Sampling for Software Process Assessments, Evaluations, and Appraisals

Dr. Mark C. Paulk
13 October 2017
What Is An Assessment?

- An appraisal of an organization's current software process for self-improvement by a trained team of experienced software professionals.

- It is based on
  - review of 4 to 6 representative projects
  - responses to the maturity questionnaire
  - in-depth discussions with project managers and practitioners
  - collective knowledge and experience of the assessment team
The “Evolution” of CMMs

- Software process maturity framework 1987
- Software CMM v1.1 1993
  - Software Acquisition CMM
  - Systems Engineering CMM
  - … over 60 different CMMs – staged and continuous
- eSCM for Service Providers v1 2001
- CMMI for Development v1.1 2002
  - CMMI for Acquisition
  - CMMI for Services
- eSCM for Service Providers v2 2004
- CMMI for Development v1.3 2010
The “Evolution” of Assessments (1 of 2)

- **Software process assessments (SPA)**
  - software process improvement, 1987 to 1995
  - “What are your problems?”
  - map problems to the maturity framework
  - emphasis on organizational intervention

- **Software capability evaluations (SCE)**
  - source selection, contract monitoring
  - audits against the maturity framework, 1988 to 1995
  - audits against the Software CMM, 1995 to 2003
  - emphasis on level playing field – fairness

- **Appraisals = assessments and evaluations**
The “Evolution” of Assessments (2 of 2)

- CMM-based appraisals for internal process improvement (CBA IPI)
  - software process improvement, 1995 to 2003
  - audit process against Software CMM
  - identify problems, map against Software CMM as appropriate
  - report CMM and non-CMM findings
  - emphasis on reliable and consistent results

- Standard CMMI appraisal method for process improvement (SCAMPI)
  - process improvement, 2003 till …
  - audits against CMMI
## Assessment Versus Evaluation

<table>
<thead>
<tr>
<th>Issue</th>
<th>Assessment</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use</td>
<td>process improvement</td>
<td>source selection</td>
</tr>
<tr>
<td>Objective</td>
<td>assess current practice</td>
<td>substantiate practice</td>
</tr>
<tr>
<td>Improvement goal</td>
<td>catalyst for improvement</td>
<td>evaluate commitment</td>
</tr>
<tr>
<td>Output</td>
<td>input for action plan</td>
<td>performance risk</td>
</tr>
<tr>
<td>Range of findings</td>
<td>non-CMM findings possible</td>
<td>CMM findings only</td>
</tr>
<tr>
<td>Style</td>
<td>collaborative</td>
<td>audit-oriented</td>
</tr>
<tr>
<td>Focus of results</td>
<td>applies to organization</td>
<td>predict next project</td>
</tr>
<tr>
<td>Status of results</td>
<td>confidential</td>
<td>known to DoD</td>
</tr>
</tbody>
</table>
SEI’s SCAMPI v1.3


• “Sampling is planned quantitatively based on
  – the diversity of unique process implementations within the appraisal scope,
  – with the goal of both ensuring a representative sample of the organizational unit and
  – optimizing the effort for collection and analysis of objective evidence.”
Organizational Analysis

- The scope of what “organization” means can cover
  - Multiple business units, product lines, domains (type of work)
  - Customer segmentation
  - Sizes of project and/or product, duration
  - Software engineering methods and environments
  - Geographical sites (location)
  - Workforce mixes (experience, …)
  - Organization vs project-level processes
SCAMPI Sampling Units

- Basic units ~ projects + support functions (organizational functions), …

- Organizational scope of the appraisal is selected as a representative sample of the organizational unit.
  - based on sampling factors that reflect meaningful differences in the conditions under which work is performed
SCAMPI Sampling Example

<table>
<thead>
<tr>
<th>City</th>
<th>Commercial customers</th>
<th>Government customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Denver</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>Total number of basic units</th>
<th>Basic units sampled</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York, Commercial</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>New York, Government</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Cincinnati, Commercial</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Denver, Commercial</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Denver, Government</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
The Problem

• Is this an adequate sampling approach?

• Performing assessments in large complex organizations quickly becomes cost prohibitive if every component (entity) must be assessed for each criterion

• Sampling methodology goals
  – Reduce the number of entities that must be assessed, especially for assessments with a large number of entities
  – Keep the risk of incorrectly certifying an organization that is not really compliant at or below 10% under most circumstances
Sampling

- Sampling allows one to specify the amount of error one can tolerate, thus maintaining the desired accuracy levels while substantially reducing the time and effort involved.

- “Four to six representative projects” might be considered a “convenience sample.”

- Covering all aspects of the organization and its projects is a distinct improvement but does not address the question of adequate coverage.
Statistical Sampling Prerequisites

- *Randomization*, every sampling unit (project) must have the same likelihood of being selected as part of any given sample.

