

Practical Software and Systems Measurement

Objective Information for Decision Makers



***Workshop - Measures for
Iterative Software
Development and Acquisition***

12 Sep 2018

***NDIA / INCOSE / PSM
Iterative Software Development and
Acquisition Working Group***

Measures for Iterative Software Development and Acquisition

Workshop Objectives:

- *Provide industry recommendations and resources to advance the use of continuous iterative software methods in DoD programs and acquisition*
- *Address recommendations of Defense Science Board (DSB) Software Design and Acquisition Task Group, and Defense Innovation Board (DIB)*

What input should we provide to DoD to support implementation of the DSB task force recommendations?

Workshop Format

- *Interactive evaluation of recommended DSB and DIB Metrics for Software Development*
- *Consider other measures to plan and manage iterative software development in defense acquisition*

Workshop Background

- *Introduction – overview of DSB and DIB recommendations, and NDIA / INCOSE / PSM working group objectives*

Intended Output

- *Progress toward consensus measures for iterative software development and acquisition*
- *Recommendations to DoD for measures and assets to support implementation of the DSB task force recommendations*

Stakeholders

- *Project Lead*
- *Program Mgr*
- *Cost analyst (internal, external)*
- *Program oversight*
- *Development / Team Agile*
- *Enterprise Mgmt*

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Iterative SW Development Measures – PSM User's Conference Workshop

Inputs

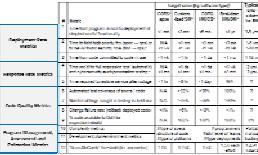
DSB SW Task Force Report



Dr. LaPlante PSM Keynote



Defense Innovation Board (DIB) Measures



Draft NDIA Goals, Questions (GQM) and 'Industry Space' Measures

Measure Type	Industry Space		GQM	
	Definition	Description	Goal	Question
Deployment Rate	Deployment Rate	Deployment Rate	How quickly can we deliver initial capability for new products?	How quickly can we add and deliver high priority capabilities for an existing operational product?
	Automated testing (unit level, system level)	Automated testing (unit level, system level)	How quickly can new security vulnerabilities be patched and deployed to fielded products?	What is the "lead time" duration from code committed to a repository to availability of code?
	Automated test coverage (new functionality)	Automated test coverage (new functionality)	What is the lead time to a product out? (TBO)	What is the code time to a product out? (TBO)
Response Rate	Quickly detect, isolate, and remove software operational defects.	Quickly detect, isolate, and remove software operational defects.	How quickly can a full automated regression test be conducted to verify capability correctness?	How quickly can a complete cybersecurity test be successfully completed to ensure adequate resistance to vulnerabilities?
	Automated test coverage (new functionality)	Automated test coverage (new functionality)	Consider other ITMs: MODS, KPIs	Consider other ITMs: MODS, KPIs
Code Quality	Ensure high quality code in development and in the field.	Ensure high quality code in development and in the field.	Is testing efficiency (e.g., automation) aligned with project plan?	Is testing efficiency (e.g., automation) aligned with project plan?
	Code reuse defines in development tracking where they are most cost efficient.	Code reuse defines in development tracking where they are most cost efficient.	Does new code functionality work as expected and not break previous functionality?	What is the reliability and availability of operational service capabilities?
			Does new product baseline updates reliable in the field such that operational service is not impacted?	What is the reliability and availability of operational service capabilities?
Program Management, Assessment, and Estimation	Achieve effective insight into management of SW programs, including cost assessment and performance estimation.	Achieve effective insight into management of SW programs, including cost assessment and performance estimation.	Does DOD have data rights to sustain the code baseline?	Is the quality of code deployed to the field?
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	How much of the code base is newly developed vs reused from other sources?	How much code has been delivered for each release?
			Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)	Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)
			How stable are the mission capability requirements?	How stable are the mission capability requirements?
			What are appropriate thresholds for SW measures that indicate serious breach relative to the integration and test progress proceeding as planned? (TBO)	What are appropriate thresholds for SW measures that indicate serious breach relative to the integration and test progress proceeding as planned? (TBO)
			Is the integration and test progress proceeding as planned? (TBO)	Are the capabilities and features being implemented and completed as planned? (TBO)

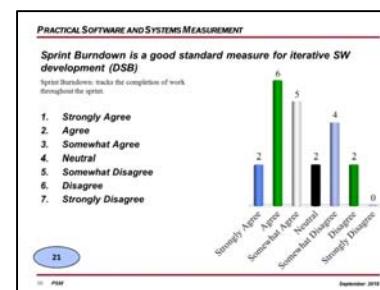
PSM Workshop

- Initial Validation of Categories, Goals, Questions
- Prioritization and Ranking

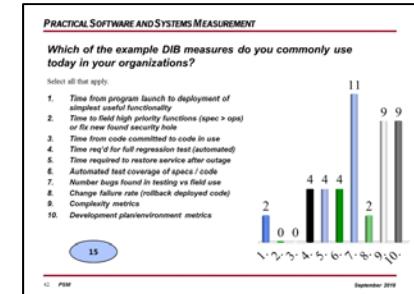
Category (DIB)	Goal (Derived from DIB by PSM/NDIA)	Questions (Derived from DIB by PSM/NDIA)	SW	H	M	L	Total
Deployment Rate	Prioritize speed in delivering value to end users through new operations capabilities.	How quickly can we deliver initial capability for new products?	10	1	1	1	10
	Automated testing (unit level, system level)	How quickly can we add and deliver high priority capabilities for an existing operational product?	12	1	1	1	15
	Automated test coverage (new functionality)	How quickly can new security vulnerabilities be patched and deployed to fielded products?	10	1	1	1	10
		What is the "lead time" duration from code committed to a repository to availability of code?	1	1	1	1	3
		What is the code time to a product out? (TBO)	1	1	1	1	3
Response Rate	Quickly detect, isolate, and remove software operational defects.	How quickly can a full automated regression test be conducted to verify capability correctness?	9	1	1	1	10
	Automated test coverage (new functionality)	How quickly can a complete cybersecurity test be successfully completed to ensure adequate resistance to vulnerabilities?	7	1	1	1	9
	Automated test coverage (new functionality)	What is the reliability and availability of operational service capabilities?	10	1	1	1	10
		Consider other ITMs: MODS, KPIs	10	1	1	1	20
		How quickly can we address bug reports from the field? (TBO)	1	1	1	1	3
Code Quality	Ensure high quality code in development and in the field.	Is testing efficiency (e.g., automation) aligned with project plan?	6	1	1	1	10
	Code reuse defines in development tracking where they are most cost efficient.	Does new code functionality work as expected and not break previous functionality?	10	1	1	1	10
		Does new product baseline updates reliable in the field such that operational service is not impacted?	10	1	1	1	10
		Does DOD have data rights to sustain the code baseline?	5	1	1	1	20
		How much of the code base is newly developed vs reused from other sources?	1	1	1	1	3
		How much code has been delivered for each release?	10	1	1	1	10
		Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)	1	1	1	1	3
		How stable are the mission capability requirements?	6	1	1	1	10
		What are appropriate thresholds for SW measures that indicate serious breach relative to the integration and test progress proceeding as planned? (TBO)	1	1	1	1	3
		Is the integration and test progress proceeding as planned? (TBO)	1	1	1	1	3
		Are the capabilities and features being implemented and completed as planned? (TBO)	1	1	1	1	3

Interactive Scoring of Draft DSB/DIB Measures

Brainstorming on Additional Potential Measures

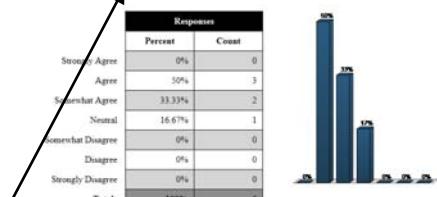


- Sampling of Organizational Adoption/Usage of Proposed Measures

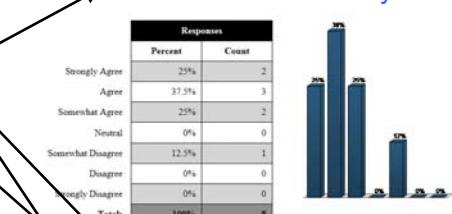


Data Slicing by Demographic

U.S. Government



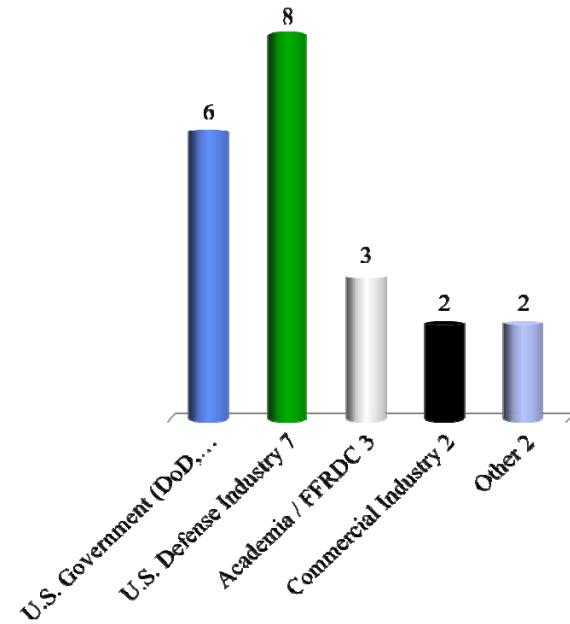
U.S. Defense Industry



Academia / FFRDC
Commercial Industry
Other

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***



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. U.S. Government (DoD, agency) 6	6	0	0	0	0
2. U.S. Defense Industry 7	0	8	0	0	0
3. Academia / FFRDC 3	0	0	3	0	0
4. Commercial Industry 2	0	0	0	2	0
5. Other 2	0	0	0	0	2

21

Recommendation 3: Risk Reduction and Metrics for New Programs



Recommendation 3: Risk Reduction and Metrics for New Programs

For all new programs, starting immediately, the following best practices should be implemented in formal program acquisition strategies.

The MDA (with the DAE, the SAE, the PEO, and the PM) should allow multiple vendors to begin work. A down-select should happen after at least one vendor has proven they can do the work, and should retain several vendors through development to reduce risk, as feasible.

The MDA with the Cost Assessment and Program Evaluation office (CAPE), the USD(R&E), the Service Cost Estimators, and others should modernize cost and schedule estimates and measurements. They should evolve from a pure SLOC approach to historical comparables as a measurement, and should adopt the National Reconnaissance Office (NRO) approach (demonstrated in Box 5) of contracting with the defense industrial base for work breakdown schedule data to include, among others, staff, cost, and productivity.

The MDA should immediately require the PM to build a program-appropriate framework for status estimation. Example metrics include:¹⁸

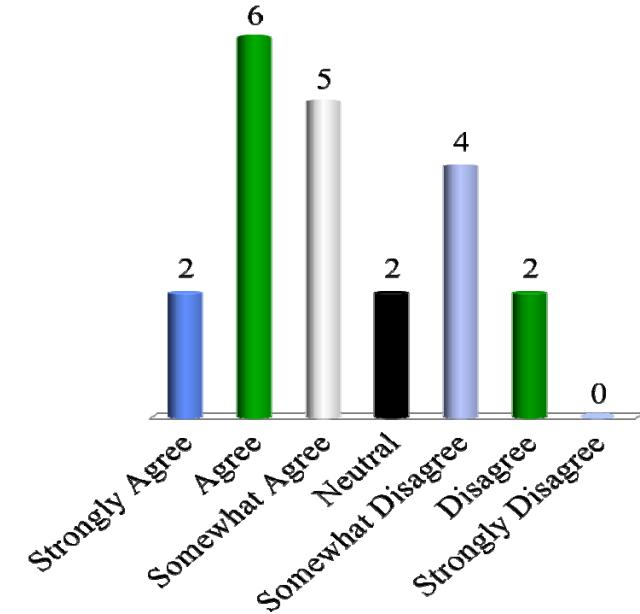
- Sprint Burndown: tracks the completion of work throughout the sprint.
- Epic and Release Burndown: tracks the progress of development over a larger body of work than a sprint.
- Velocity: the average amount of work a team completes during a sprint.
- Control Chart: focus on the cycle time of individual issues—the total time from “in progress” to “complete.”
- Cumulative Flow Diagram: shows whether the flow of work across the team is consistent; visually points out shortages and bottlenecks.

There may be short-term costs in transitioning to iterative development (e.g., software factory, training). However, based on the experience of the commercial sector, net costs can be expected to decrease after adopting iterative development.

Sprint Burndown is a good standard measure for iterative SW development (DSB)

Sprint Burndown: tracks the completion of work throughout the sprint.

