Delivering Systems That Work: Institutionalizing Performance Measures across the DoD Acquisition Life Cycle

Mr. Sean Brady
Office of the Deputy Assistant Secretary of Defense for Systems Engineering

83rd MORS Symposium
June 22-25, 2015
Overview

- PMO and OSD Decision Analysis Support: why do we measure?
  - Data-driven decisions at every level (IPT Lead, Chief SE, PM, PEO, DAE, Congress)
  - Statue, Policy and Guidance

- DASD(SE)’s Development of a Metrics Assessment Framework
  - toward consistent performance measurement across the DoD Acquisition System
  - toward improved decision making

- Impacting Decisions: DASD(SE) Metrics Efforts
  - Schedule Realism and System Maturity
  - Knowledge Point: Performance Measures for OT Readiness
  - Agile Metrics
  - Enterprise Benchmarking
  - Software Parametric Statistical Analysis

- Conclusions & Future Goals
DASD, Systems Engineering

Stephen Welby
Principal Deputy Kristen Baldwin

Major Program Support
James Thompson

Supporting USD(AT&L) Decisions with Independent Engineering Expertise

- Engineering Assessment / Mentoring of Major Defense Programs
- Program Support Assessments
- Overarching Integrated Product Team and Defense Acquisition Board Support
- Systems Engineering Plans
- Systemic Root Cause Analysis
- Development Planning/Early SE
- Program Protection

Engineering Enterprise
Robert Gold

Leading Systems Engineering Practice in DoD and Industry

- Systems Engineering Policy and Guidance
- Technical Workforce Development
- Specialty Engineering (System Safety, Reliability and Maintainability, Quality, Manufacturing, Producibility, Human Systems Integration)
- Security, Anti-Tamper, Counterfeit Prevention
- Standardization
- Engineering Tools and Environments

Providing technical support and systems engineering leadership and oversight to USD(AT&L) in support of planned and ongoing acquisition programs
Performance Measurement is Crucial to DoD

- **The DoD engineering challenge:** acquire, within budget and schedule, the beyond-cutting-edge systems that will provide U.S. warfighters overwhelming superiority in the full spectrum of conflicts

- DoD develops & delivers incredibly effective but increasingly complex weapon systems to the Warfighter

- **Increasing complexity** of Warfighting capability demands more effective life cycle SE and quantitative insight

- **DoD’s senior leadership depends on data** to drive multi-billion dollar decisions which impact warfighting capability

- **DASD(SE) is committed to using a quantitative SE approach to**
  - mentor major PMOs and system developers; shape program plans; monitor execution
  - inform DoD leadership of technical risks, opportunities, and impacts to schedule & performance at major decisions
  - reduce time-to-Warfighter and cost for System and Software acquisition
Knowledge Points

– What information is relevant at decisions maker engagement points?

– DoDI 5000.02 provides some key knowledge points including acq milestones, decisions points, and SE Technical Reviews

Inflection Points

– What changes in metrics alert decision makers to emerging problems?

Both are relevant in the context of:

• Knowledge sought
• Decisions to be made
• Data / documentation available

Represented as Goals, associated Questions, and performance Measures (GQM)

MPS supports data-driven decisions to reduce risk.
Why do we measure?

• **Weapon Systems Acquisition Reform Act (2009)**
  - S.454-10; d.(1): The development and tracking of detailed measurable performance criteria as part of the systems engineering master plans...

• **Systems Engineering Plan (SEP) Outline, 20 April 2011**
  - Directs programs to present their strategy for identifying, prioritizing, and selecting metrics for monitoring and tracking program SE activities and performance
  - **Sect 3.6. - “Technical Performance Measures and Metrics”**
    - Provides an overview of measurement planning and metrics selection process
    - Include approach to monitor execution-to-plan and identification of roles, responsibilities, and authorities
    - Minimum set of TPMs and intermediate goals and plan to achieve them with dates
    - Examples include TPMs in areas of software, reliability, manufacturing, integration, and test

Performance measures are foundational to DASD(SE)’s mission.
Performance Measurement Shortfalls & DASD(SE) Initiatives

• **Background**
  – By statute, DASD(SE) is responsible for “the development and tracking of detailed measureable performance criteria as part of the [Systems Engineering Plans]”
    – Performance measurement over a system’s lifecycle is at the core of our mission

