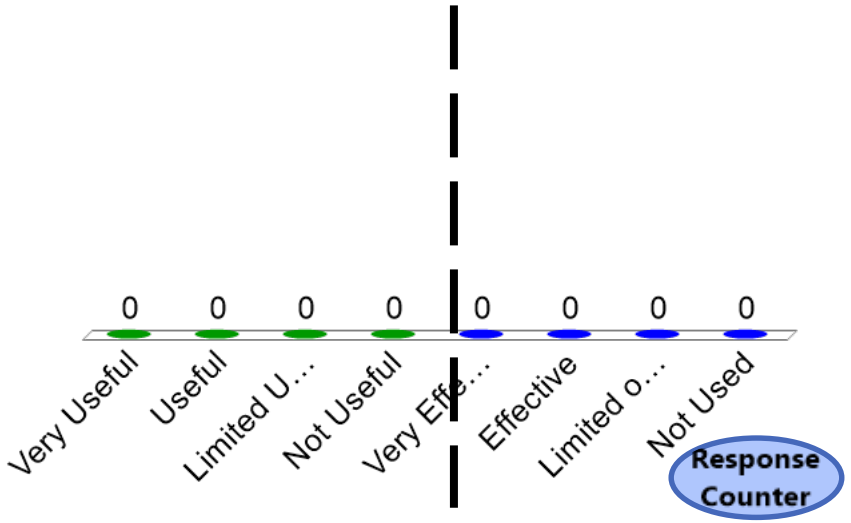
Defect Escapes

Release	Defects: Internal	Defects: Delivered	Defects: Total Defects	Ratio
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%

Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

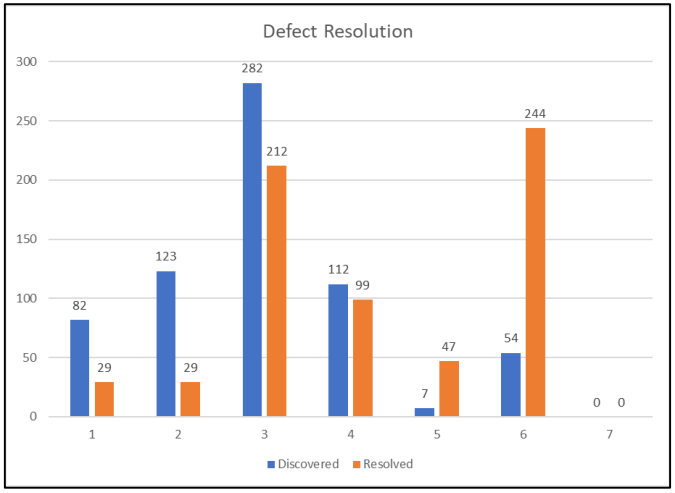
Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

(select 1 from each)



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

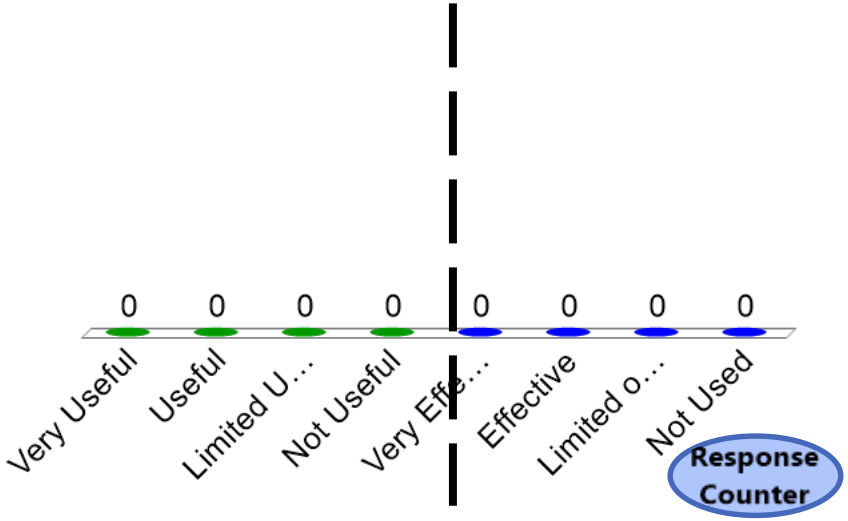
Information Need and Measure Description	
Information Need	• When were discovered defects resolved? How effective was the defect resolution process? • How can we resolve defects earlier in the product lifecycle?
Base Measure 1	Number of defects discovered per iteration (integer scale)
Base Measure 2	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