1. ***Strongly Agree***
2. ***Agree***
3. ***Somewhat Agree***
4. ***Neutral***
5. ***Somewhat Disagree***
6. ***Disagree***
7. ***Strongly Disagree***



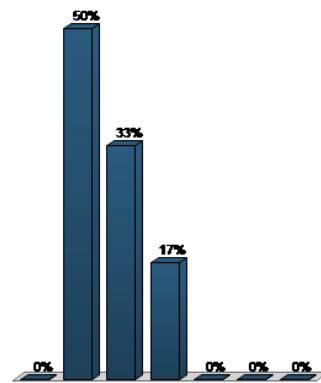
	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	2	0	0	0
2. Agree	3	3	0	0	0
3. Somewhat Agree	2	2	1	0	0
4. Neutral	1	0	0	1	0
5. Somewhat Disagree	0	1	1	1	1
6. Disagree	0	0	1	0	1
7. Strongly Disagree	0	0	0	0	0

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Sprint Burndown is a good standard measure for iterative SW development (DSB)

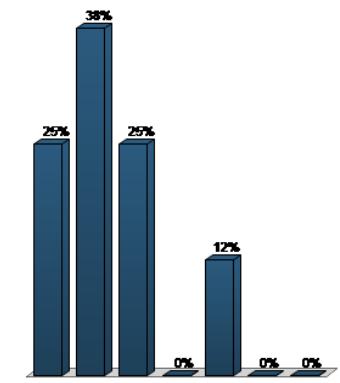
U.S. Government

Responses	
Percent	Count
0%	0
50%	3
33.33%	2
16.67%	1
0%	0
0%	0
0%	0
Totals	6



U.S. Defense Industry

Responses	
Percent	Count
25%	2
37.5%	3
25%	2
0%	0
12.5%	1
0%	0
0%	0
Totals	8



Academia / FFRDC

Responses	
Percent	Count
0%	0
0%	0
33.33%	1
0%	0
33.33%	1
33.33%	1
0%	0
Totals	3

Commercial Industry

Responses	
Percent	Count
0%	0
0%	0
0%	0
50%	1
50%	1
0%	0
0%	0
Totals	2

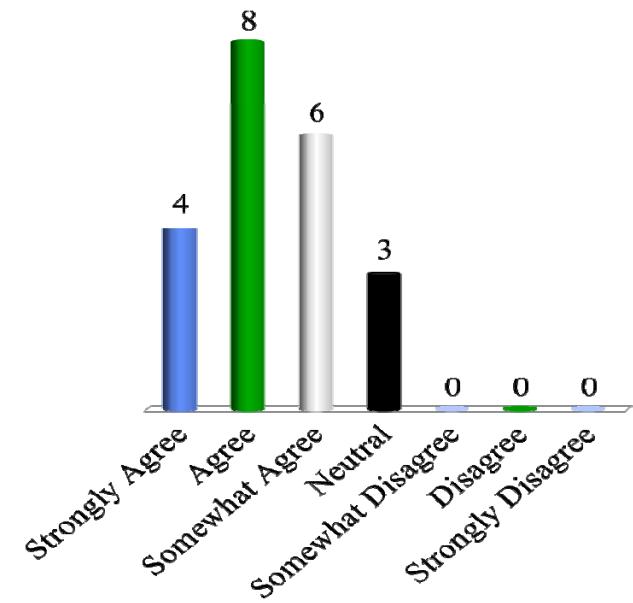
Other

Responses	
Percent	Count
0%	0
0%	0
0%	0
50%	1
50%	1
0%	0
0%	0
Totals	2

Epic and Release Burndown is a good standard measure for iterative SW development (DSB)

Epic and Release Burndown: tracks the progress of development over a larger body of work than a sprint.

- 1. *Strongly Agree***
- 2. *Agree***
- 3. *Somewhat Agree***
- 4. *Neutral***
- 5. *Somewhat Disagree***
- 6. *Disagree***
- 7. *Strongly Disagree***



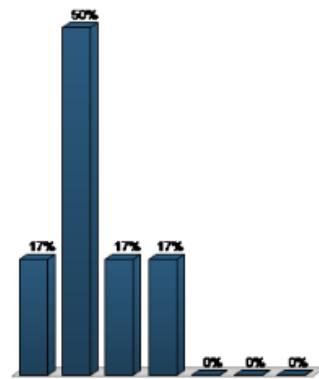
	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	1	3	0	0	0
2. Agree	3	4	0	1	0
3. Somewhat Agree	1	1	2	0	2
4. Neutral	1	0	1	1	0
5. Somewhat Disagree	0	0	0	0	0
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

21

Epic and Release Burndown is a good standard measure for iterative SW development (DSB)

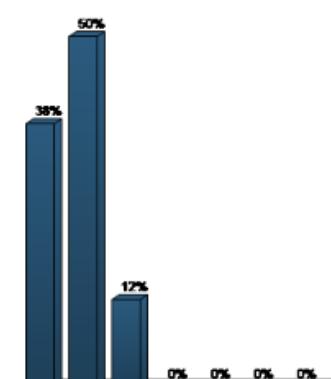
U.S. Government

Responses		
	Percent	Count
Strongly Agree	16.67%	1
Agree	50%	3
Somewhat Agree	16.67%	1
Neutral	16.67%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6

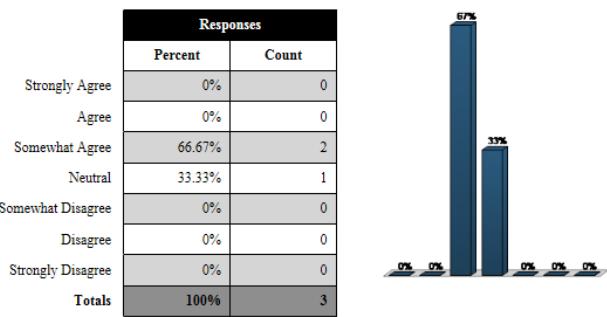


U.S. Defense Industry

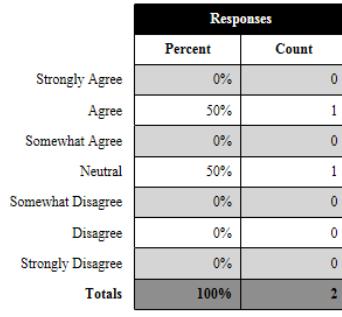
Responses		
	Percent	Count
Strongly Agree	37.5%	3
Agree	50%	4
Somewhat Agree	12.5%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	8



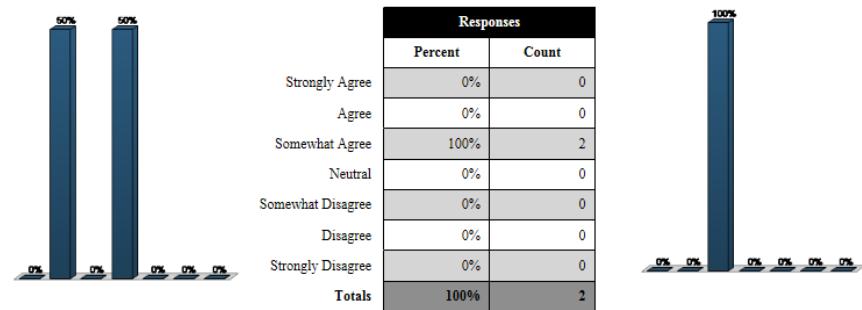
Academia / FFRDC



Commercial Industry



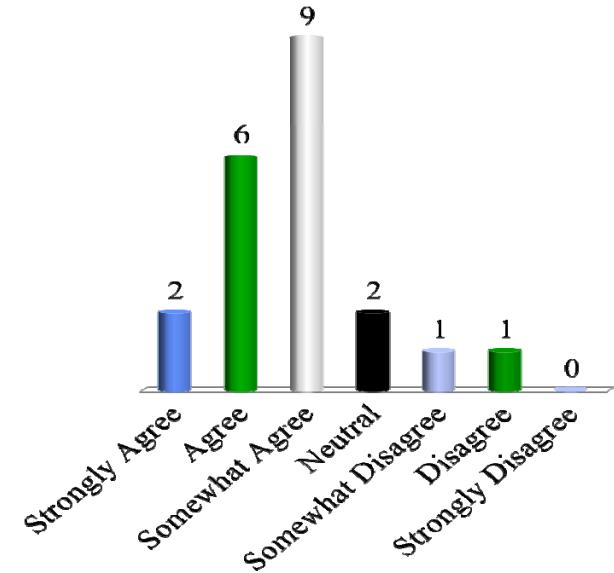
Other



Velocity is a good standard measure for iterative SW development (DSB)

Velocity: the average amount of work a team completes during a sprint.

- 1. *Strongly Agree***
- 2. *Agree***
- 3. *Somewhat Agree***
- 4. *Neutral***
- 5. *Somewhat Disagree***
- 6. *Disagree***
- 7. *Strongly Disagree***



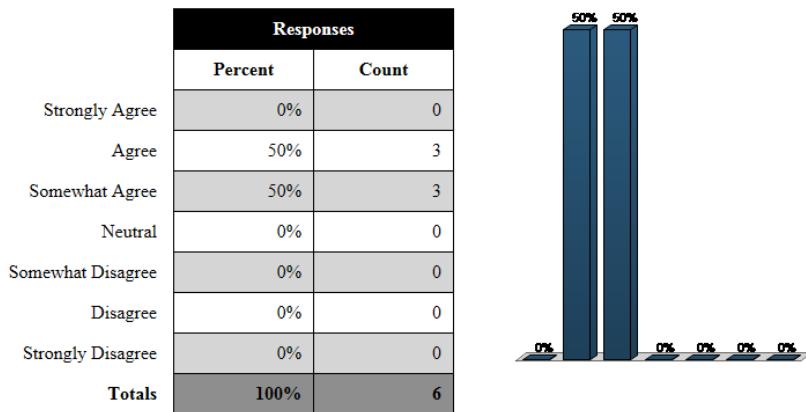
	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	2	0	0	0
2. Agree	3	2	1	0	0
3. Somewhat Agree	3	3	1	1	1
4. Neutral	0	1	1	0	0
5. Somewhat Disagree	0	0	0	1	0
6. Disagree	0	0	0	0	1
7. Strongly Disagree	0	0	0	0	0

21

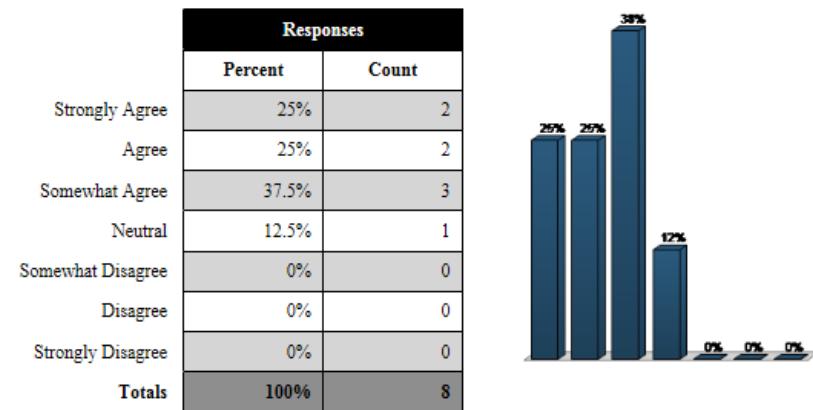
PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Velocity is a good standard measure for iterative SW development (DSB)

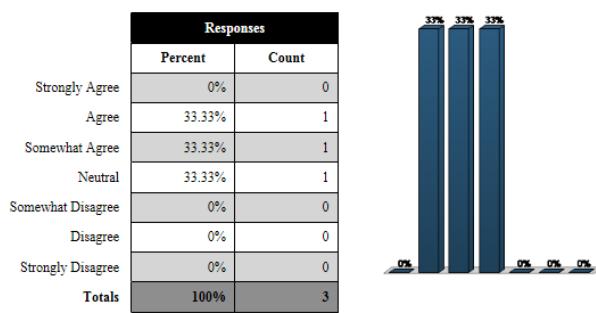
U.S. Government



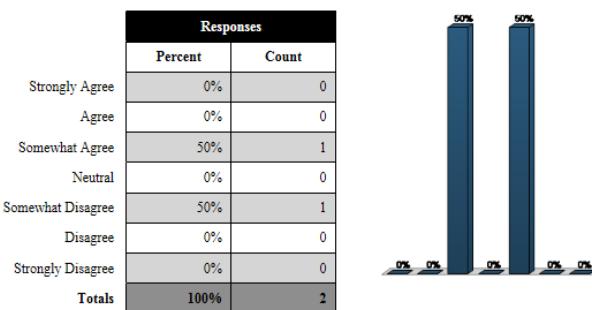
U.S. Defense Industry



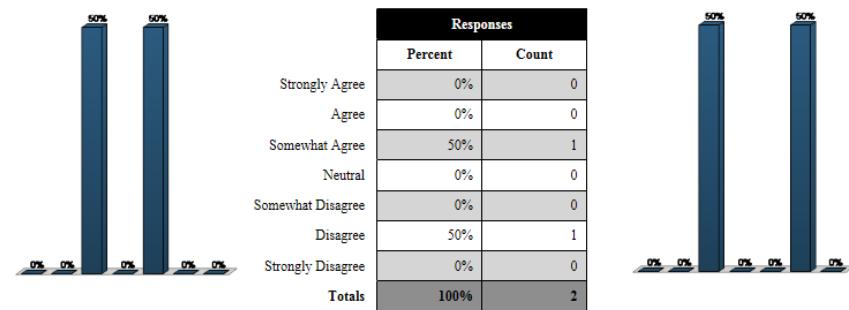
Academia / FFRDC



Commercial Industry



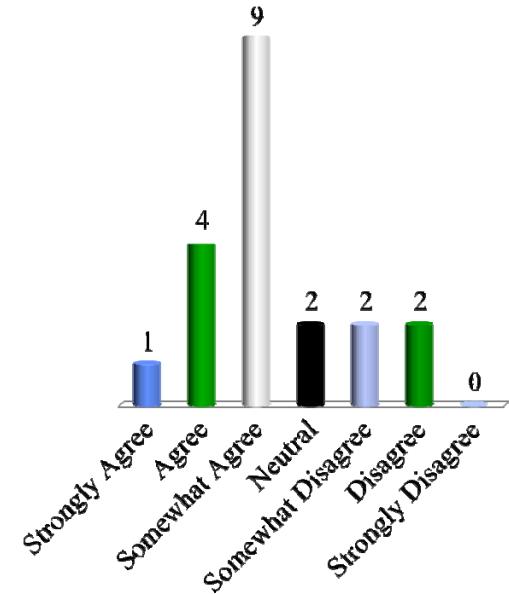
Other



Cycle Time (Control Chart) is a good standard measure for iterative SW development (DSB)

Control Chart: focus on the cycle time of individual issues
 – the total time from “in progress” to “complete”

