Understanding CMMI Measurement Capabilities & Impact on Performance: Results from the 2007 SEI State of the Measurement Practice Survey

Dennis R. Goldenson
Software Engineering Institute

CMMI Technology Conference
14 November 2007
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Understanding the State of Measurement Practice

Careful & well executed use of measurement & analysis

- Is a well accepted tenet in many fields of endeavor
- Including of course CMMI

Basic aims

- To inform management & technical decisions based on empirical evidence
- & to judge the results of those decisions once made

But, how well, and how frequently, are measurement practices put into effect in our own field?
Benchmarking: The current state

- Some professional & consulting organizations maintain repositories they use for establishing benchmarks & facilitating benchmarking activities
- However, their measures & measurement definitions differ in many ways
- In that sense, one cannot speak confidently about "industry standards"
- Which is why the SEI has launched the Performance Benchmarking Consortium {as described at last year's CMMI Technology Conference}

The state of the practice surveys

- Aim to provide data that's not yet widely available
  - Updates of trends in typical use of measurement in software & systems engineering
  - To help projects & organizations judge their progress relative to others
- But there also will be a continuing need to track qualitative as well as quantitative descriptions about the quality & frequency use of measurement in our field
New this year

- Screening question to identify respondents whose organizations develop software but rarely if ever do measurement
- Questions about
  - Resources & infrastructure devoted to measurement
  - Practices to ensure data quality & integrity
  - Value added by doing measurement
  - The kinds of measures used by the responding organizations

Among other things, these questions allow us to make some useful comparisons by CMMI maturity level
1st survey described at last year's CMMI technology Conference

Similar results this year

- Moderately strong relationships exist when comparing the replies of respondents based on:
  - Management versus staff roles
  - Industry \textit{versus} government organizations
  - The United States \textit{versus} other countries
  - Organization size

But that's a topic for another time
Outcomes

Today’s focus

- Provide evidence about the circumstances under which measurement capabilities and performance outcomes are likely to vary
- As a consequence of achieving higher levels of CMMI maturity

Most differences are consistent with expectations based on CMMI

- Which provides confidence in the validity of the model structure & content

However, the results also highlight areas where sometimes considerable room for improvement remains

- Even at maturity levels 4 and 5
- For example
  - A rather strong overall relationship between maturity level & use of measures about quality attributes
  - Little attention to quality attributes at the lower maturity levels
  - Yet, almost half of maturity level 4 & 5 respondents’ organizations track quality attributes only occasionally at best
Random sample of SEI customers

• 944 valid email invitations to participate

Data collected 20 February through 10 April 2007

• Two reminders

Response rate

• 41% completed all or part of the questionnaire

• N = 384

• Individual questions answered by 75-97% of respondents
  ∆ ~29% 39% of the sample invitees
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Understanding CMMI Measurement Capabilities

Role in the Organization

- Executive: 42%
- Program manager: 10%
- Project manager: 10%
- Engineer: 13%
- Analyst: 9%
- Programmer: 4%
- Other: 12%

N = 366
Understanding CMMI Measurement Capabilities

Dennis R. Goldenson, 14 November 2007

© 2007 Carnegie Mellon University

Who are the others?

- Quality: 26%
- Process: 24%
- Process + Quality: 20%
- Consultant: 15%
- Management: 9%
- Other Others: 6%

N = 155

= 8% of all those responding
<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process + Measurement</td>
<td>3</td>
</tr>
<tr>
<td>Measurement Specialist</td>
<td>1</td>
</tr>
<tr>
<td>Process + Quality + Measurement + Training</td>
<td>6</td>
</tr>
<tr>
<td>Quality + Process + Measurement</td>
<td>1</td>
</tr>
<tr>
<td>Training</td>
<td>6</td>
</tr>
<tr>
<td>Architect</td>
<td>4</td>
</tr>
<tr>
<td>Security</td>
<td>2</td>
</tr>
<tr>
<td>Testing</td>
<td>2</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
</tr>
</tbody>
</table>

One each:
- Administrative support
- Coach
- Consultant + researcher
- Engineering Manager + Process
- Process + Project engineer
- Program / team lead
- Program manager + Quality + Process
- Project manager + Quality
- Project manager + Engineer
- Not specified
Understanding CMMI Measurement Capabilities