• **Performance Measurement Shortfalls**
  – DASD(SE) identified systemic issues in Annual Report to Congress and various forums (e.g., NDIA, MORS)
  – Lack of sufficient **predictive metrics** and **quantitative management**
    – DASD(SE) mentoring:
      – Introduced programs to defect prediction and measurement techniques as a means to plan, measure and control software quality, and assess maturity
  – Lack of **end-to-end performance measurement**, **developer/tester disconnect** and **insufficient integration testing**
    – DASD(SE) mentoring:
      – MPS developed a framework that aids programs’ identification of effective measures to track;
      – Helped programs develop the right performance measures
Enterprise-level Insights & Trends:
-- Measurement Categories in Practice --
Community vs. SEP Samples (FY14 vs. FY15)

<table>
<thead>
<tr>
<th>Community Sources:</th>
<th>FY14 SEPs:</th>
<th>FY15 SEPs:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Staffing</td>
<td>5. Software</td>
<td>5. Integration</td>
</tr>
</tbody>
</table>

Differences between Community recommended measures and “real world” measures in SEPs may be driven by law, DoD policy & guidance (e.g., reliability, NR KPP).

*Community: Govt (P&G, reports, training) / Industry (reports, standards) / Academia (papers)

Developing the right measures and tracking performance is a challenge in DoD programs. Identifying high-level trends across portfolio and enabling improved guidance.
**Performance Measures Study**

Approach to Develop & Maintain a Practical Assessment Framework

---

**Lean Six Sigma project is focusing on improving metrics in SEPs.**

1. What metrics categories exist in DoD, Academia & Industry literature?

2. Created an Analysis Framework to evaluate SEPs’ performance measures.

3. Applied framework to select programs & study.

4. Identified high-level trends across portfolio and recommended guidance.

---

**Compare/contrast Literature TPMs to Program SEPs’ TPMs**

---

**Attributes of Good TPMs**

- **Relevance:** Only select measures that do not have numerous interpretations and that are pertinent to an end result you are trying to obtain.
- **Completeness:** Be sure you identify a balanced set of measures and that your emphasis does not become skewed.
- **Timeliness:** Be sure collection and analysis will provide the needed information in time to allow corrective action to be initiated.
- **Simplicity:** Keep it as simple and logical as possible. The measures should be easy to collect, analyze, and understand.
- **Cost Effectiveness:** Use data that is economical to collect. Use organizational or customer required data to address other program issues, where applicable. Leverage data collected for current management practices.
- **Repeatability:** This is important for comparing measures across projects.
- **Accuracy:** Make sure that your measures are accurate and the resulting analysis accurately serves the intended purpose of the measure.

---

**Overall Program Measure Assessment**

<table>
<thead>
<tr>
<th>Category</th>
<th>Specific</th>
<th>Measurable</th>
<th>Actionable</th>
<th>Repeatable</th>
<th>Timeliness</th>
<th>Relevant</th>
<th>Cost Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design/development</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements Management</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Management</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schedule</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staffing</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Assurance</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Quality</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology Maturity</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mission Performance</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Ready KPP</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supportability - Maintainability</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Performance</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Acceptance</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Total # of TPMs**: 16

---

**Program Assessment Framework**

Framework = categories + criteria

---

**SMART DAU Criteria**

- **Attributes of Good TPMs**
  - **Relevance**: Only select measures that do not have numerous interpretations and that are pertinent to an end result you are trying to obtain.
  - **Completeness**: Be sure you identify a balanced set of measures and that your emphasis does not become skewed.
  - **Timeliness**: Be sure collection and analysis will provide the needed information in time to allow corrective action to be initiated.
  - **Simplicity**: Keep it as simple and logical as possible. The measures should be easy to collect, analyze, and understand.
  - **Cost Effectiveness**: Use data that is economical to collect. Use organizational or customer required data to address other program issues, where applicable. Leverage data collected for current management practices.
  - **Repeatability**: This is important for comparing measures across projects.
  - **Accuracy**: Make sure that your measures are accurate and the resulting analysis accurately serves the intended purpose of the measure.
Study Outcomes & Framework Outputs “Inside the Black Box”

High-level analysis and trends

30+ SEPs with over 1,000 metrics

Detailed-level

Practical application

Not enough TPMs; No threshold / objective values; Measuring too late; Limited ability to influence program; Expensive to collect

No mission performance metrics; exclusively focused on “Product” measures; NR KPP unmeasurable
SEP Performance Measures
Assessment Framework Purpose & Effect

Measures Assessment Framework

21 Categories

MPS tool assesses 2 basic questions:
1. Does the program have measures in the appropriate categories?
2. How good are the measures the program selected?