- 1. Strongly Agree**
- 2. Agree**
- 3. Somewhat Agree**
- 4. Neutral**
- 5. Somewhat Disagree**
- 6. Disagree**
- 7. Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	1	0	0	0
2. Agree	1	2	0	1	0
3. Somewhat Agree	2	4	2	1	0
4. Neutral	1	1	0	0	0
5. Somewhat Disagree	1	0	0	0	1
6. Disagree	0	0	1	0	1
7. Strongly Disagree	0	0	0	0	0

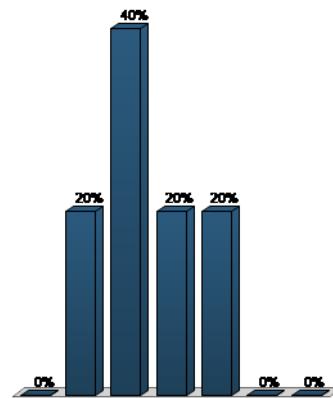
20

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Cycle Time (Control Chart) is a good standard measure for iterative SW development (DSB)

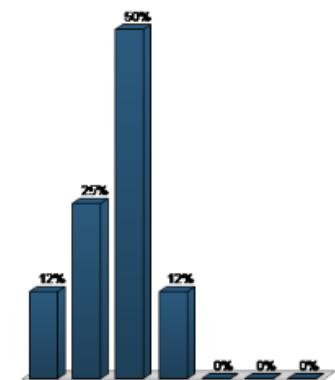
U.S. Government

Responses	
Percent	Count
0%	0
20%	1
40%	2
20%	1
20%	1
0%	0
0%	0
100%	5



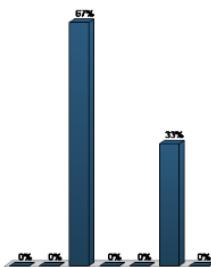
U.S. Defense Industry

Responses	
Percent	Count
12.5%	1
25%	2
50%	4
12.5%	1
0%	0
0%	0
0%	0
100%	8



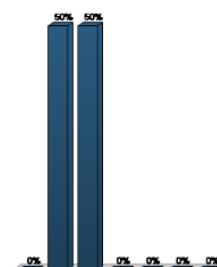
Academia / FFRDC

Responses	
Percent	Count
0%	0
0%	0
66.67%	2
0%	0
0%	0
33.33%	1
0%	0
100%	3



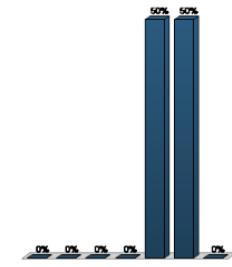
Commercial Industry

Responses	
Percent	Count
0%	0
50%	1
50%	1
0%	0
0%	0
0%	0
0%	0
100%	2



Other

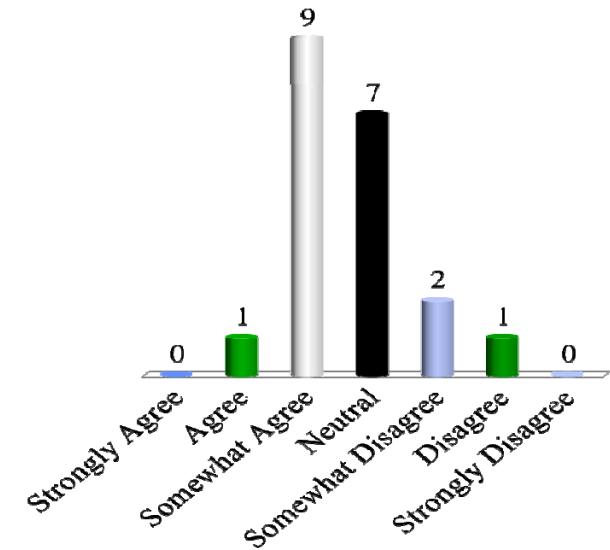
Responses	
Percent	Count
0%	0
0%	0
0%	0
50%	1
50%	1
0%	0
100%	2



Cumulative Flow Diagram is a good standard measure for iterative SW development (DSB)

Cumulative Flow Diagram: shows whether the flow of work across the team is consistent; visually points out shortages and bottlenecks

- 1. Strongly Agree**
- 2. Agree**
- 3. Somewhat Agree**
- 4. Neutral**
- 5. Somewhat Disagree**
- 6. Disagree**
- 7. Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	0	0	0	0
2. Agree	0	0	0	1	0
3. Somewhat Agree	2	4	1	1	1
4. Neutral	3	2	1	0	1
5. Somewhat Disagree	1	1	0	0	0
6. Disagree	0	1	0	0	0
7. Strongly Disagree	0	0	0	0	0

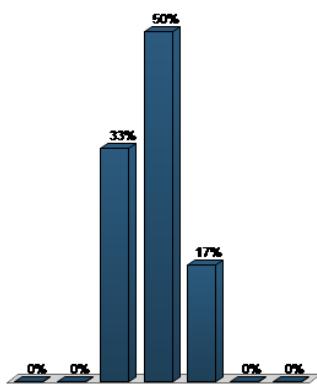
19

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Cumulative Flow Diagram is a good standard measure for iterative SW development (DSB)

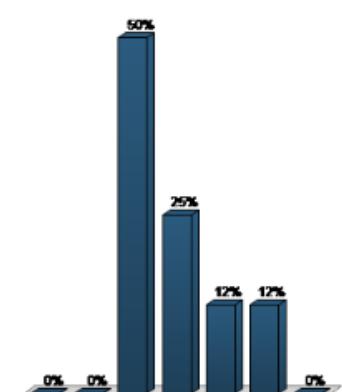
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	33.33%	2
Neutral	50%	3
Somewhat Disagree	16.67%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



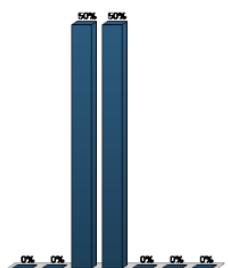
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	4
Neutral	25%	2
Somewhat Disagree	12.5%	1
Disagree	12.5%	1
Strongly Disagree	0%	0
Totals	100%	8



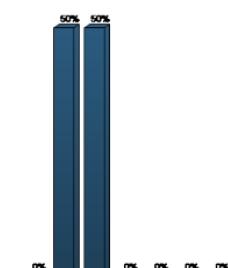
Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	1
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



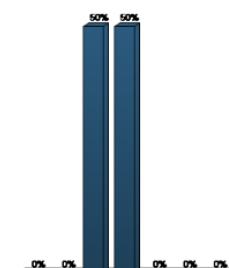
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	50%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

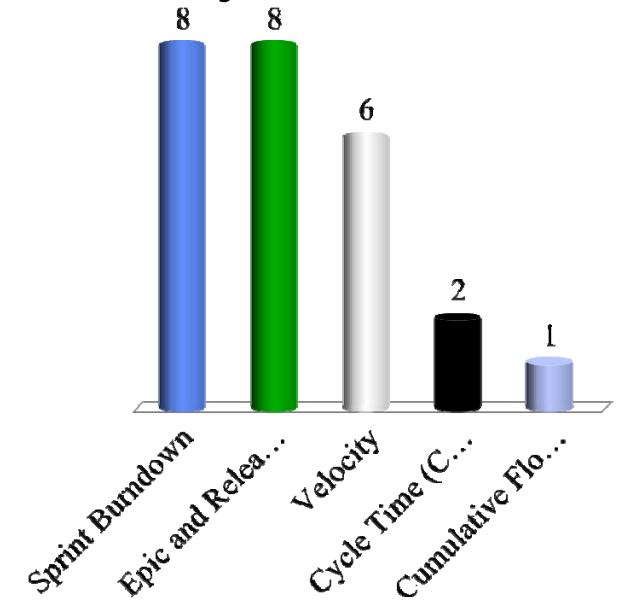
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	1
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Which of the example DSB measures do you commonly use today in your organizations?

Select all that apply.

1. **Sprint Burndown**
2. **Epic and Release Burndown**
3. **Velocity**
4. **Cycle Time (Control Chart)**
5. **Cumulative Flow Diagram**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Sprint Burndown	1	5	2	0	0
2. Epic and Release Burndown	2	5	1	0	0
3. Velocity	1	4	1	0	0
4. Cycle Time (Control Chart)	0	2	0	0	0
5. Cumulative Flow Diagram	0	0	1	0	0

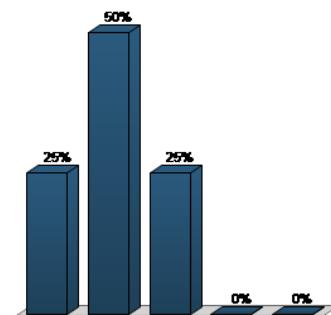
12

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Which of the example DSB measures do you commonly use today in your organizations?

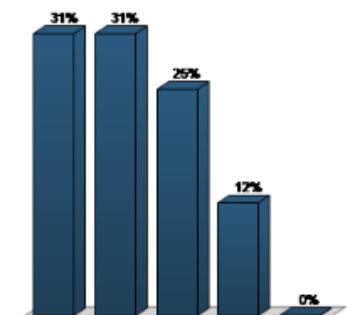
U.S. Government

Responses		
	Percent	Count
Sprint Burndown	25%	1
Epic and Release Burndown	50%	2
Velocity	25%	1
Cycle Time (Control Chart)	0%	0
Cumulative Flow Diagram	0%	0
Totals	100%	4



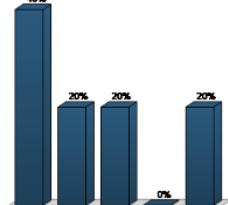
U.S. Defense Industry

Responses		
	Percent	Count
Sprint Burndown	31.25%	5
Epic and Release Burndown	31.25%	5
Velocity	25%	4
Cycle Time (Control Chart)	12.5%	2
Cumulative Flow Diagram	0%	0
Totals	100%	16



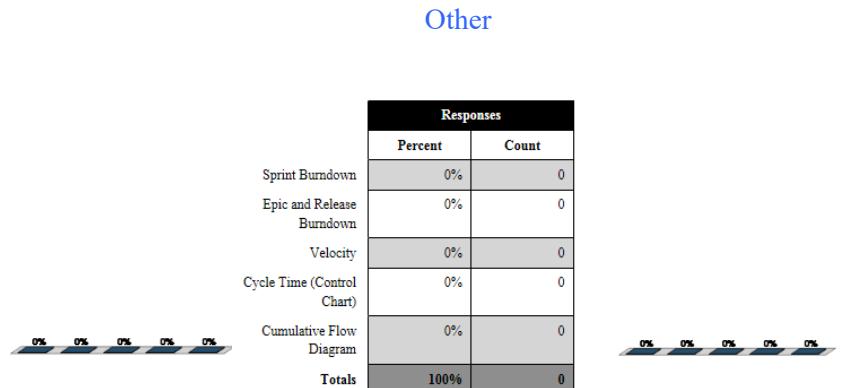
Academia / FFRDC

Responses		
	Percent	Count
Sprint Burndown	40%	2
Epic and Release Burndown	20%	1
Velocity	20%	1
Cycle Time (Control Chart)	0%	0
Cumulative Flow Diagram	20%	1
Totals	100%	5



Commercial Industry

Responses		
	Percent	Count
Sprint Burndown	0%	0
Epic and Release Burndown	0%	0
Velocity	0%	0
Cycle Time (Control Chart)	0%	0
Cumulative Flow Diagram	0%	0
Totals	100%	0



Other

DIB Proposed Software Metrics for DoD

https://media.defense.gov/2018/Jul/10/2001940937/-1/-1/0/DIB_METRICS_FOR_SOFTWARE_DEVELOPMENT_V0.9_2018.07.10.PDF

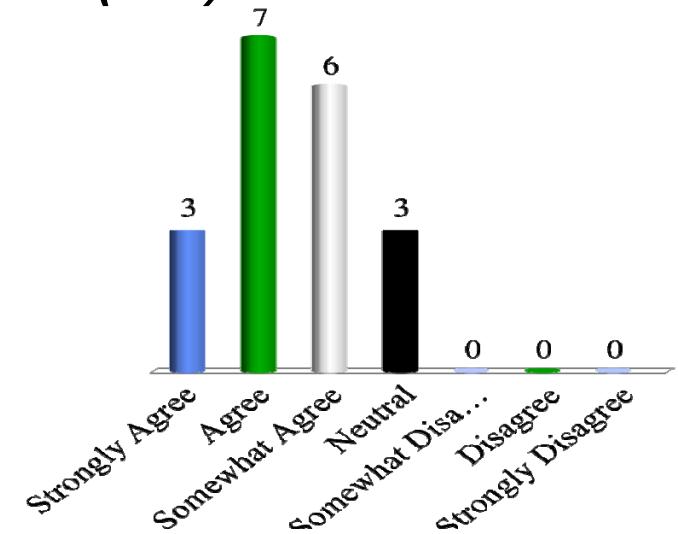


#	Metric	Target value (by software type) ⁱ				Typical DoD values for SW
		COTS ⁱⁱ apps	Custom -ized SW ⁱⁱⁱ	COTS HW/OS ^{iv}	Real-time HW/SW ^v	
Deployment Rate Metrics	1 Time from program launch to deployment of simplest useful functionality	<1 mo	<3 mo	<6 mo	<1 yr	3-5 yrs
	2 Time to field high priority fcn (spec → ops) or fix newly found security hole (find → ops) ^{vi}	N/A <1 wk	<1 mo <1 wk	<3 mo <1 wk	<3 mo <1 wk	1-5 yrs 1-18 m
	3 Time from code committed to code in use	<1 wk	<1 hr	<1 da	<1 mo	1-18 m
Response Rate Metrics	4 Time req'd for full regression test (automat'd) and cybersecurity audit/penetration testing ^{vii}	N/A <1 mo	<1 da <1 mo	<1 da <1 mo	<1 wk <3 mo	2 yrs 2 yrs
	5 Time required to restore service after outage	<1 hr	<6 hr	<1 day	N/A	?
Code Quality Metrics	6 Automated test coverage of specs / code	N/A	>90%	>90%	100%	?
	7 Number of bugs caught in testing vs field use	N/A	>75%	>75%	>90%	?
	8 Change failure rate (rollback deployed code)	<1%	<5%	<10%	<1%	?
	9 % code available to DoD for inspection/rebuild	N/A	100%	100%	100%	0%
Program Management, Assessment and Estimation Metrics	10 Complexity metrics	#/type of specs structure of code #/type of platforms			# programmers #/skill level of teams #/type deployments	Partial/manual tracking Various: Static code analysis measures New vs. NDI % Estimating parameters/assumptions Other TBD
	11 Development plan/environment metrics					
	12 "Nunn-McCurdy" threshold (for any metric)	1.1X	1.25X	1.5X	1.5X each effort	
(N/A – DoD acquisition measure)						