Dennis R. Goldenson, 14 November 2007

© 2007 Carnegie Mellon University

N = 366

- Commercial shrink-wrap: 37%
- Custom software development: 13%
- In-house or proprietary: 11%
- Defense contractor: 7%
- Other government contractor: 4%
- Defense or military organization: 4%
- Other government agency: 3%
- Consultancy: 3%
- Other: 16%

Software Engineering Institute | Carnegie Mellon
Country

23%

48%

12%

N = 363

United States
India
Japan
France
Germany
United Kingdom
Canada
Netherlands
All others
Understanding CMMI Measurement Capabilities

Dennis R. Goldenson, 14 November 2007

© 2007 Carnegie Mellon University

N = 364

Percent

<table>
<thead>
<tr>
<th>Category</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 or fewer</td>
<td></td>
</tr>
<tr>
<td>51-100</td>
<td></td>
</tr>
<tr>
<td>101-200</td>
<td></td>
</tr>
<tr>
<td>201-500</td>
<td></td>
</tr>
<tr>
<td>501-2000</td>
<td></td>
</tr>
<tr>
<td>More than 2000</td>
<td></td>
</tr>
</tbody>
</table>
Maturity level

Percent

N = 365

Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Don't Know
Maturity Level: Use of Measurement in the Organization

<table>
<thead>
<tr>
<th>Maturity Level</th>
<th>Don't know</th>
<th>Rare or never</th>
<th>Occasional</th>
<th>Routine</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML1&amp;DK (N = 151)</td>
<td>8%</td>
<td>30%</td>
<td>34%</td>
<td>28%</td>
</tr>
<tr>
<td>ML2 (N = 84)</td>
<td>2%</td>
<td>19%</td>
<td>70%</td>
<td>7%</td>
</tr>
<tr>
<td>ML3 (N = 59)</td>
<td>3%</td>
<td>22%</td>
<td>75%</td>
<td>3%</td>
</tr>
<tr>
<td>ML4&amp;5 (N = 71)</td>
<td>3%</td>
<td>96%</td>
<td>2%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Gamma = .73  \( p < .0001 \)
The results:
The Respondents’ Measurement Roles

<table>
<thead>
<tr>
<th>ML1&amp;DK</th>
<th>N = 151</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>8%</td>
</tr>
<tr>
<td>Neither</td>
<td>20%</td>
</tr>
<tr>
<td>User</td>
<td>50%</td>
</tr>
<tr>
<td>Provider</td>
<td>12%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ML2</th>
<th>N = 84</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>13%</td>
</tr>
<tr>
<td>Neither</td>
<td>10%</td>
</tr>
<tr>
<td>User</td>
<td>38%</td>
</tr>
<tr>
<td>Provider</td>
<td>17%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ML3</th>
<th>N = 59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>8%</td>
</tr>
<tr>
<td>Neither</td>
<td>51%</td>
</tr>
<tr>
<td>User</td>
<td>17%</td>
</tr>
<tr>
<td>Provider</td>
<td>14%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ML4&amp;5</th>
<th>N = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>7%</td>
</tr>
<tr>
<td>Neither</td>
<td>61%</td>
</tr>
<tr>
<td>User</td>
<td>9%</td>
</tr>
<tr>
<td>Provider</td>
<td>9%</td>
</tr>
</tbody>
</table>

p = .04
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
How Measurement Work is Staffed

- ML1&DK (N = 78) 9%
- ML2 (N = 60) 13%
- ML3 (N = 58) 12%
- ML4&5 (N = 60) 7%

- Project Level:
  - 34% other
  - 34% don't know
  - 28% organization wide group
  - 20% project level
  - 13% other
  - 50% project level

- Other:
  - 41% other
  - 33% don't know

- A few key experts:
  - 3%, 1%, 2% & 3% respectively

p < .006
Budgets for Measurement

<table>
<thead>
<tr>
<th>Category</th>
<th>ML1&amp;DK</th>
<th>ML2</th>
<th>ML3</th>
<th>ML4&amp;5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>76</td>
<td>68</td>
<td>50</td>
<td>61</td>
</tr>
<tr>
<td>Don't know</td>
<td>7%</td>
<td>18%</td>
<td>18%</td>
<td>38%</td>
</tr>
<tr>
<td>No</td>
<td>72%</td>
<td>65%</td>
<td>56%</td>
<td>34%</td>
</tr>
<tr>
<td>Yes</td>
<td>21%</td>
<td>18%</td>
<td>22%</td>
<td>22%</td>
</tr>
</tbody>
</table>