Attributes of Good TPMs

● Relevance: Only select measures that do not have numerous interpretations and that are pertinent to an end result you are trying to obtain.
● Completeness: Be sure you identify a balanced set of measures and that your emphasis does not become skewed.
● Timeliness: Be sure collection and analysis will provide the needed information in time to allow corrective action to be initiated.
● Simplicity: Keep it as simple and logical as possible. The measures should be easy to collect, analyze, and understand.
● Cost Effectiveness: Use data that is economical to collect. Use organizational or customer required data to address other program issues, where applicable. Leverage data collected for current management practices.
● Repeatability: This is important for comparing measures across projects.
● Accuracy: Make sure that your measures are accurate and the resulting analysis accurately serves the intended purpose of the measure.

Top-level criteria

Program reports TPMs IAW SEP? [Y / N & report frequency]
Source document (CDRL or report name)
Entered in MPS dB as of DATE
Program is using TPMs as part of Risk Management process? SEP includes TPM status?
Do TPMs show progress against a plan? [Y / N]
Does a risk w/mitigation plan exist for under performing TPMs? [Y / N]

SEP TPM Table Compliance

SEP TPM section exists? [Y / N]
TPMs include product and process measures? [Y / N]
Are TPMs devoid of qualitative or existence (yes / no) metrics? [Y / N]
Compliance with SEP Guidance mandated table: (% of TPMs)

Owner defined
Requirements reference
Target value(s) (threshold/objective)
Plan to achieve over time

Before

After (for SEP Signature)

In FY14, MPS Performance Measure assessment tool assisted 20 programs in improving their individual measurement plans.
Performance Measures
Recommended Measurement Categories for DoD Acquisition Programs

**Quantitative Process Measures**

**Software**
- Demographics
- Effort
- Productivity, Agile Velocity
- Schedule
- Staff
- Test

**Staffing**
- Quantity
- Effort Hours
- Experience
- Turnover Rate

**Schedule**
- Technology Maturity

**Requirements Management**
- % DoDAF drawings complete
- Quality Attributes
- Flexibility, Stability

**Cost**
- Affordability
- Resources
- Dollars/Funding
- CPI

**Risk Management**
- Exposure
- Burndown

**Production**
- Build-to-Package Completions
- Traveled Work
- Supplier/Sub Quality Tests
- Scrap, Rework and Repair Hours
- Touch Labor Hours
- Yield

**Design/Development**
- Architecture
- COTS/GOTS/NDI Components
- Interface Definition
- Interface Verification
- Interface Stability

**Manufacturing**
- System Assurance
- Infrastructure

**System Performance**
- Accuracy
- Lethality
- Bandwidth
- System Latency
- System Throughput
- System Response Time
- Utilization—Data bus, CPU, Memory
- SWAP-C
- Range

**Integration**
- COTS/GOTS/NDI Components
- Interface Definition
- Interface Verification
- Interface Stability

**System Quality**
- Reliability
  - # unscheduled reboots
  - Time between reboots (MTBCF)
- Time to reboot (MTRCF)
- MTBF, MTTF

**System Assurance**
- System Quality
- User Acceptance
  - User questionnaire scores
- User acceptance scores

**Supportability/Maintainability**
- Maintainability Characteristics
- Mean time to repair

**Legend**
- MDAP-centric
- Included on SRDR

* Staffing, Quality & Schedule are also included in the Software Category.
“Brady Matrix”
A Key Output of Measure Assessment Framework

Metrics Assessment Analysis
September version of SEP Rev 1.1, April 2014

Overall Program Measure Assessment
SL Integration / Infrastructure: to be addressed in SEP update post contract award
Software Requirements Stability TPMS will be addressed in SEP update post contract award
Contractor IIS and Risks to be addressed in SEP update post contract award
Contractor Software TPMS will be addressed

Practical framework highlights measurement risk areas & tester-developer disconnect
## SEP Coverage of Metrics Categories
### FY14 & 15