'Time from program launch to deployment of simplest useful functionality' is a good standard SW measure (DIB)

Aka: Initial capability cycle time (v1.0 release)

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	3	0	0	0
2. Agree	1	1	2	2	1
3. Somewhat Agree	3	2	0	0	1
4. Neutral	1	1	1	0	0
5. Somewhat Disagree	0	0	0	0	0
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

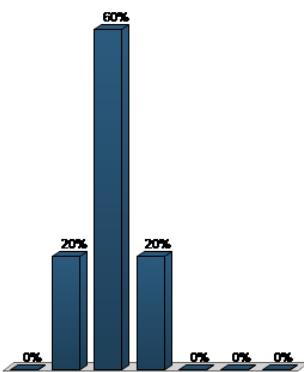
19

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time from program launch to deployment of simplest useful functionality' is a good standard SW measure (DIB)

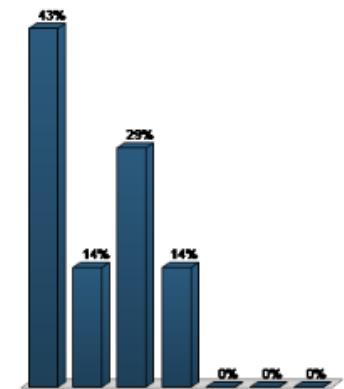
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	20%	1
Somewhat Agree	60%	3
Neutral	20%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	5



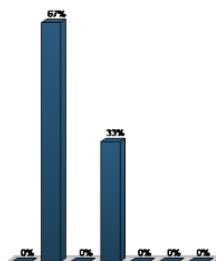
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	42.86%	3
Agree	14.29%	1
Somewhat Agree	28.57%	2
Neutral	14.29%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	7



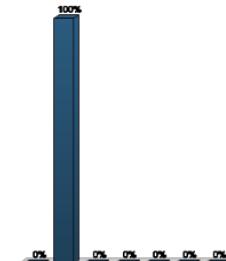
Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	2
Somewhat Agree	0%	0
Neutral	33.33%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3



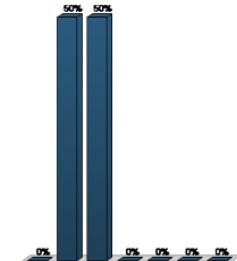
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	100%	2
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

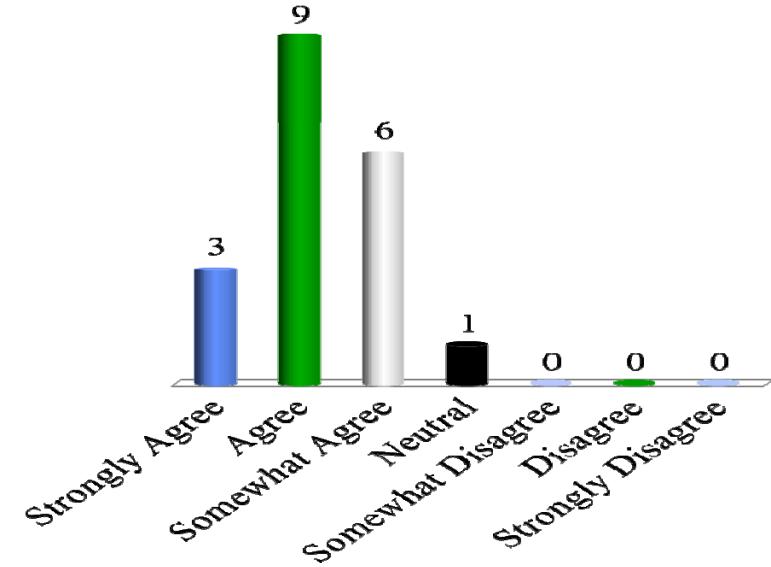
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	50%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



'Time to field high priority functions (spec > ops)' is a good standard SW measure (DIB)

Aka: Incremental capability cycle time (v1.x)

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	2	0	1
2. Agree	3	3	1	0
3. Somewhat Agree	3	0	1	0
4. Neutral	0	1	0	0
5. Somewhat Disagree	0	0	0	0
6. Disagree	0	0	0	0
7. Strongly Disagree	0	0	0	0

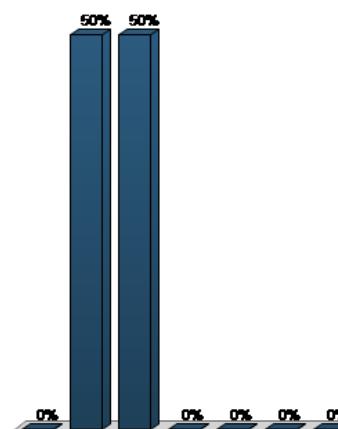
19

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time to field high priority functions (spec > ops)' is a good standard SW measure (DIB)

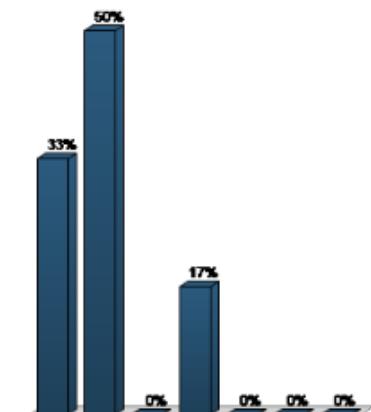
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	3
Somewhat Agree	50%	3
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	33.33%	2
Agree	50%	3
Somewhat Agree	0%	0
Neutral	16.67%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6

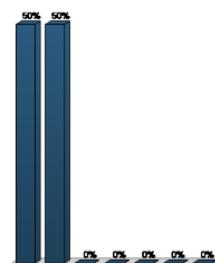


Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	2
Somewhat Agree	33.33%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3

Commercial Industry

Responses		
	Percent	Count
Strongly Agree	50%	1
Agree	50%	1
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



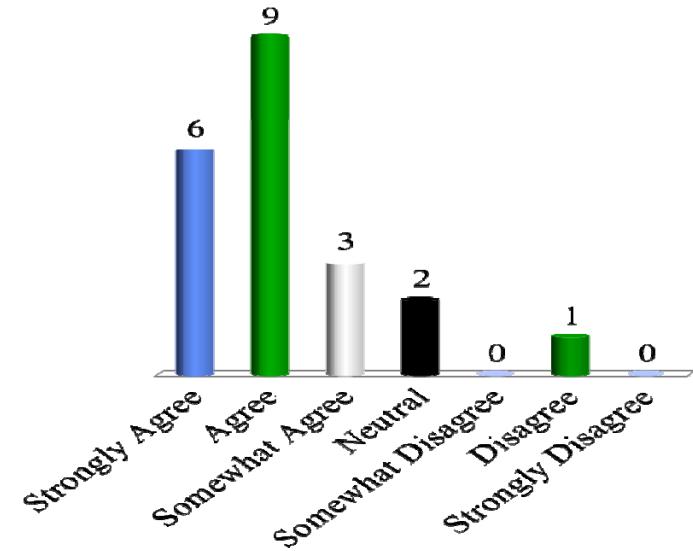
Other

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	100%	2
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2

'Time to fix newly found security hole (find > ops)' is a good standard SW measure (DIB)

Aka: Patch cycle time (v.1.x.y)

- 1. Strongly Agree**
- 2. Agree**
- 3. Somewhat Agree**
- 4. Neutral**
- 5. Somewhat Disagree**
- 6. Disagree**
- 7. Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	2	2	0	2	0
2. Agree	4	3	2	0	0
3. Somewhat Agree	0	2	0	0	1
4. Neutral	0	1	0	0	1
5. Somewhat Disagree	0	0	0	0	0
6. Disagree	0	0	1	0	0
7. Strongly Disagree	0	0	0	0	0

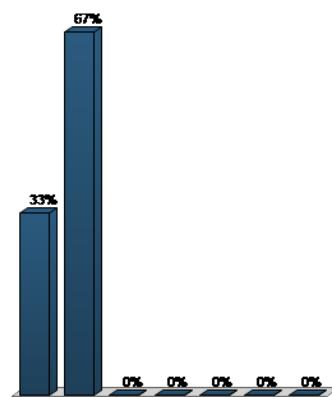
21

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time to fix newly found security hole (find > ops)' is a good standard SW measure (DIB)

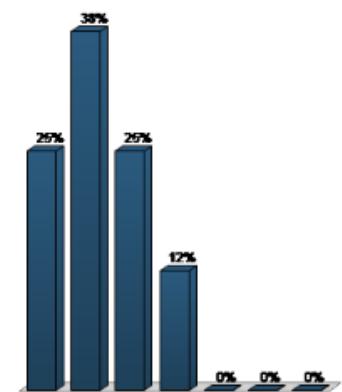
U.S. Government

Responses		
	Percent	Count
Strongly Agree	33.33%	2
Agree	66.67%	4
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	25%	2
Agree	37.5%	3
Somewhat Agree	25%	2
Neutral	12.5%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	8

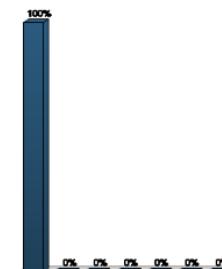


Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	2
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	33.33%	1
Strongly Disagree	0%	0
Totals	100%	3

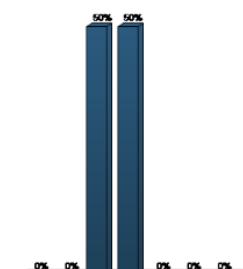
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	100%	2
Agree	0%	0
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

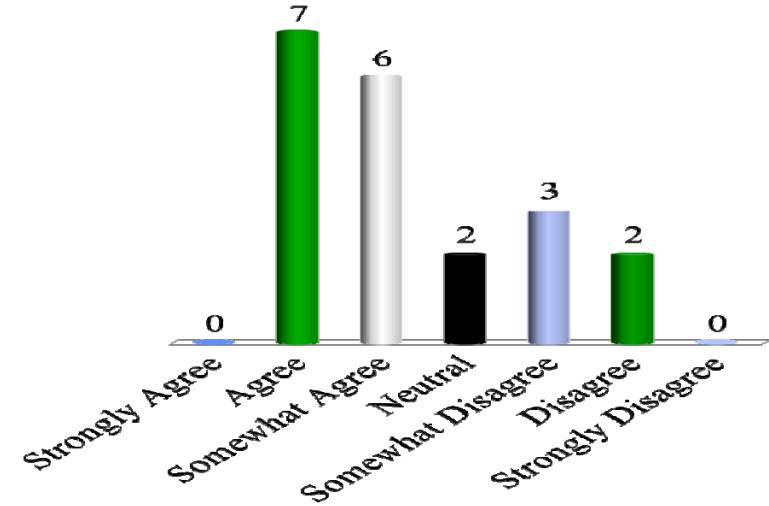
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	1
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



'Time from code committed to code in use' is a good standard SW measure (DIB)

Aka: Factory cycle time

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	0	0	0	0
2. Agree	0	6	1	0	0
3. Somewhat Agree	3	1	2	0	0
4. Neutral	1	0	0	1	0
5. Somewhat Disagree	1	1	0	0	1
6. Disagree	1	0	0	1	0
7. Strongly Disagree	0	0	0	0	0

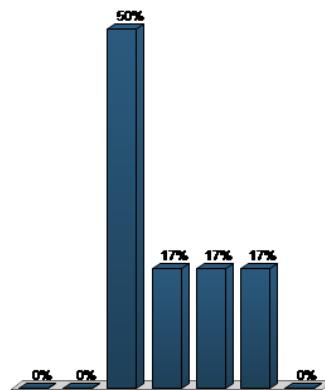
20

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time from code committed to code in use' is a good standard SW measure (DIB)

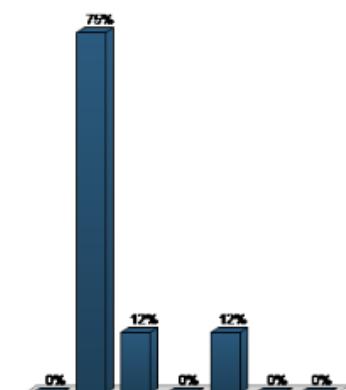
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	3
Neutral	16.67%	1
Somewhat Disagree	16.67%	1
Disagree	16.67%	1
Strongly Disagree	0%	0
Totals	100%	6



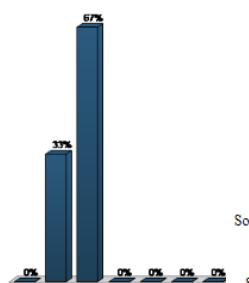
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	75%	6
Somewhat Agree	12.5%	1
Neutral	0%	0
Somewhat Disagree	12.5%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	8



Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	33.33%	1
Somewhat Agree	66.67%	2
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3



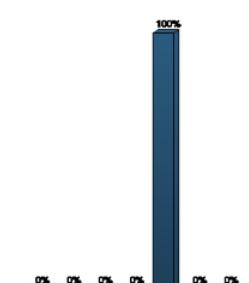
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	0%	0
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	50%	1
Strongly Disagree	0%	0
Totals	100%	2