- *Stratified random sample*, a sample obtained by separating the population elements into non-overlapping groups (strata), and then selecting a simple random sample from each stratum.
No Guarantees Assumption

- The ability within a stratum to detect a single entity that is not compliant was not a design goal.
  - A high likelihood of finding at least one noncompliant entity when at least two exist within a stratum.
  - It does not guarantee finding a single non-compliant entity.
  - It does not guarantee finding multiple non-compliant entities.
Perfect Team Assumption

- Any entity that is truly not compliant and is selected in the sample will be detected.

The Ultimate Sampling Objective

- Incorrectly passing a single Practice, while undesirable, is not as serious as incorrectly passing an organization at any given Capability Level.
Sampling Method

• Uses a variation of Stratified Random Sampling – entities are placed in similar strata by Practice (84 Practices in eSCM-SP)

• Stratifying at the Practice level allows us to differentiate while still achieving maximum efficiency for Practices that are implemented across large parts of the organization (e.g., HR, Contracting, etc.)
Stratifying the Population

• Identify the entities that are in scope for the assessment.

• Organize entities into strata that have similar characteristics (organizational analysis).
  – Consistent process implementations within a strata (homogeneous subgroups)
  – Responsibility of Lead Assessor
  – Approved by certification body
Example Stratification for 63 Entities

- 63 Entities
- 28 Practices
- 17 Practices
- 39 Practices

63 Entities

- 40
- 23
- 15
- 25
- 17
- 6

28 Practices
17 Practices
39 Practices
Sampling Table Design

• Designed to control the probability of an organization incorrectly achieving a Capability Level for which they were not truly compliant.

• Set the sample sizes such that the conditional probability is less than 10% that no Practice noncompliances will be detected, given that Practice noncompliances exist.

• If any sample would have a 70% or better chance of detecting at least one noncompliant entity, when there are two or more noncompliant entities in the stratum, we will achieve a minimum of a 91% probability of detecting at least one Practice noncompliance.
If any sample would have a 70% or better chance of detecting at least one noncompliant entity, when there are two or more noncompliant entities in the stratum, we will achieve a minimum of a 91% probability of detecting at least one Practice noncompliance.

<table>
<thead>
<tr>
<th>Number of Entities</th>
<th>Sample Size</th>
<th>Number of Entities</th>
<th>Sample Size</th>
<th>Number of Entities</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>All</td>
<td>21-22</td>
<td>10</td>
<td>39-40</td>
<td>18</td>
</tr>
<tr>
<td>4-7</td>
<td>3</td>
<td>23-24</td>
<td>11</td>
<td>41-42</td>
<td>19</td>
</tr>
<tr>
<td>8-9</td>
<td>4</td>
<td>25-27</td>
<td>12</td>
<td>43-44</td>
<td>20</td>
</tr>
<tr>
<td>10-11</td>
<td>5</td>
<td>28-29</td>
<td>13</td>
<td>45-46</td>
<td>21</td>
</tr>
<tr>
<td>12-13</td>
<td>6</td>
<td>30-31</td>
<td>14</td>
<td>47-49</td>
<td>22</td>
</tr>
<tr>
<td>14-16</td>
<td>7</td>
<td>32-33</td>
<td>15</td>
<td>50 and up</td>
<td>23</td>
</tr>
<tr>
<td>17-18</td>
<td>8</td>
<td>34-35</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-20</td>
<td>9</td>
<td>36-38</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

eSCM Sampling
A Simple eSCM Sampling Example

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Number of Entities in Stratum</th>
<th>Number of Practices</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Org</td>
<td>93</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>S1</td>
<td>43</td>
<td>67</td>
<td>20</td>
</tr>
<tr>
<td>S2</td>
<td>25</td>
<td>67</td>
<td>12</td>
</tr>
<tr>
<td>S3</td>
<td>17</td>
<td>67</td>
<td>8</td>
</tr>
<tr>
<td>S4</td>
<td>8</td>
<td>67</td>
<td>4</td>
</tr>
</tbody>
</table>
Failed Level Detection

Because a single failed practice is sufficient to fail at any Level, detecting Level failure is the same as detecting Practice failure.
Typical Results

- With a full census, the earlier example would have required 5,292 Entity-Practice pairs to be reviewed.
- With the sampling plan, only 2,307 need to be reviewed – a 56.4% reduction.

![Diagram showing comparison between full census and sampling plan.]

- **63 Entities**: 28 Practices, 17 Practices, 39 Practices
- **23 Practices**
- **15 Practices**
- **25 Practices**
- **17 Practices**
- **6 Practices**
Conclusions

• It is possible to significantly reduce the effort required to evaluate organizations with multiple entities while still controlling for the number of false positives

• While the reduction is significant in relative terms, the remaining effort for large organizations is still substantial in absolute terms
eSCM


  - “By sampling we mean the process of drawing an inference about some population of interest by examining only a subset of the population.”
Questions?