<table>
<thead>
<tr>
<th>Category</th>
<th>Quantitative Process Measures</th>
<th>Technical Performance Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Software</td>
<td>cost</td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 SEP 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 SEP 14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Opportunity to increase visibility into development progress**

### Quantitative Process Management
How far have you progressed in developing the product? (e.g. design, cost, schedule)

### Technical Performance Measures
How well does your product do what it is supposed to do? (e.g. throughput, CPU/memory use)
Metrics Distribution in FY14 SEPs
FY14 Software Focus Area Summary

Distribution of Gathered Metrics

Cost Metrics:
- Domain: Cost
- Systems: DBS, Ship, Space, UAS

Net Ready KPP Metrics:
- Domain: Net Ready KPP
- Systems: DBS, Ship, Space, UAS

Reliability Metrics:
- Domain: Reliability
- Systems: DBS, Ship, Space, UAS

Mission Performance Metrics:
- Domain: Mission Performance
- Systems: DBS, Ship, Space, UAS

System Quality Metrics:
- Domain: System Quality
- Systems: DBS, Ship, Space, UAS

Software Metrics:
- Domain: Software
- Systems: DBS, Ship, Space, UAS

Color shows category and proportional size shows gathered metrics.
What if? MPS aids decisions, helps establish realistic schedules

### Estimates (yellow, red, green) of T&E Build overlaid on program schedule

1. Balanced Probabilities / 11. Same + DT
2. Solve for Size
3. Solve for Schedule
4. Code Growth
5. Unconstrained
6. Highest PI
7. AT&L PI / 8. AT&L PI w/ Code Growth
9. Industry Avionics PI
10. Industry Avionics PI w/ Code Growth

#### Scenario Prob%*

<table>
<thead>
<tr>
<th>Prime’s Plan</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Balanced Probabilities</td>
<td>5%</td>
</tr>
<tr>
<td>2. Solve for Size</td>
<td>90%</td>
</tr>
<tr>
<td>3. Solve for Schedule</td>
<td>50%</td>
</tr>
<tr>
<td>4. Code Growth</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>5. Unconstrained</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>6. Highest Productivity</td>
<td>&gt;99%</td>
</tr>
<tr>
<td>7. AT&amp;L (Rotary/UAS) PI</td>
<td>&gt;99%</td>
</tr>
<tr>
<td>8. AT&amp;L w/ Code Growth</td>
<td>99%</td>
</tr>
<tr>
<td>9. Industry (Avionics) PI</td>
<td>85%</td>
</tr>
<tr>
<td>10. Industry PI Code Growth</td>
<td>50%</td>
</tr>
</tbody>
</table>

---

**Extending project to remove defects:** ~37 months predicted to ensure SW is mature for T&E

**SW quality must be used as early-warning system for release management & Test Readiness**
System Maturity Analysis

- Defects are not closing at a rate that will ensure no CAT Is by entry into System Acceptance Test (SAT) per contractor’s plan
  - Contractor defect metric is currently yellow/red; re-planning to “get well”
- OSD defect projection exceeds contractor’s capacity to fix by SAT and indicates risk of delivering immature software to OT&E
- Recommendations
  - PMO open a risk for CAT I defect closure by SAT and DT/OT events
  - Status: PMO and contractor agreed to increase projection by 100 defects

OSD analysis predicts defect burn down risk to test events; up to 300 additional defects and 3-4 months to fix per OSD’s CAT I defects predictions.
Complicating Factors

- Late Formal CM (and CR process)
- Lab (late stand up)
- QA
- SW Supplier Quality
- M&S (Lab Fidelity)
- Staffing: Loss of Knowledge, Skill, Ability and its effect on test
- Loosening of Allowable Defects for IOT&E

**Defects Stabilizing?**
System Maturity

Alternate defect burn-down.
MPS-added trend-lines to system burn-down.
We considered staffing relative to current trends. Based on exponential trends, we projected more time required for burn down.