Other

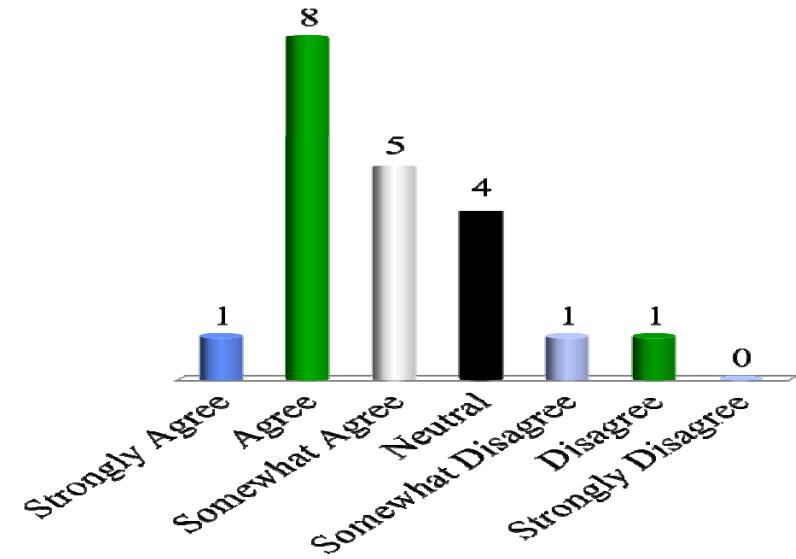
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	100%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	1



'Time required for full regression test (automated)' is a good standard SW measure (DIB)

Aka: Regression test cycle time /
Cybersecurity testing cycle time

- 1. Strongly Agree**
- 2. Agree**
- 3. Somewhat Agree**
- 4. Neutral**
- 5. Somewhat Disagree**
- 6. Disagree**
- 7. Strongly Disagree**



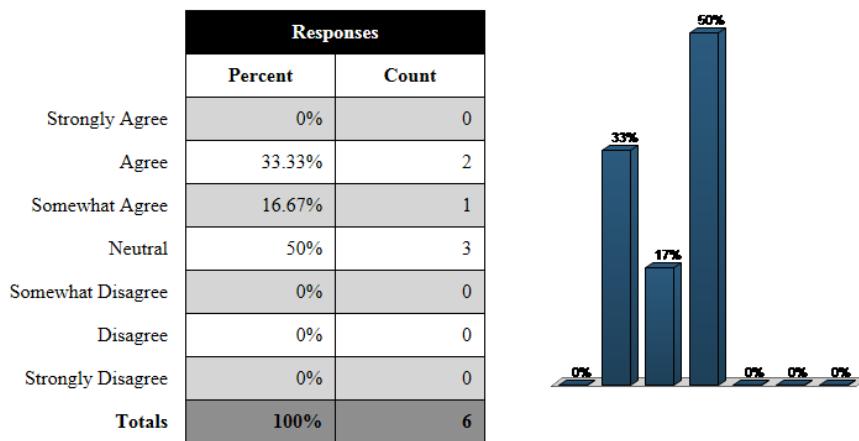
	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	1	0	0	0
2. Agree	2	3	1	1	1
3. Somewhat Agree	1	2	1	1	0
4. Neutral	3	0	0	0	1
5. Somewhat Disagree	0	0	1	0	0
6. Disagree	0	1	0	0	0
7. Strongly Disagree	0	0	0	0	0

20

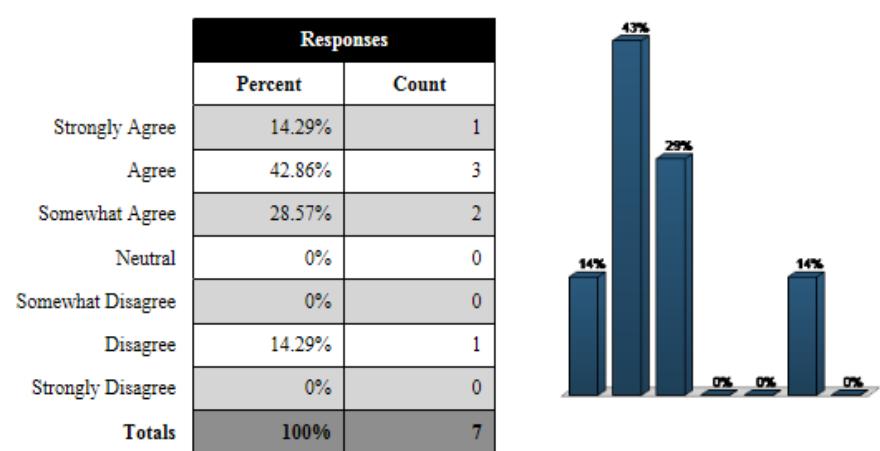
PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time required for full regression test (automated)' is a good standard SW measure (DIB)

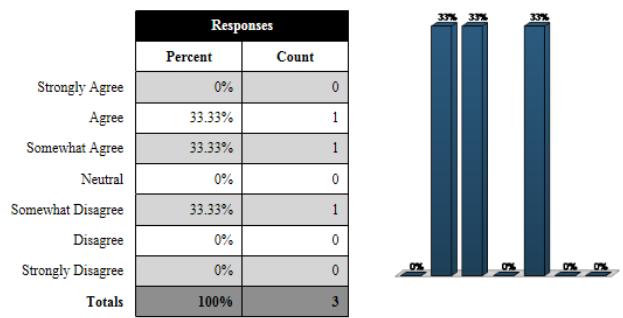
U.S. Government



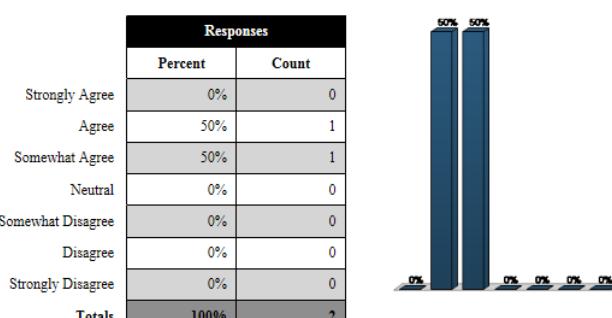
U.S. Defense Industry



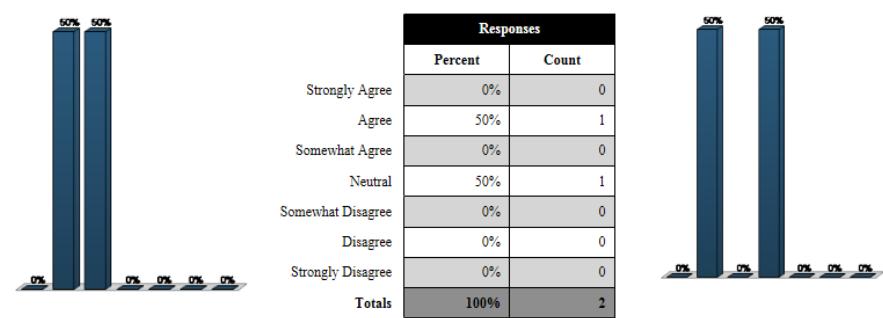
Academia / FFRDC



Commercial Industry



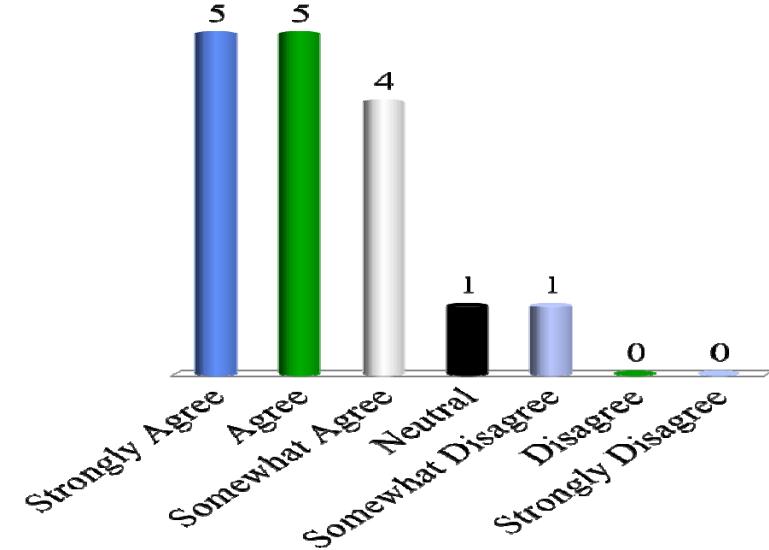
Other



'Time required to restore service after outage' is a good standard SW measure (DIB)

Aka: Mean Time to Repair (MTTR)

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	1	3	1	0	0
2. Agree	2	0	2	1	0
3. Somewhat Agree	2	1	0	1	0
4. Neutral	0	1	0	0	0
5. Somewhat Disagree	0	0	0	0	1
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

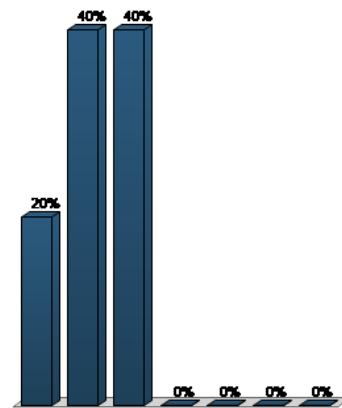
16

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Time required to restore service after outage' is a good standard SW measure (DIB)

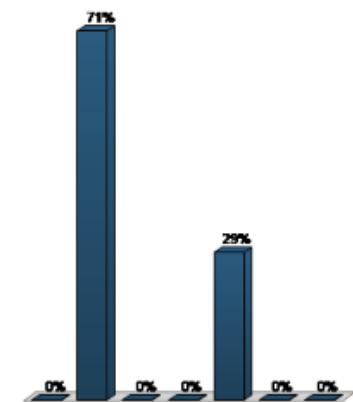
U.S. Government

Responses		
	Percent	Count
Strongly Agree	20%	1
Agree	40%	2
Somewhat Agree	40%	2
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	5



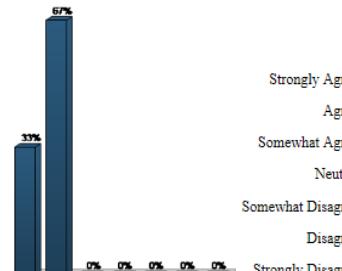
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	71.43%	5
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	28.57%	2
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	7



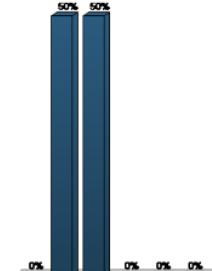
Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	33.33%	1
Agree	66.67%	2
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3



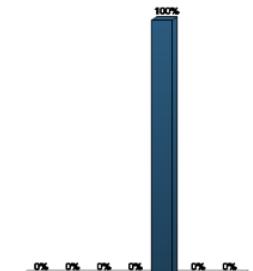
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	50%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

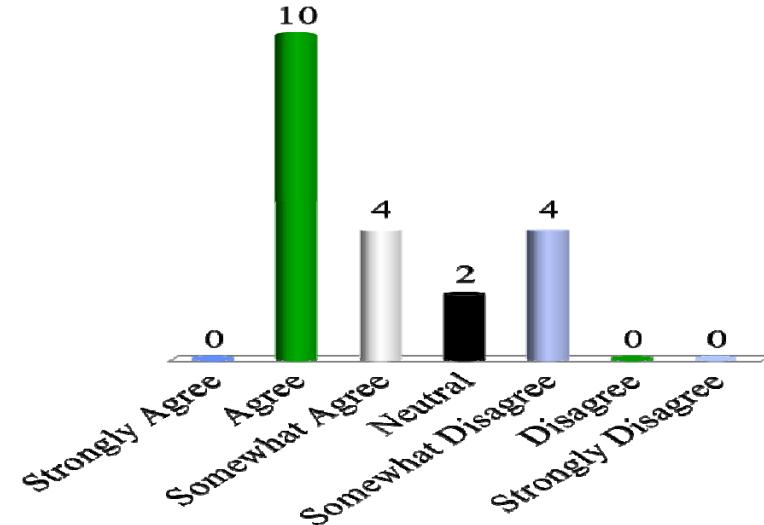
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	100%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	1



'Automated test coverage of test specs / code' is a good standard SW measure (DIB)

Aka: Automated test coverage %

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	0	0	0	0
2. Agree	2	5	2	1	0
3. Somewhat Agree	2	0	1	1	0
4. Neutral	1	0	0	0	1
5. Somewhat Disagree	1	2	0	0	1
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

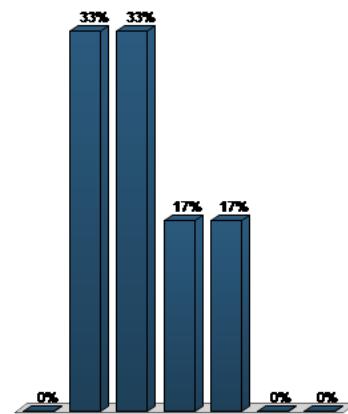
20

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Automated test coverage of test specs / code' is a good standard SW measure (DIB)

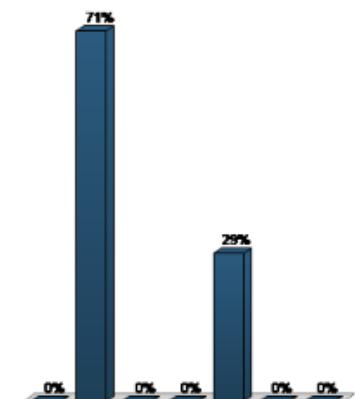
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	33.33%	2
Somewhat Agree	33.33%	2
Neutral	16.67%	1
Somewhat Disagree	16.67%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