p < .0001
### Qualified Measurement Staff

**Availability of Qualified Measurement Staff**

<table>
<thead>
<tr>
<th>Capability</th>
<th>ML1&amp;DK (N = 76)</th>
<th>ML2 (N = 65)</th>
<th>ML3 (N = 50)</th>
<th>ML4&amp;5 (N = 61)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML1</td>
<td>51%</td>
<td>26%</td>
<td>26%</td>
<td>11%</td>
</tr>
<tr>
<td>ML2</td>
<td>38%</td>
<td>34%</td>
<td>40%</td>
<td>61%</td>
</tr>
<tr>
<td>ML3</td>
<td>35%</td>
<td>40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML4&amp;5</td>
<td>18%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gamma** = .44  \( p < .0001 \)
For:

- Automated measurement support for data collection, data management, data analysis & reporting
- Use of commercial measurement packages & tools
- Existence of common, integrated organizational measurement repositories
- Availability of measurement related training

Proportions sometimes vary across the distributions.

But there are consistent differences by maturity level.
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- **Value added by measurement**
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
### Better Project Performance

- **ML1&DK**
  - Rare, never, worse, DK or NA: 26%
  - Half time or on occasion: 50%
  - Always or frequently: 24%

- **ML2**
  - Rare, never, worse, DK or NA: 12%
  - Half time or on occasion: 40%
  - Always or frequently: 35%

- **ML3**
  - Rare, never, worse, DK or NA: 20%
  - Half time or on occasion: 40%
  - Always or frequently: 35%

- **ML4&5**
  - Rare, never, worse, DK or NA: 27%
  - Half time or on occasion: 40%
  - Always or frequently: 26%

**Gamma** = .41

**p < .0001**

### Better Product Quality

- **ML1&DK**
  - Rare, never, worse, DK or NA: 26%
  - Half time or on occasion: 38%
  - Always or frequently: 26%

- **ML2**
  - Rare, never, worse, DK or NA: 13%
  - Half time or on occasion: 44%
  - Always or frequently: 38%

- **ML3**
  - Rare, never, worse, DK or NA: 22%
  - Half time or on occasion: 34%
  - Always or frequently: 44%

- **ML4&5**
  - Rare, never, worse, DK or NA: 7%
  - Half time or on occasion: 63%
  - Always or frequently: 30%

**Gamma** = .34

**p < .0002**
Better Tactical Decisions

- ML1&DK: 27% (Gamma = .35, p = .0001)
- ML2: 20% (58%)
- ML3: 26% (36%)
- ML4&5: 9% (38%)

Better Strategic Decisions

- ML1&DK: 38% (Gamma = .31, p = .0008)
- ML2: 39% (41%)
- ML3: 35% (39%)
- ML4&5: 13% (49%)

Legend:
- Rare, never, worse, DK or NA
- Half time or on occasion
- Always or frequently
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- **Software measures used**
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Organizational Measurement Results

**Cost Performance**

- **Rarely, never, DK, or NA**
  - ML1&DK: 21% (Gamma = .25, p < .03)
  - ML2: 24% (Gamma = .25, p < .03)
  - ML3: 11% (Gamma = .25, p < .03)
  - ML4&5: 10% (Gamma = .25, p < .03)

- **Occasionally**
  - ML1&DK: 23% (Gamma = .25, p < .03)
  - ML2: 24% (Gamma = .25, p < .03)
  - ML3: 27% (Gamma = .25, p < .03)
  - ML4&5: 25% (Gamma = .25, p < .03)

- **Frequently**
  - ML1&DK: 33% (Gamma = .25, p < .03)
  - ML2: 38% (Gamma = .25, p < .03)
  - ML3: 38% (Gamma = .25, p < .03)
  - ML4&5: 53% (Gamma = .25, p < .03)

- **Regularly**
  - ML1&DK: 23% (Gamma = .25, p < .03)
  - ML2: 38% (Gamma = .25, p < .03)
  - ML3: 38% (Gamma = .25, p < .03)
  - ML4&5: 11% (Gamma = .25, p < .03)