**SW Maturity Risk**
- Stability
- Behind on Testing
- Defect Burn-down

Un-modeled factors contributing to maturity risk

**Execsum:** Enough capacity? Sufficient schedule? Is X months to fix bugs/test insufficient mitigation?
Risk: If SW maturity problem is discovered, it is too late to add staff (Brooke's Law)

**First Order Impact** – Unsuccessful IOT&E

**Recommend**
(1) Maintain/add staff
(2) Monitor metrics submitted to OSD -- until adequately demo maturity
OSD Recommended Agile Metrics - Frequent Learning & Decision Support

- Team and aggregate velocity
- Team task & Aggregate story point burn-down
- Capabilities/Features completed
- Capabilities/Features accepted
- Product Backlog Stability
- Rate of closure vs discovery of defects
- Rate of change of risk exposure
- Story point efficiency
- Percentage of key stakeholder groups represented at sprint demos
- Peer review effectiveness
- Hours per defect to resolve
- Staffing planned vs actual (by month/sprint)
- Staffing stability
- Agile competency

Core

DoD: Definition of Done

At end of every iteration, the only work that is declared “Done” is ... work that has gone through the proper engineering activities and could be employed by End-Users

Agile tools and team & exec-level Agile metrics support rapid, near real-time decisions at IPT, PMO, CEO/CIO and MDA levels

Fibonacci series often used for story points
To meet Better Buying Power goals to inform decision makers of program risk, we need to characterize a system’s technical and SW maturity (e.g. quality and stability) and quantitatively assess its readiness at key knowledge points in the development life cycle. OT readiness is one of several critical points.” (Mr. James Thompson, Director, MPS)
Enterprise Benchmarking and Software Parametric Statistical Analysis

- **SW benchmarks:** a foundation supporting DASD(SE) decisions
  - underpin OSD independent parametric models
  - identifying statistically anomalous behavior
- **OSD SW Team gathers data and extracts core SW metrics**
  - for SW assessment and parametric analyses
  - e.g. size, effort, staffing, schedule, defects, demographics, etc.
- **Validated metrics form critical basis of comparison; benchmarking DoD trend-lines**
- **Parametric analyses of supported top findings in 21 program engagements**
  - Analyzed 52 SW releases
- **Analytical support to Congressional reporting on key Warfighting capability**
  - Unprecedented support: 413 excursions on complex SW helped deliver RTC predictions

DoD Trend—Schedule vs. Size

Program Optimism Range?

FY14 Software Team Engagements by Domain

- **Air Launched Weapon, 1**
- **UAV, 5**
- **Electronics, 3**
- **Space, 3**
- **Missile Defense, 1**
- **Land Warfare, 2**
- **Fixed Wing, 1**
- **Rotary Wing, 1**
- **Ships, 0**
- **Comms, 2**
- **MAAS, 0**
Sample of Software Benchmarks

Increasing statistical power year over year — adding dozens of closed and planned projects.

FY14 projects averaged 13% larger than historic trends.

DoD defied industry trends with lower productivity for larger software staffs.

Improving DoD SW practice via program benchmarking.

Based on SRDR data for 374 subsystems DoD SW programs roughly follow the industry productivity heuristic of 1-2 SLOC per hour.
Conclusions & Future Goals

1. We defined best practice performance measures
   – 21 categories of expected performance measures

2. We developed the SEP measurement assessment framework
   – In use internally to support ODASD(SE)/MPS assessment of SEPs’ metrics

3. We quantitatively analyzed programs to identify trends and gaps
   – Most MAIS programs measuring Process but not Product performance; MDAPs are the reverse
   – General lack of quantifiable threshold/objective values & end-to-end performance

4. Goal: Strengthen **consistent** use of SE & SW metrics DOD-wide
   – Characterize SE & SW development status at key knowledge points & define inflection points
   – Address tracking of metrics & enable path to growth (flexibility, tech refresh, modularity, quality…)

5. Goal: Quantitative, enterprise wide-assessments about state of DoD software practice
   – Employ new analysis tools and OSD benchmarks

6. Goal: Promote & advance **quantitative SE approach** by enhancing
   – performance forecasting throughout the life cycle and support rapid decision making
   – structuring & leveraging of program data; trend analyses; correlation of results across programs

Performance measurement remains critical to DoD
Additional Reference Material

- Abstract
- Who are we?
- Performance Measures
- Zipper Chart: Tester Developer Disconnect
- DASD(SE) Portfolio
Abstract

This presentation details how the Office of the Deputy Assistant Secretary of Defense for System Engineering uses metrics and quantitative analysis to shape program plans, monitor execution, and assess technical risk in support of the acquisition decision making process. We discuss our analysis of defense, industry, and academia recommendations for process and product performance measures.