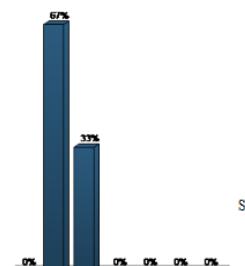
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	71.43%	5
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	28.57%	2
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	7



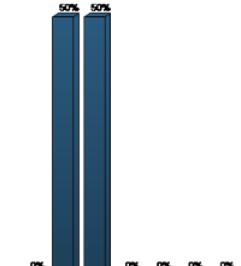
Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	2
Somewhat Agree	33.33%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3



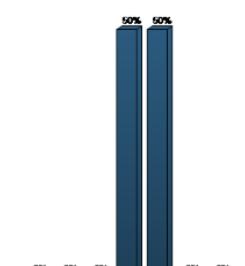
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	50%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

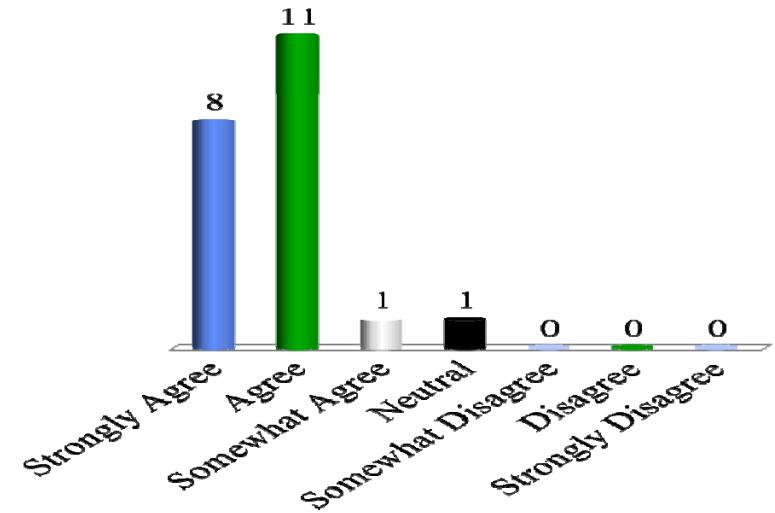
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	0%	0
Neutral	50%	1
Somewhat Disagree	50%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



'Number of bugs caught in testing vs. field use' is a good standard SW measure (DIB)

Aka: Defect profile; Defect Containment Efficiency

1. **Strongly Agree**
2. **Agree**
3. **Somewhat Agree**
4. **Neutral**
5. **Somewhat Disagree**
6. **Disagree**
7. **Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	0	0	0	0
2. Agree	4	4	2	1	1
3. Somewhat Agree	0	1	0	0	1
4. Neutral	1	2	1	1	0
5. Somewhat Disagree	1	0	0	0	0
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

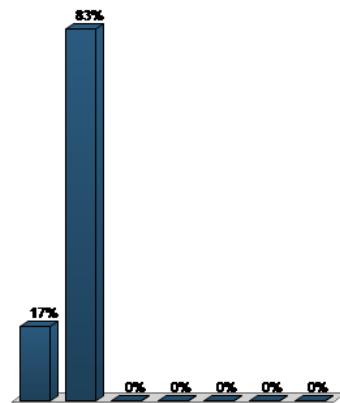
21

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Number of bugs caught in testing vs. field use' is a good standard SW measure (DIB)

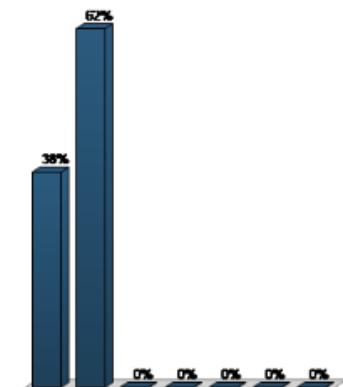
U.S. Government

Responses		
	Percent	Count
Strongly Agree	16.67%	1
Agree	83.33%	5
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



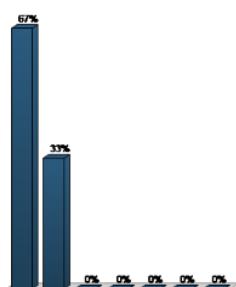
U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	37.5%	3
Agree	62.5%	5
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	8



Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	66.67%	2
Agree	33.33%	1
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3



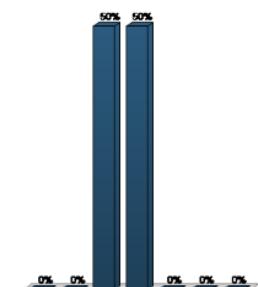
Commercial Industry

Responses		
	Percent	Count
Strongly Agree	100%	2
Agree	0%	0
Somewhat Agree	0%	0
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



Other

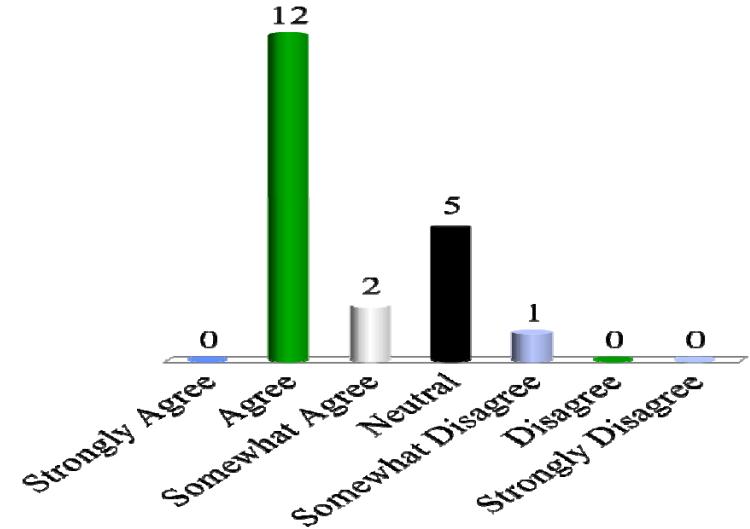
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	0%	0
Somewhat Agree	50%	1
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2



'Change failure rate (rollback deployed code)' is a good standard SW measure (DIB)

Aka: Change failure rate (rollback deployed code)

- 1. Strongly Agree**
- 2. Agree**
- 3. Somewhat Agree**
- 4. Neutral**
- 5. Somewhat Disagree**
- 6. Disagree**
- 7. Strongly Disagree**



	U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Strongly Agree	0	0	0	0	0
2. Agree	4	4	2	1	1
3. Somewhat Agree	0	1	0	0	1
4. Neutral	1	2	1	1	0
5. Somewhat Disagree	1	0	0	0	0
6. Disagree	0	0	0	0	0
7. Strongly Disagree	0	0	0	0	0

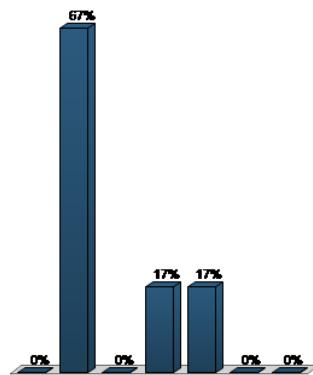
20

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

'Change failure rate (rollback deployed code)' is a good standard SW measure (DIB)

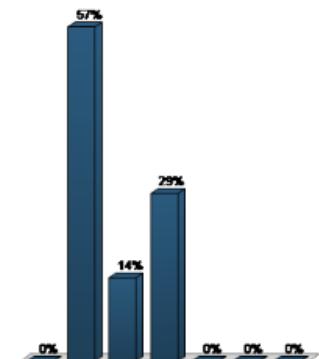
U.S. Government

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	4
Somewhat Agree	0%	0
Neutral	16.67%	1
Somewhat Disagree	16.67%	1
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	6



U.S. Defense Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	57.14%	4
Somewhat Agree	14.29%	1
Neutral	28.57%	2
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	7



Academia / FFRDC

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	66.67%	2
Somewhat Agree	0%	0
Neutral	33.33%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	3

Commercial Industry

Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	0%	0
Neutral	50%	1
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2

Other

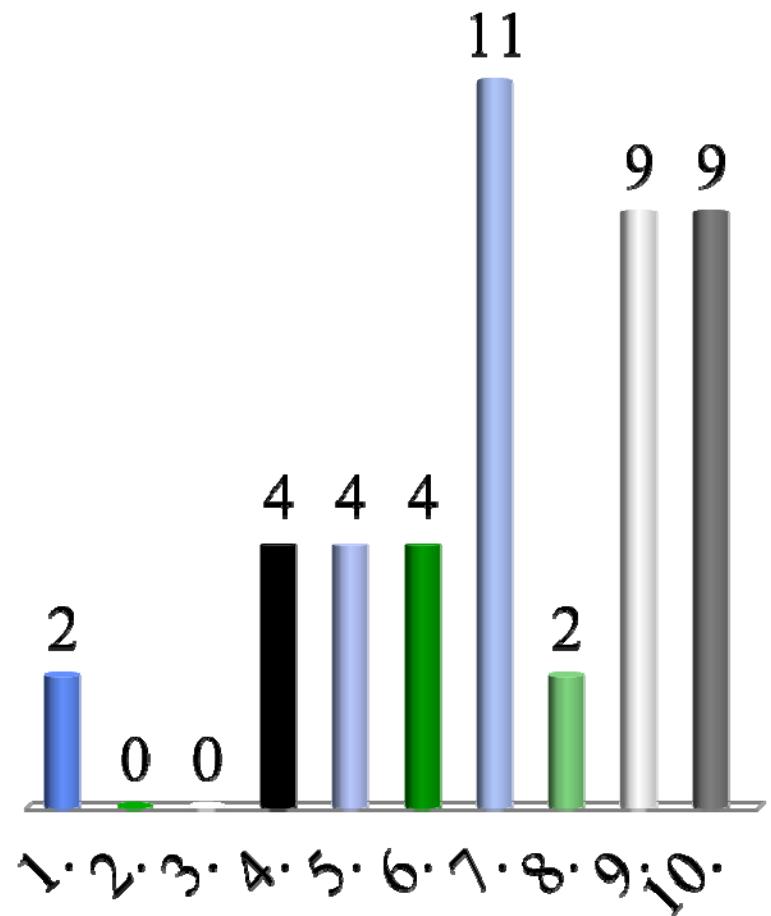
Responses		
	Percent	Count
Strongly Agree	0%	0
Agree	50%	1
Somewhat Agree	50%	1
Neutral	0%	0
Somewhat Disagree	0%	0
Disagree	0%	0
Strongly Disagree	0%	0
Totals	100%	2

Which of the example DIB measures do you commonly use today in your organizations?

Select all that apply.

1. *Time from program launch to deployment of simplest useful functionality*
2. *Time to field high priority functions (spec > ops) or fix new found security hole*
3. *Time from code committed to code in use*
4. *Time req'd for full regression test (automated)*
5. *Time required to restore service after outage*
6. *Automated test coverage of specs / code*
7. *Number bugs found in testing vs field use*
8. *Change failure rate (rollback deployed code)*
9. *Complexity metrics*
10. *Development plan/environment metrics*

15



PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

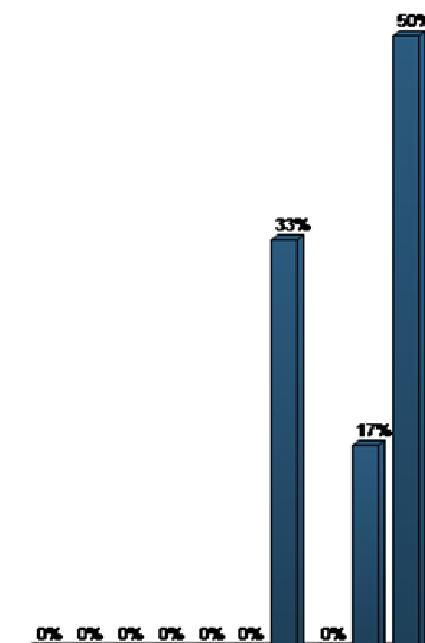
18. Which of the example DIB measures do you commonly use today in your organizations? (Multiple Choice – Multiple Response)

U.S. Government (DoD, agency) 6	U.S. Defense Industry 7	Academia / FFRDC 3	Commercial Industry 2	Other 2
1. Time from program launch to deployment of simplest useful functionality	0	0	1	0
2. Time to field high priority functions (spec > ops) or fix new found security hole	0	0	0	0
3. Time from code committed to code in use	0	0	0	0
4. Time req'd for full regression test (automated)	0	4	0	0
5. Time required to restore service after outage	0	3	1	0
6. Automated test coverage of specs / code	0	3	1	0
7. Number bugs found in testing vs field use	2	7	2	0
8. Change failure rate (rollback deployed code)	0	1	0	1
9. Complexity metrics	1	5	3	0
10. Development plan/environment metrics	3	4	2	0

Which of the example DIB measures do you commonly use today in your organizations?