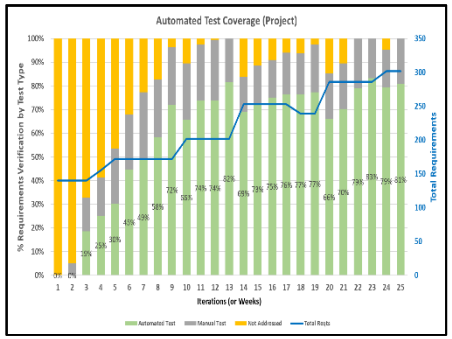
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

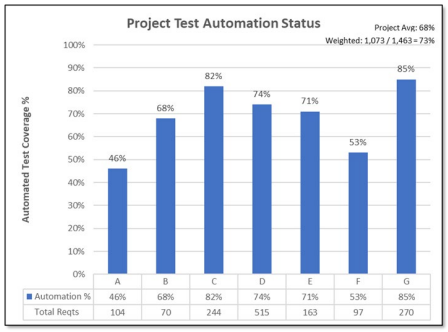
(select 1 from each)



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



Project



Enterprise

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)

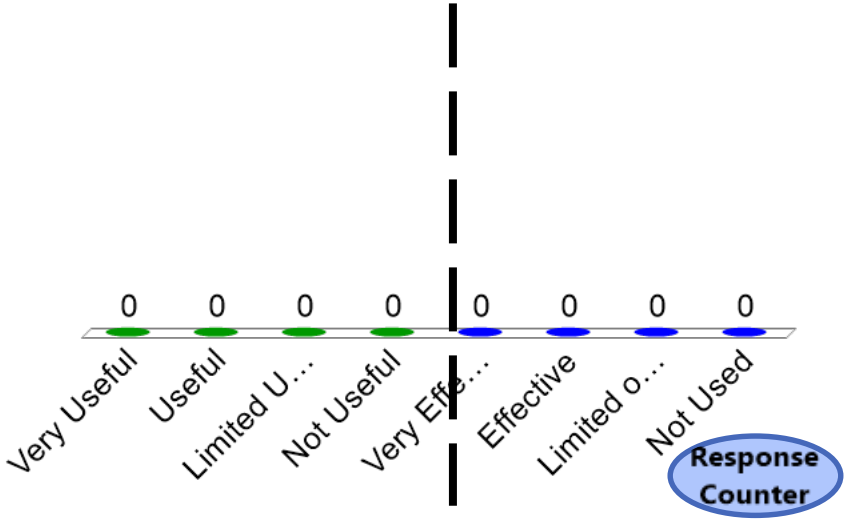
Insight

1. *Very Useful*
2. *Useful*
3. *Limited Usefulness*
4. *Not Useful*

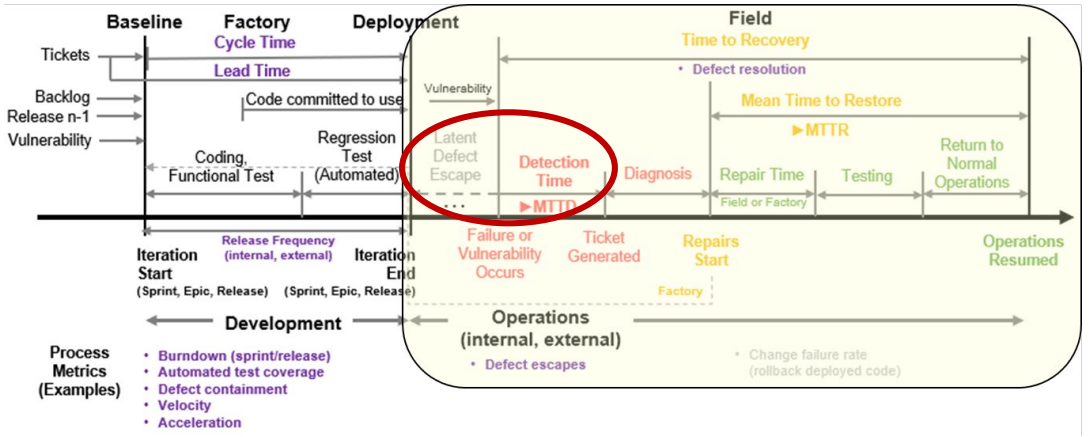
Impact

5. *Very Effective*
6. *Effective*
7. *Limited or No Effectiveness*
8. *Not Used*

(select 1 from each)