Based on these analyses, we developed a set of expected performance measure categories and an assessment framework to analyze acquisition programs’ performance. Over the last year, we assessed performance measures on a variety of programs to identify best practices and improve System Engineering Plans (SEPs). Given our framework, we provided guidance and mentored a wide variety of major programs on performance measures, including Agile metrics and software maturity and test readiness metrics. The framework also serves as the basis for enterprise benchmarking, to include software parametric statistical analyses.
A Robust Systems Engineering Capability Across the Department Requires Attention to Policy, People and Practice
Where DASD(SE) sits in OSD

- USD (Acquisition, Technology, & Logistics)
- USD (Policy)
- USD (Comptroller) / Chief Financial Officer
- USD (Personnel & Readiness) / Chief Human Capital Officer
- USD (Intelligence)
- USD Deputy Chief Management Officer

- Inspector General, DoD ***
- Director, Cost Assessment & Program Evaluation
- Director, Operational Test & Evaluation
- General Counsel, DoD
- ASD (Legislative Affairs)
- ATSD (Public Affairs) *
- DoD Chief Information Officer *
- Director, Administration & Management *
- ATSD (Intelligence Oversight) *
- Director, Net Assessment *

** All positions shown are Presidentially Appointed, Senate-confirmed (PAS) except those with * which are SES positions

** As of February 2013
D,MPS Mission

“Inform Decision Makers to Understand and Mitigate Risk”
- Better Buying Power 3.0

- Emphasize quantitative understanding consistent with industry practice of systems engineering
- Make visible relationships between system/equipment design objectives and performance
- Harness and use existing information for timely and better decisions at the appropriate levels
- Enable data-driven decisions

“To meet Better Buying Power goals to inform decision makers of program risk, we need to characterize a system’s technical maturity and quantitatively assess its readiness at key knowledge points in the development life cycle.

-Mr. James Thompson, Director, MPS

Reduce Risk: Data-driven Decisions at Every Level.
Recommended Metrics for SW Development

- **Requirements**
  - Periodic (e.g., monthly) & Cumulative volatility
  - Total # SW requirements
  - Requirements deferred to later builds
  - SW requirements growth

- **Technical Performance**
  - Metrics related to KPPs, KSAs, MOEs, MOPs
  - Latency by CSCI and aggregate
  - CPU utilization
  - Bandwidth usage as percent of planned hardware capacity
  - Quality attributes (-ilities)
  - End-to-end mission performance (e.g., time to perform mission critical function, # simultaneous users)

- **Cost**
  - EVM
  - Tradeoff analysis (as related to performance and schedule)

- **Risk**
  - Risk Burn Down
  - Risk Cube
  - Risk Curve

- **Progress/Schedule**
  - Size (Planned vs. Actual)
  - SW build completion date (plan v. actual)
  - Capabilities/features (plan v. actual)
  - SW requirements verified (plan v. actual)
    - Safety-critical SW requirements
  - Requirements documents status (e.g., # ICDs defined)
  - Design artifacts (e.g., # use cases complete—actual v. plan)
  - Tests (completed v. planned)
  - Productivity (planned v. actual)

- **Resources**
  - Staffing levels (planned vs actual)

- **Product Quality**
  - Defects by severity/priority
  - Defect backlog
  - Defect/maturity target
  - Defects open and closed
  - Defects aging
  - Average resolution time in hours
  - Technical debt
  - Defects inherited/deferred by build
  - Life-cycle phase containment

**Top 4 Metrics to Assess Readiness to transition from DT to OT**
1. Capabilities met
2. Quality:
   a. Defect Profile Risk Assessment
   b. Defect Backlog (No Pri 1 or 2 defects)
   c. Key Quality Attributes “-ilities”

**Common set of metrics for all programs.**
Recommendations
Performance Measures for OT Readiness

1. **Capabilities met**
   - Status of end-to-end performance / SW-related mission thread TPMs (from SEP)
     - Go/No-go criteria: all measures have met threshold; Present any “non-green” measures
     - Examples: Time to decision, time to perform mission critical function, latency by CSCI and aggregate, bandwidth usage as percent of planned hardware capacity, # simultaneous users
   - Status of MOEs/MOSs/CTPs (from TEMP)
     - Go/No-go criteria: all measures have met threshold; Present any “non-green” measures
     - Related measures already linked to TEMP
     - Status of KPPs/KSAs/TPMs
     - Requirements verification status [Go/No-go : X% of all operational & system requirements met]