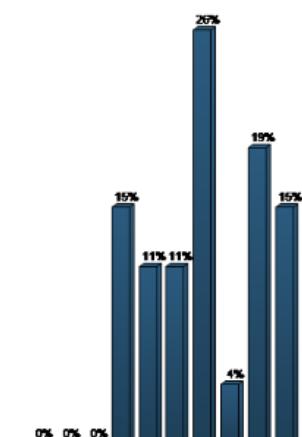
Responses	
Percent	Count
Time from program launch to deployment of simplest useful functionality	0%
Time to field high priority functions (spec > ops) or fix new found security hole	0%
Time from code committed to code in use	0%
Time req'd for full regression test (automated)	0%
Time required to restore service after outage	0%
Automated test coverage of specs / code	0%
Number bugs found in testing vs field use	33.33%
Change failure rate (rollback deployed code)	0%
Complexity metrics	16.67%
Development plan/environment metrics	50%
Totals	100%

U.S. Government



Responses	
Percent	Count
Time from program launch to deployment of simplest useful functionality	0%
Time to field high priority functions (spec > ops) or fix new found security hole	0%
Time from code committed to code in use	0%
Time req'd for full regression test (automated)	14.81%
Time required to restore service after outage	11.11%
Automated test coverage of specs / code	11.11%
Number bugs found in testing vs field use	25.93%
Change failure rate (rollback deployed code)	3.7%
Complexity metrics	18.52%
Development plan/environment metrics	14.81%
Totals	100%

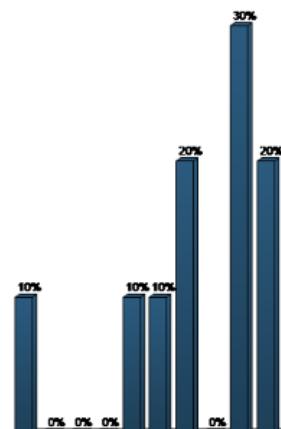
U.S. Defense Industry



Which of the example DIB measures do you commonly use today in your organizations?

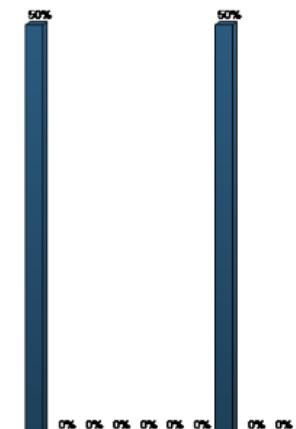
Responses		
	Percent	Count
Time from program launch to deployment of simplest useful functionality	10%	1
Time to field high priority functions (spec > ops) or fix new found security hole	0%	0
Time from code committed to code in use	0%	0
Time req'd for full regression test (automated)	0%	0
Time required to restore service after outage	10%	1
Automated test coverage of specs / code	10%	1
Number bugs found in testing vs field use	20%	2
Change failure rate (rollback deployed code)	0%	0
Complexity metrics	30%	3
Development plan/environment metrics	20%	2
Totals	100%	10

Academia / FFRDC



Responses		
	Percent	Count
Time from program launch to deployment of simplest useful functionality	50%	1
Time to field high priority functions (spec > ops) or fix new found security hole	0%	0
Time from code committed to code in use	0%	0
Time req'd for full regression test (automated)	0%	0
Time required to restore service after outage	0%	0
Automated test coverage of specs / code	0%	0
Number bugs found in testing vs field use	0%	0
Change failure rate (rollback deployed code)	50%	1
Complexity metrics	0%	0
Development plan/environment metrics	0%	0
Totals	100%	2

Other



***General Comments/Feedback on Measures Recommended by
DSB or DIB?***

- ???

What other effective measures for iterative software development should be considered as common or standard measures for acquisition and management of software systems?

- ***Team measures?***
- ***Product measures?***
- ***Enterprise measures?***

Next steps...

- Interested in working further on measures for defense software acquisition and response to DSB task force recommendations?
 - Engage in the NDIA / INCOSE / PSM Iterative Software Development and Acquisition Working Group (ISDAWG)
 - Contacts: Geoff Draper (gdraper@harris.com), Joe Elm (Joseph.Elm@L3T.com), Cheryl Jones (cheryl.l.jones128.civ@mail.mil), Garry Roedler (garry.j.roedler@lmco.com), Robin Yeman (robin.yeman@lmco.com)
- Follow-on panel discussion at NDIA Systems Engineering Conference (October 22-25, Tampa FL)
<http://www.ndia.org/events/2018/10/22/9870---21st-systems-engineering-conference>
- Support development of ICM tables for DSB/DIB measures or derived measures for recommendations to DoD

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

NDIA Systems Engineering Conference, October 2018

Abstract: A Path Toward Consensus Measures for Iterative Software Development

Presenter: NDIA/INCOSE/PSM Iterative Software Development & Acquisition Working Group

Summary:

Traditional measures for managing software development are not keeping pace with industry trends toward iterative development, agile, and DevOps. Modernization of measurement frameworks has been recommended by independent DoD studies and reports. NDIA, INCOSE, and PSM are collaborating on potential solutions.

Text:

Traditional measures used to plan and manage software programs based largely on waterfall development and software lines of code-based estimates are not keeping pace with trends in the defense industry toward methods based in a software factory environment including automated testing, continuous integration, and rapid iterative development and deployment of new capabilities. A recent report from the Defense Science Board Task Force on Software Design and Acquisition recommends modernizing measurement frameworks and utilizing historical data in areas such as schedule, cost, productivity, and staffing. The Defense Innovation Board similarly considers many current software metrics obsolete and proposed new metrics for DoD/industry software development based on DevOps methods and focused on deployment rate, response rate, code quality, and modern development plan attributes.

NDIA, INCOSE, and PSM are collaborating to consider measures such as these to develop consensus recommendations to DoD for improved measures in software design, development, testing, and acquisition. Members of the joint Iterative Software Development & Acquisition Working Group have initiated workshops through the Practical Software and Systems Measurement (PSM) User's Group conference and will provide a summary of work in progress towards an improved measurement set adapted to modern software development practices, such as DevOps. The working group is soliciting input from across the defense community.

Workshop Report Out

***Measures for Iterative Software
Development and Acquisition***

Workshop Participants

Geoff Draper (Harris)	Bob Stoddard (SEI)
Cheryl Jones (U.S. Army / PSM)	Joe Dean
Garry Roedler (Lockheed Martin)	Roz Singh (Raytheon)
Mark Cornwell (ODASD SE)	Tony Powell
Gary Palosaari (Aerospace)	Kevin McBride (Lockheed Martin)
Lauren Lindros (Aerospace)	Steve Verga (Harris)
Bernard Reger (DoD)	Jason McDonald (Harris)
Paul Janusz (DoD)	Robin Yeman (Lockheed Martin)
Julie DiStefano (Tecolote)	...
David Becker (Tecolote)	

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Iterative SW Development Measures – PSM User's Conference Workshop

Inputs

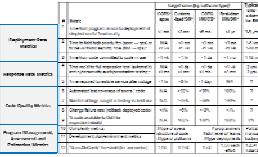
DSB SW Task Force Report



Dr. LaPlante PSM Keynote



Defense Innovation Board (DIB) Measures



Draft NDIA Goals, Questions (GQM) and 'Industry Space' Measures

Measure Type	Industry Space		GQM	
	Definition	Description	Goal	Question
Deployment Rate	Deployment Rate	Deployment Rate	How quickly can we deliver initial capability for new products?	How quickly can we add and deliver high priority capabilities for an existing operational product?
	Automated test coverage (unit level, system level)	Automated test coverage (unit level, system level)	How quickly can new security vulnerabilities be patched and deployed to fielded products?	What is the "lead time" duration from code committed to a repository to availability of code?
	Automated test coverage (unit level, system level)	Automated test coverage (unit level, system level)	What is the lead time to a product out? (TBO)	What is the lead time to a product out? (TBO)
Response Rate	Quickly detect, isolate, and remove software operational defects.	Quickly detect, isolate, and remove software operational defects.	How quickly can a full automated regression test be conducted to verify capability correctness?	How quickly can a full automated regression test be conducted to verify capability correctness?
	Automated test coverage (unit level, system level)	Automated test coverage (unit level, system level)	Can a complete cybersecurity test be successfully completed to ensure adequate resistance to vulnerabilities?	What is the reliability and availability of operational service capabilities?
	Automated test coverage (unit level, system level)	Automated test coverage (unit level, system level)	Consider other TPMs, MODS, KPIs	Consider other TPMs, MODS, KPIs
Code Quality	Ensure high quality code in development and in the field.	Ensure high quality code in development and in the field.	Is testing efficiency (e.g., automation) aligned relative to project plans? How does new code functionally work and expected and not break previous functionality?	Is testing efficiency (e.g., automation) aligned relative to project plans? How does new code functionally work and expected and not break previous functionality?
	Code quality definitions in development and in the field	Code quality definitions in development and in the field	Does new code update reliable in the field such that operational service is not impacted?	What is the quality of code deployed to the field?
	Code quality definitions in development and in the field	Code quality definitions in development and in the field	Does DOD have data rights to sustain the code baseline?	Does DOD have data rights to sustain the code baseline?
Program Management, Assessment, and Estimation	Achieve effective insight into management of SW programs, including cost assessment and performance estimation.	Achieve effective insight into management of SW programs, including cost assessment and performance estimation.	Does new code work as expected and maintainable?	How much of the code base is newly developed vs. reused from other sources?
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	How much capability has been delivered for each release?	How much capability has been delivered for each release?
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)	Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	How stable are the mission capability requirements?	How stable are the mission capability requirements?
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	What are appropriate thresholds for SW measures that indicate serious relative to minor performance issues?	What are appropriate thresholds for SW measures that indicate serious relative to minor performance issues?
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Is the integration and test progress proceeding as planned? (TBO)	Is the integration and test progress proceeding as planned? (TBO)
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Are the capabilities and features being implemented and completed as planned? (TBO)	Are the capabilities and features being implemented and completed as planned? (TBO)

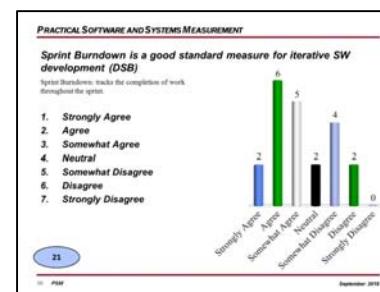
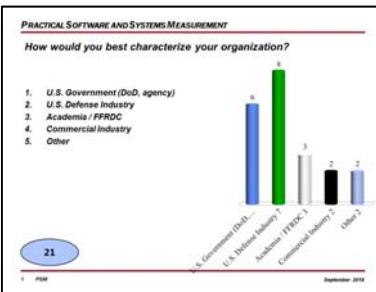
PSM Workshop

- Initial Validation of Categories, Goals, Questions
- Prioritization and Ranking

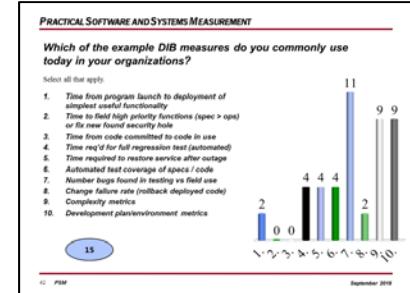
Category (DIB)	Goal (Derived from DIB by PSM/NDIA)	Questions (Derived from DIB by PSM/NDIA)	SW	H	M	L	Total
Deployment Rate	Prioritize speed in delivering value to end users through new operations capabilities.	How quickly can we deliver initial capability for new products?	10	1	1	1	10
	Automated test coverage (unit level, system level)	How quickly can we add and deliver high priority capabilities for an existing operational product?	12	1	1	1	15
	Automated test coverage (unit level, system level)	How quickly can new security vulnerabilities be patched and deployed to fielded products?	10	1	1	1	10
	Automated test coverage (unit level, system level)	What is the "lead time" duration from code committed to a repository to availability of code?	1	1	1	1	3
	Automated test coverage (unit level, system level)	What is the lead time to a product out? (TBO)	1	1	1	1	3
Response Rate	Quickly detect, isolate, and remove software operational defects.	How quickly can a full automated regression test be conducted to verify capability correctness?	9	1	1	1	10
	Automated test coverage (unit level, system level)	Can a complete cybersecurity test be successfully completed to ensure adequate resistance to vulnerabilities?	7	1	1	1	9
	Automated test coverage (unit level, system level)	What is the reliability and availability of operational service capabilities?	1	1	1	1	3
	Automated test coverage (unit level, system level)	Consider other TPMs, MODS, KPIs	10	1	1	1	20
	Automated test coverage (unit level, system level)	How quickly can we address bug reports from the field? (TBO)	1	1	1	1	3
Code Quality	Ensure high quality code in development and in the field.	Is testing efficiency (e.g., automation) aligned relative to project plans? How does new code functionally work and expected and not break previous functionality?	6	1	1	1	10
	Code quality definitions in development and in the field	Does new code functionality work as expected and not break previous functionality?	10	1	1	1	10
	Code quality definitions in development and in the field	Does new code update reliable in the field such that operational service is not impacted?	5	1	1	1	20
	Code quality definitions in development and in the field	Does DOD have data rights to sustain the code baseline?	1	1	1	1	3
	Code quality definitions in development and in the field	Is the quality of code deployed to the field?	1	1	1	1	3
	Code quality definitions in development and in the field	Does new code work as expected and maintainable?	1	1	1	1	3
	Code quality definitions in development and in the field	How much of the code base is newly developed vs. reused from other sources?	0	1	1	1	4
	Code quality definitions in development and in the field	How much capability has been delivered for each release?	10	1	1	1	10
	Code quality definitions in development and in the field	Are all resources available to execute the SW development plan? (staff, skills, tools, facilities)	6	1	1	1	10
	Code quality definitions in development and in the field	How stable are the mission capability requirements?	0	1	1	1	3
	Code quality definitions in development and in the field	What are appropriate thresholds for SW measures that indicate serious relative to minor performance issues?	0	1	1	1	3
	Code quality definitions in development and in the field	Is the integration and test progress proceeding as planned? (TBO)	0	1	1	1	3
	Code quality definitions in development and in the field	Are the capabilities and features being implemented and completed as planned? (TBO)	0	1	1	1	3
Program Management, Assessment, and Estimation	Achieve effective insight into management of SW programs, including cost assessment and performance estimation.	Is the integration and test progress proceeding as planned? (TBO)	0	1	1	1	3
	Activity descriptions that should be required to provide insight into the planning and producing time, cost, and quality.	Are the capabilities and features being implemented and completed as planned? (TBO)	0	1	1	1	3

Interactive Scoring of Draft DSB/DIB Measures

Brainstorming on Additional Potential Measures

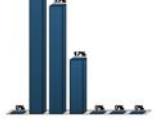
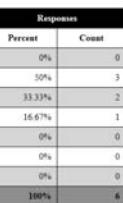