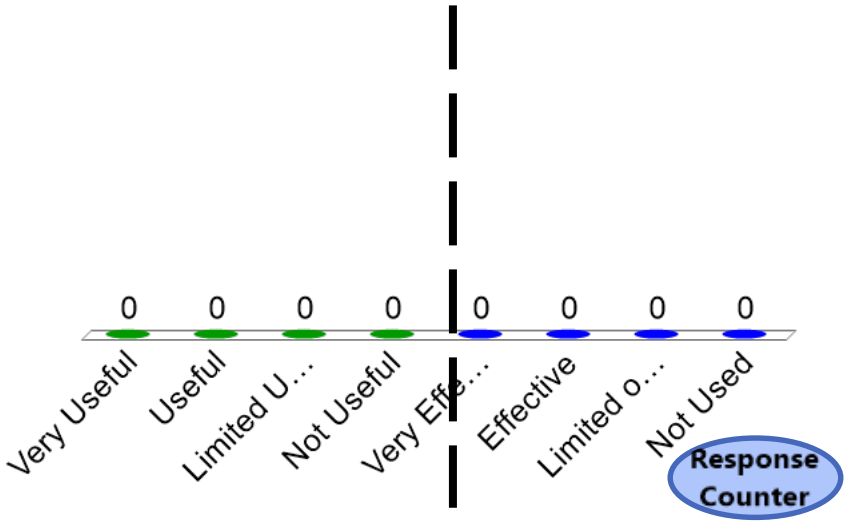
Is Mean Time to Detect (MTTD) a useful measure, and how effectively is it used to provide insight and impactful action in your organization?



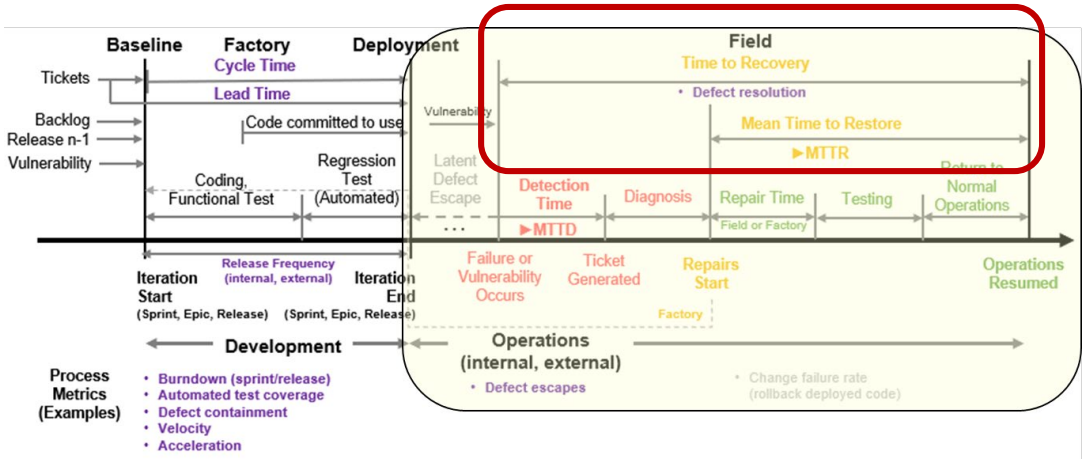
Insight
1. Very Useful
2. Useful
3. Limited Usefulness
4. Not Useful

Impact
5. Very Effective
6. Effective
7. Limited or No Effectiveness
8. Not Used

(select 1 from each)



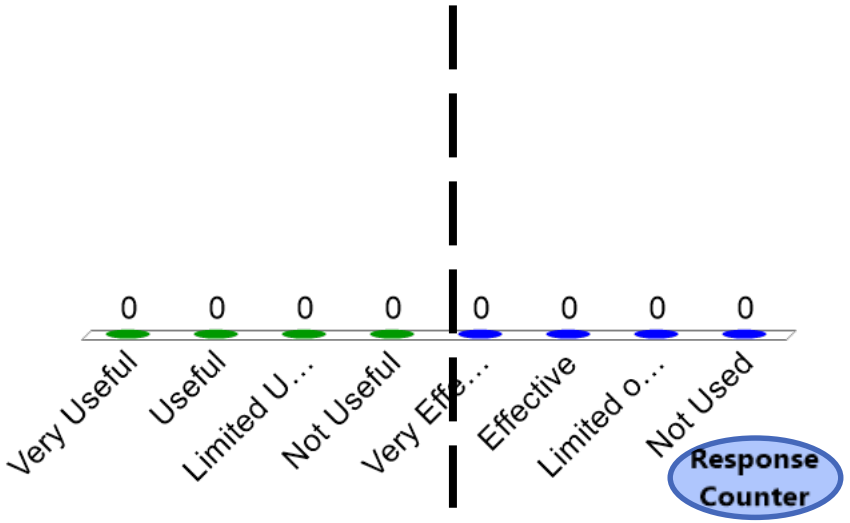
Is Mean Time to Restore (or Recover) (MTTR) a useful measure, and how effectively is it used to provide insight and impactful action in your organization?



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!

Workshop Outbrief….

Workshop Title

Workshop Participants

Summary

Conclusions, Recommendations, and Results

Next Steps/Action Items