---

**Schedule Performance**

- **Rarely, never, DK, or NA**
  - ML1&DK: 14% (Gamma = .37, p = .0006)
  - ML2: 3% (Gamma = .37, p = .0006)
  - ML3: 11% (Gamma = .37, p = .0006)
  - ML4&5: 7% (Gamma = .37, p = .0006)

- **Occasionally**
  - ML1&DK: 19% (Gamma = .37, p = .0006)
  - ML2: 34% (Gamma = .37, p = .0006)
  - ML3: 34% (Gamma = .37, p = .0006)
  - ML4&5: 11% (Gamma = .37, p = .0006)

- **Frequently**
  - ML1&DK: 34% (Gamma = .37, p = .0006)
  - ML2: 48% (Gamma = .37, p = .0006)
  - ML3: 73% (Gamma = .37, p = .0006)
  - ML4&5: 61% (Gamma = .37, p = .0006)

- **Regularly**
  - ML1&DK: 33% (Gamma = .37, p = .0006)
  - ML2: 48% (Gamma = .37, p = .0006)
  - ML3: 33% (Gamma = .37, p = .0006)
  - ML4&5: 33% (Gamma = .37, p = .0006)
Organizational Measurement Results

Business Growth & Profitability

<table>
<thead>
<tr>
<th></th>
<th>ML1 &amp; DK</th>
<th>ML2</th>
<th>ML3</th>
<th>ML4 &amp; 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>70</td>
<td>55</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>40% Rarely</td>
<td>21%</td>
<td>15%</td>
<td>24%</td>
<td>31%</td>
</tr>
<tr>
<td>33% Occasionally</td>
<td>31%</td>
<td>22%</td>
<td>24%</td>
<td>31%</td>
</tr>
<tr>
<td>33% Frequently</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20% Regularly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gamma = .20
p = .2244
Defect Density

- Rarely, never, DK, or NA: 4%
- Occasionally: 22%
- Frequently: 31%
- Regularly: 51%

Gamma = .41
p < .0001

ML1&DK: N = 70
ML2: N = 56
ML3: N = 45
ML4&5: N = 52

Defect Phase Containment

- Rarely, never, DK, or NA: 8%
- Occasionally: 13%
- Frequently: 27%
- Regularly: 49%

Gamma = .44
p < .0001

ML1&DK: N = 70
ML2: N = 56
ML3: N = 45
ML4&5: N = 51

- Product & Quality Measurement Results Reported
- Gamma = .41
- p < .0001
- Gamma = .44
- p < .0001
Customer Satisfaction

ML1 & DK
N = 70
Gamma = .31

ML2
N = 56

ML3
N = 45

ML4 & 5
N = 52

Gamma = .31
p < .005

Rarely, never, DK, or NA
Occasionally
Frequently
Regularly
For:

- Adherence to work processes
- Effort applied to task
- Estimation accuracy
- Cycle time

Proportions sometimes vary across the distributions.

But there are consistent differences by maturity level.
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- **Data quality & integrity**
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Maturity Level: Practices to Ensure Data Quality

Statistical estimates of measurement error

- ML1&DK: N = 74, 61% (Gamma = .44, p < .0001)
- ML2: N = 56, 59%
- ML3: N = 47, 47%
- ML4&5: N = 51, 14%

Checks for inconsistent interpretation

- ML1&DK: N = 74, 27% (Rarely, never, or DK)
- ML2: N = 57, 23%
- ML3: N = 48, 30%
- ML4&5: N = 50, 37%

- ML1&DK: N = 74, 27% (Half time or on occasion)
- ML2: N = 57, 31%
- ML3: N = 48, 46%
- ML4&5: N = 50, 25%

- ML1&DK: N = 74, 27% (Always or frequently)
- ML2: N = 57, 30%
- ML3: N = 48, 38%
- ML4&5: N = 50, 74%
## Maturity Level: Practices to Ensure Data Quality

### Checks for unusual distribution patterns

<table>
<thead>
<tr>
<th>Maturity Level</th>
<th>Gamma</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML1 &amp; DK</td>
<td>.46</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ML2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML4 &amp; 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Frequencies

- **Rarely, never, or DK**
  - ML1 & DK: 12%
  - ML2: 31%
  - ML3: 31%
  - ML4 & 5: 86%

- **Half time or on occasion**
  - ML1 & DK: 28%
  - ML2: 33%
  - ML3: 44%
  - ML4 & 5: 12%