- Sampling of Organizational Adoption/Usage of Proposed Measures

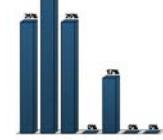
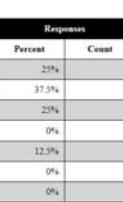


- Data Slicing by Demographic

U.S. Government



U.S. Defense Industry



Academia / FFRDC
Commercial Industry
Other

Recommended SW Measures from Defense Innovation Board (DIB)

#	Metric	Target value (by software type) ⁱ				Typical DoD values for SW
		COTS ⁱⁱ apps	Custom-sized SW ⁱⁱⁱ	COTS HW/OS ^{iv}	Real-time HW/SW ^v	
Deployment Rate Metrics	1 Time from program launch to deployment of simplest useful functionality	<1 mo	<3 mo	<6 mo	<1 yr	3-5 yrs
	2 Time to field high priority fcn (spec → ops) or fix newly found security hole (find → ops) ^{vi}	N/A <1 wk	<1 mo <1 wk	<3 mo <1 wk	<3 mo <1 wk	1-5 yrs 1-18 m
	3 Time from code committed to code in use	<1 wk	<1 hr	<1 da	<1 mo	1-18 m
Response Rate Metrics	4 Time req'd for full regression test (automat'd) and cybersecurity audit/penetration testing ^{vii}	N/A <1 mo	<1 da <1 mo	<1 da <1 mo	<1 wk <3 mo	2 yrs 2 yrs
	5 Time required to restore service after outage	<1 hr	<6 hr	<1 day	N/A	?
	6 Automated test coverage of specs / code	N/A	>90%	>90%	100%	?
Code Quality Metrics	7 Number of bugs caught in testing vs field use	N/A	>75%	>75%	>90%	?
	8 Change failure rate (rollback deployed code)	<1%	<5%	<10%	<1%	?
	9 % code available to DoD for inspection/rebuild	N/A	100%	100%	100%	0%
Program Management, Assessment and Estimation Metrics	10 Complexity metrics	#/type of specs structure of code #/type of platforms			# programmers #/skill level of teams #/type deployments	Partial/ manual tracking
	11 Development plan/environment metrics					
	12 "Nunn-McCurdy" threshold (for any metric)	1.1X	1.25X	1.5X	1.5X each effort	1.25X Total \$

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Validation and Ranking of DIB Measures > Draft PSM/NDIA-Derived Categories, Goals, Questions to Initiate a Program Measurement Framework

Category (DIB)	Goal (Derived from DIB by PSM/NDIA)	Questions (derived from DIB by PSM/NDIA; see GQM table for details)	VH	H	M	L	Total
Deployment Rate	Prioritize speed in delivering value to end users through new operational capabilities. -Automated development and deployment -Automated testing (unit level, system level) -Iterative deliver-value-now mentality	How quickly can we deliver initial capability for new products?	10	9			19
		How quickly can we add and deliver high priority capabilities for an existing operational product?	12	7			19
		How quickly can new security vulnerabilities be patched and deployed to fielded products?	16	3			19
		What is the "lead time" duration from code committed to a repository to availability of tested functionality?	3	4	12		19
		What is the cycle time to get a product out? (TBD)					0
Response Rate	Quickly detect, isolate, and remove software operational defects. How gracefully SW fails, how many errors are caught and resolved in testing, and how rapidly developers patch bugs are excellent measures of SW development prowess.	How quickly can a full automated regression test be conducted to verify capability correctness?	9	7	3		19
		How quickly can a complete cybersecurity test be successfully completed to ensure adequate resistance to vulnerabilities?	7	9	3		19
		What is the reliability and availability of operational service capabilities?	10	7	3		20
		How quickly can we address bug reports from the field? (TBD)					0
Code Quality	Ensure high quality code in development and in the field. Find and remove defects in developmental testing where they are most cost-efficient to fix.	Is testing efficiency (e.g., automation, coverage) appropriate relative to project plans? How long does it take to conduct testing within planned constraints?	6	6	6	1	19
		Does new code functionality work as expected and not break previous functionality?	16	2	1		19
		What is the quality of code deployed to the field?	10	9			19
		Are product baseline updates reliable in the field such that operational service is not impacted?	5	8	6	1	20
		Does DoD have data rights to sustain the code baseline?	1	2	13	2	18
Program Management, Assessment, and Estimation	Achieve effective insight into management of SW programs, including cost assessment and performance estimation. These metrics describe a list of 'features' (performance metric, contract terms, project plans, activity descriptions) that should be required to provide better tools for monitoring and predicting time, cost, and quality.	Is the code and development platform well structured and maintainable?	7	10	1		18
		How much of the code base is newly developed vs. reused from other sources?	0	5	9	4	18
		How much capability has been delivered for each release?	10	7	1	1	19
		Are sufficient resources available to execute the SW development plan? (staff, skills, tools, suppliers)	6	8	4		18
		How stable are the mission capability requirements?	4	9	4		17
		What are appropriate thresholds for SW measures that indicate serious breach relative to baseline plans such that re-evaluation is needed? (TBD)					0
		Is the integration and test progress proceeding as planned? (TBD)					0
		Are the capabilities and features being implemented and completed as planned? (TBD)					0

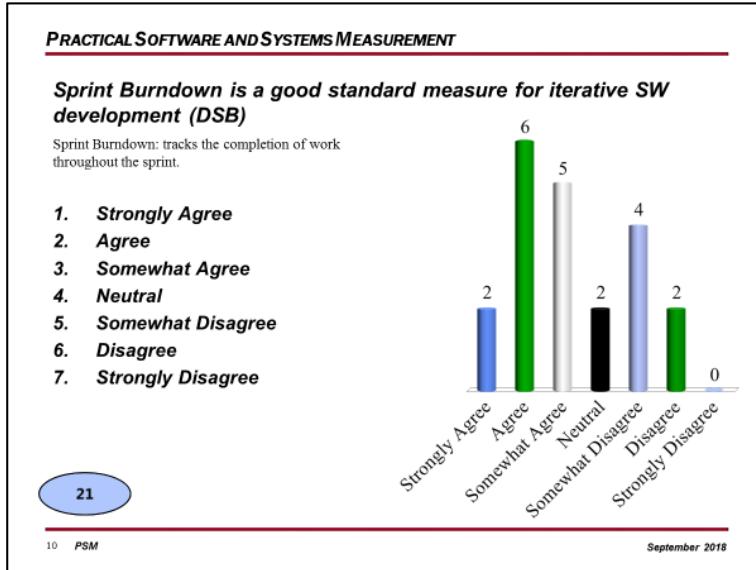
See PSM workshop spreadsheet for details and mappings to candidate measures

PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Example Evaluation and Scoring of DSB/DIB Measures

Scoring by Participant Demographic

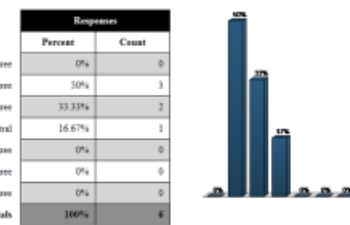
Overall PSM Workshop Participant Scoring



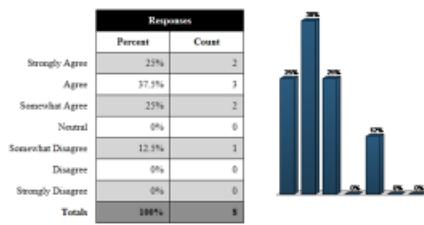
PRACTICAL SOFTWARE AND SYSTEMS MEASUREMENT

Sprint Burndown is a good standard measure for iterative SW development (DSB)

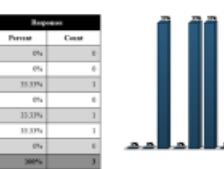
U.S. Government



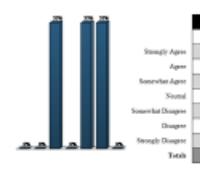
U.S. Defense Industry



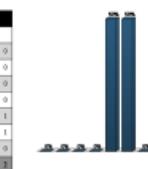
Academia / FFRDC



Commercial Industry



Other

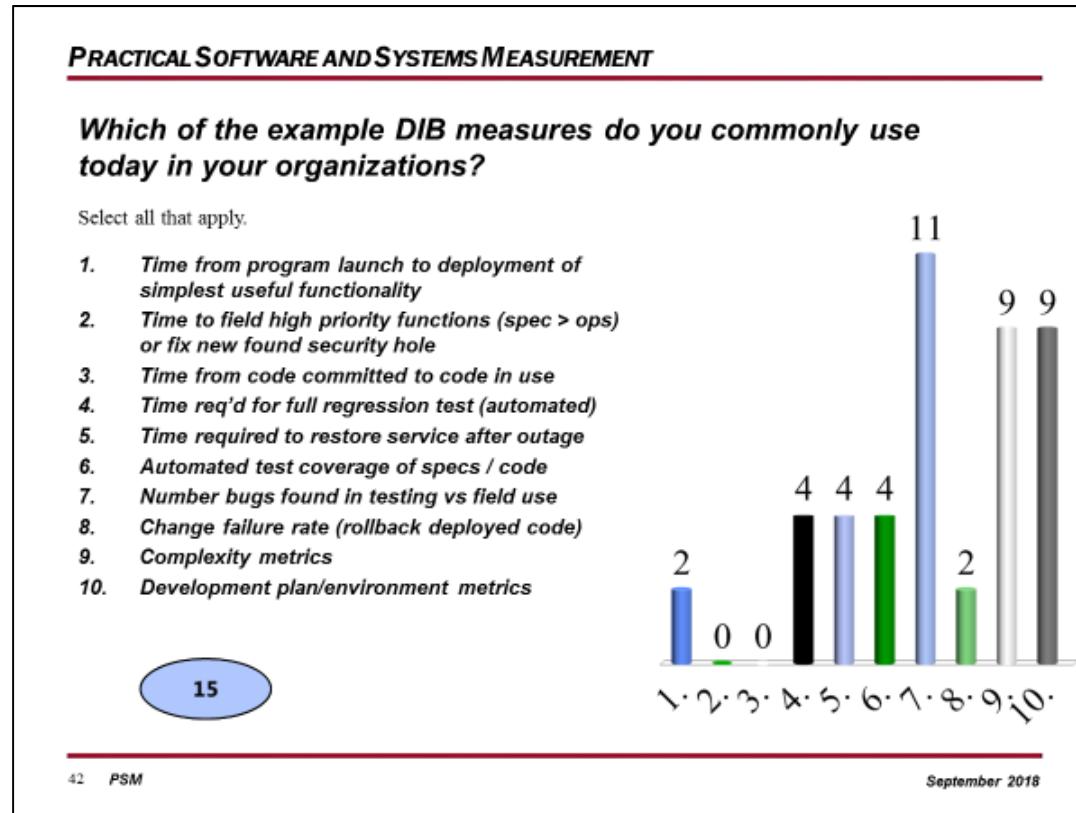


11 PSM

September 2018

PSM workshop participants were asked to interactively score measures suggested by DSB and DIB using a Likert scale

Sampling of current adoption and usage of proposed DIB metrics



As anticipated by the DSB report, we can likely expect some learning curve for DoD and the defense industry to transition into consistent widespread adoption of iterative SW development / DevOps processes and measures

Summary Scoring of DSB/DIB Candidate SW Measures

Question: [Measure name] is a good standard measure for iterative SW development.

	Points:	10	8	6	5	4	2	0				
	Candidate or Example SW Metrics Proposed by DSB and DIB:	Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree	Total Voters	Weighted Score	In Common Use?	Usage %
DSB	Good standard SW measure?											
	Sprint Burndown	2	6	5	2	4	2		21	6.10	8	38%
	Epic and Release Burndown	4	8	6	3				21	7.38	8	38%
	Velocity	2	6	9	2	1	1		21	6.57	6	29%
	Cycle Time (Control Chart)	1	4	9	2	2	2		20	5.90	2	10%
DIB	Cumulative Flow Diagram	1	9	7	2	1			20	5.35	1	5%
	Time from program launch to deployment of simplest useful functionality	3	7	6	3				19	7.21	2	11%
	Time to field high priority functions (spec > ops)	3	9	6	1				19	7.53	0	0%
	Time to fix newly found security hole (find > ops)	6	9	3	2		1		21	7.71	4	19%
	Time from code committed to code in use	7	6	2	3	2			20	5.90	4	20%
	Time required for full regression test (automated)	1	8	5	4	1	1		20	6.50	4	20%
	Time required to restore service after outage	5	5	4	1	1			16	7.69	11	69%
	Automated test coverage of test specs / code		10	4	2	4			20	6.50	2	10%
	Number of bugs caught in testing vs. field use	8	11	1	1				21	8.52	9	43%
	Change failure rate (rollback deployed code)		12	2	5	1			20	6.85	9	45%

PSM workshop participants generally found value in many of the candidate DSB / DIB measures – with suggestions for additional follow-on consideration

Next Steps/Action Items

- *Continue analysis of measures collected from PSM workshop*
- *Develop PSM ICM table and other products as a foundation for a proposed SW measurement framework (DSB recommendation #3)*
 - *Adapt or integrate with PSM draft ICM table for agile development (see other PSM workshop results)*
- *Engage in joint NDIA/INCOSE/PSM Iterative Software Development and Acquisition Working Group (ISDAWG)*
 - *Encourage PSM leadership in addressing measurement-related findings*
 - *Consider attending (or leading) follow-on workshop at NDIA Systems Engineering Conference (Oct 22-25, Tampa FL)
<http://www.ndia.org/events/2018/10/22/9870---21st-systems-engineering-conference>*
 - *Participate in development or review of suggestions to DoD for implementation of DSB recommendations
(products, source selection criteria, guidance, measures, education, ...)*