2. **Quality**
   - **Risk Assessment based on Defect Profile**
     - Defect profile – discovered vs. closed over time (includes defect backlog)
     - Other defect information to understand maturity risk: e.g. defect aging
     - Go/No-go criteria
       » Developmental test readiness level: 95% of expected defects have been found & fixed
       » Operational test readiness level: 99% of expected defects have been found & fixed
   - **Defect Backlog**
     - No open Priority 1 or 2 (showstopper) defects
     - No excessive (user-acceptable) number of open Priority 3 & 4 (workaround/nuisance) defects
   - **Key Quality attributes on track**
     - e.g. System stability (# anomalies per test hour)

For transitioning from DT to OT, the 2-4 metrics are related to results of DT testing.

Tester-developer disconnect & the “SE trace”
TPMs should trace to the relevant requirements (operational and system), KPPs, KSAs, and capability measures (e.g. MOEs). Program managers will ensure consistency between system TPMs and operational Critical Test Parameters.
Performance Measurement Disconnects
What can (and does) happen?

From JCIDs and User

User Context / User Capabilities

ICD

AoA (MOEs/MOSs)

CDD/CPD

KPPs

KSAs

COICs

Transition

From SEP

Operational Requirements / Architecture

System Requirements / Architecture

System / Subsystem Design

Development / Implementation

Lack of correspondence / communication between developer and user/ tester perspectives

From TEMP

Evaluation Framework (MOEs/MOSs/MOPs)

Verification

Integrators

Acronyms:
AoA – Analysis of Alternatives
COIC – Critical Operational Issues and Criteria
CTP – Critical Technical Parameter
KPP – Key Performance Parameter
KSA – Key System Attribute
MOE – Measure of Effectiveness
MOS – Measure of Suitability
MOP – Measure of Performance
TPM – Technical Performance Measure
Top 10 Metric Categories Over Time
Community vs. FY14 vs. FY15

Community vs. FY14 vs. FY15 identified Top 10 Metrics in SEPs
1= Community sources identified rankings for metrics

- Software
- Requirements management
- System Performance
- Mission Performance
- Reliability
- System Quality
- Schedule
- Staffing
- Integration
- Net Ready KPP
- Cost
- Risk Management
- Technology Maturity
- Design/development
- User Acceptance
- Supportability - Maintainability
• **World’s largest engineering organization**
  – Over 99,000 Uniformed and Civilian Engineers
  – Over 39,000 in the Engineering (ENG) Acquisition Workforce

• **DoD Systems Engineering focuses on engineering excellence**
  – Design, develop, construct and operate complex systems
  – Forecast their behavior under specific operating conditions
  – Deliver their intended function while addressing economic efficiency, environmental stewardship and safety of life and property

A Robust Systems Engineering Capability Across the Department Requires Attention to Policy, People and Practice
Preponderance of Acquisition Funding is for ACAT 1D and 1C Programs

• 97% of acquisition funding is in MDAP 1D and 1C programs
• Software applications are components or sub-components of large, complex systems
• Software must:
  – Support system / component / sub-component requirements
  – Support overall program / component / sub-component schedules
  – Support integration with other system software and hardware
• Software acquisition for 1D and 1C programs poses some of our toughest systems engineering challenges

Program data – SE oversight; Cost data from DAMIR

<table>
<thead>
<tr>
<th>Number of Programs by Category</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>60</td>
</tr>
<tr>
<td>1C</td>
<td>70</td>
</tr>
<tr>
<td>IAM</td>
<td>31</td>
</tr>
<tr>
<td>IAC</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Funding by Category ($Billion, %)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IAM</td>
<td>625, 34%</td>
</tr>
<tr>
<td>1C</td>
<td>28, 2%</td>
</tr>
<tr>
<td>1D</td>
<td>1150, 63%</td>
</tr>
<tr>
<td>IAM</td>
<td>21, 1%</td>
</tr>
</tbody>
</table>
DASD(SE) Portfolio

- System Engineering oversight of 182 Programs with acquisition costs of $1.8T
- Approve System Engineering Plans (SEPs)
- Assess preliminary and critical design reviews (PDR, CDR)
"So-what": OSD’s quantitative estimates of contractor performance in 2012 was very close to 2014 actual performance.