- **Always or frequently**
  - ML1 & DK: 32%
  - ML2: 36%
  - ML3: 33%
  - ML4 & 5: 25%
For:

- Out of range & illegal values ... Number & distribution of missing data
- Missing data not treated as zero ... Precision & accuracy tests
- Other aspects of alignment & coordination of measurement activities
  - Understandable & consistent measurement definitions
  - Understandable & interpretable measurement results
  - Use of "standard" measurement methods
  - Measurable product & service criteria
  - Measurement used to understand product & service quality
  - Documented data collection process
  - Documented process for reporting results
  - Corrective action taken when thresholds exceeded
  - Understands purposes of the data collected/reported

Proportions sometimes vary across the distributions.
But there are consistent differences by maturity level.
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Perspectives

Not Relevant for Decision Making

- ML1&DK: 23% (N = 102)
- ML2: 39% (N = 61)
- ML3: 44% (N = 41)
- ML4&5: 55% (N = 53)

Gamma = .27
p = .0002

Onerous or Burdensome

- ML1&DK: 4% (N = 110)
- ML2: 20% (N = 67)
- ML3: 25% (N = 45)
- ML4&5: 11% (N = 52)

Gamma = .17
p < .45
For:

- Stated negatively
  - Inappropriate collection & use of data
  - Resistance to “extra” work
- Stated positively
  - Understandable & interpretable results
  - Data collected are regularly analyzed
  - Measurement an integral part of the business
  - Objective results highly valued

Once again:

- Proportions sometimes vary across the distributions.
- But there are consistent differences by maturity level.

Yet resistance to measurement still exists in our field.

- Even in high maturity organizations
Purpose & scope of the survey

Results

- The respondents & their organizations
- Measurement resources & infrastructure
- Value added by measurement
- Software measures used
- Data quality & integrity
- Organizational perspectives on software measurement

Summary, lessons learned & next steps
Characteristic differences associated with CMMI Maturity level achieved

- Measurement capability & performance outcomes
- Common stair step pattern up the maturity levels
- Some quite substantial

Still, some of the results imply room for improvement

- Sometimes substantial room

Even in higher maturity organizations

- Although the expectations for quality & “goodness” may well be higher there too
- Jim Herbsleb & I saw a similar pattern years ago
  - For process champions *versus* practitioners & managers
Lower than desired response rate

- Lower maturity level respondents less likely to finish the questionnaire
- Some drop off in higher maturity level respondents later in questionnaire

Not surprising in a relatively long questionnaire ... but exacerbated by:

- Spoofed email invitations & reminder message errors
- Related problems with incremental saving
  - Cookie flushing & assignment of multiple URLs by COTS web survey product
  - Leading to lost information
- & (possibly) lack of feedback on time/length remaining

Recurring anomalous dip at maturity level 3

- May be due to bias from relatively small number of ML3 respondents
- Or learning curve effects ... or higher expectations
There *always* is noise in survey (and other measurement) data, e.g.

- Differing interpretations of intended meaning of questions
- Use of *vague quantifiers* in closed ended response categories
- *Don’t know* & other off scale responses
  - Most common at lower maturity levels
  - But they also exist at the higher maturity levels
  - Perhaps because some folks in larger organizations truly *don’t know*

Regardless, the survey results are consistent with expectations based on CMMI

- a.k.a. predictive validity
Relatively little data yet exist for meaningful comparisons among software & systems engineering projects & organizations

Â Hence tendency to cover too much at once in a single sample survey

Considering variants on matrix sampling strategies for 2008 survey

Â Answer only a subset of questions ... to avoid over-burdening the respondents

State of the practice can refer to very different target populations

Â The SEI customer base ... the broader software & systems engineering community ... or those organizations that more routinely use measurement?

Â Of course, the answer depends on the purposes of the survey
Our plans

- We will track change over time & go into further depth about focused topics from the perspective of current measurement practitioners

Considering parallel samples for 2008

- A short set of questions for tracking the diffusion of measurement through the broader software & systems engineering community
- Possible focus on issues faced with respect to the adoption & use of high maturity measurement practices

Also fielding a survey on Program Office acquisition capabilities (early 2008)

Of course, there is no shortage of additional topics for the future

- In the SEI series or in those that we hope to see done by others
Dennis R. Goldenson
dg@sei.cmu.edu

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA