Affiliate Presentation:

*Process Maturity / Cost Analysis*

*Brad Clark*

March 11, 1996

---

**Presentation Outline:**

- **Overview**
  - Data Collection Status
  - Preliminary Results
  - Future Work
Question:

• What are the effects of Software Process Maturity on Development Effort?

• 2 Answers:

- There is an effect
- Can not tell

### Process Maturity:

<table>
<thead>
<tr>
<th>CMM Level</th>
<th>Key Process Areas (KPA)</th>
<th>KPA #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 Initial</td>
<td>Requirements Management</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Software Project Planning</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Software Project Tracking and Oversight</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Software Subcontract Management</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Software Quality Assurance</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Software Configuration Management</td>
<td>6</td>
</tr>
<tr>
<td>Level 2 Repeatable</td>
<td>Organization Process Focus</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Organization Process Definition</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Training Program</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Integrated Software Management</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Software Product Engineering</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Intergroup Coordination</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Peer Reviews</td>
<td>13</td>
</tr>
<tr>
<td>Level 3 Defined</td>
<td>Quantitative Process Management</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Software Quality Management</td>
<td>15</td>
</tr>
<tr>
<td>Level 4 Managed</td>
<td>Defect Prevention (8)</td>
<td>16</td>
</tr>
<tr>
<td>Level 5 Optimizing</td>
<td>Technology Change Management (8)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Process Change Management (10)</td>
<td>18</td>
</tr>
</tbody>
</table>
Affiliate Presentation:

Data Collection and Calibration Update

Brad Clark

March 11, 1996

Presentation Outline:

- Updated Data Collection Form
  - Data Collection Status
  - Repository Structure
  - Calibration
  - USC COCOMO Post-Architecture Software
  - Data Collection Call
Data Collection Form Update

- Minor changes:
  - Application Development Type: New or Maintenance
  - UNFM added
- Major change: Defect Insertion and Removal
  - What methods are used to prevent and detect defects?
- New form is Version 1.4
- Postscript file of data collection form is available in the ‘cocomo’ account on sunset.usc.edu

---

Top Portion of Defect Collection

<table>
<thead>
<tr>
<th>Method</th>
<th>Thoroughness of Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Project Reviews</td>
<td></td>
</tr>
<tr>
<td>Systems Requirements</td>
<td></td>
</tr>
<tr>
<td>System Architecture</td>
<td></td>
</tr>
<tr>
<td>Software Requirements</td>
<td></td>
</tr>
<tr>
<td>Software Architecture</td>
<td></td>
</tr>
<tr>
<td>Detailed Design</td>
<td></td>
</tr>
<tr>
<td>User Documentation</td>
<td></td>
</tr>
<tr>
<td>Maintenance Documentation</td>
<td></td>
</tr>
</tbody>
</table>
Assess KPA Goal Compliance:

- 52 goals for 18 KPAs

Goal attainment based on 7 ratings:
- **Almost Always** (over 90% of the time) when the goals are consistently achieved and are well established in standard operating procedures.
- **Frequently** (about 60 to 90% of the time) when the goals are achieved relatively often, but sometimes are omitted under difficult circumstances.
- **About Half** (about 40 to 60% of the time) when the goals are achieved about half of the time.
- **Occasionally** (about 10 to 40% of the time) when the goals are sometimes achieved, but less often.
- **Rarely If Ever** (less than 10% of the time) when the goals are rarely if ever achieved.
- **Does Not Apply** when you have the required knowledge about your project or organization and the KPA, but you feel the KPA does not apply to your circumstances (e.g. Subcontract Management).
- **Don’t Know** when you are uncertain about how to respond for the KPA.

For Other Factors:

- Use COCOMO 2.0 Cost Drivers:
  - Product: RELY, DATA, RUSE, CPLX, DOCU
  - Platform: TIME, STOR, PVOL
  - Personnel: ACAP, PCAP, AEXP, PEXP, LTEX, PCON
  - Project: TOOL, SITE, SCED
Research Cost Model:

\[ PM = A (X_1)^{b_1} (X_2)^{b_2} (X_3)^{b_3} (X_4)^{b_4} (X_5)^{b_5} \ldots (X_n)^{n} \]

- Each X is a factor that is believed to influence effort.
  - Size (KSLOC)
  - COCOMO Product cost drivers
  - COCOMO Platform cost drivers
  - COCOMO Personnel cost drivers
  - COCOMO Project cost drivers
  - 18 Key Process Areas
- That's a lot!

Not all KPA's may be significant:

- COCOMO 2.0 May 1995 Workshop break-out group discussion results:

<table>
<thead>
<tr>
<th>Key Process Area</th>
<th>Rating</th>
<th>Number of Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements Management</td>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>Project Tracking and Oversight</td>
<td>2.0</td>
<td>3</td>
</tr>
<tr>
<td>Software Product Engineering</td>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>Peer Reviews</td>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>Training Program</td>
<td>1.8</td>
<td>3</td>
</tr>
<tr>
<td>Software Configuration Management</td>
<td>1.6</td>
<td>4</td>
</tr>
<tr>
<td>Intergroup Coordination</td>
<td>1.4</td>
<td>3</td>
</tr>
<tr>
<td>Organization Process Definition</td>
<td>1.2</td>
<td>2</td>
</tr>
<tr>
<td>Software Subcontract Management</td>
<td>1.2</td>
<td>4</td>
</tr>
<tr>
<td>Integrated Software Management</td>
<td>1.0</td>
<td>2</td>
</tr>
<tr>
<td>Organizational Process Focus</td>
<td>0.8</td>
<td>3</td>
</tr>
<tr>
<td>Software Project Planning</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>Software Quality Assurance</td>
<td>0.2</td>
<td>4</td>
</tr>
</tbody>
</table>
Hypothesis Testing:

- The null hypothesis is that Process Maturity factors have no influence on Development Effort, i.e. the coefficient for those factors are zero (or near zero).
- The alternative is that the Process Maturity factors have coefficients large enough to be 95% certain that the null hypothesis can be rejected

\[ H_0 : b_{X_n} = 0 \]
\[ H_1 : b_{X_n} \neq 0 \]

- T statistic is used to assess the significance of coefficient:

\[ t_{\text{computed}} = \frac{b_{X_n}}{\text{est}(\sigma_{X_n})} \]
\[ |t_{\text{computed}}| \geq t(df, \alpha) \]

Status:

- Number of CMM-level observations: 42
- Number of KPA-level observations: 17
- Average frequency of KPA-level observations was high for CMM Level 2 and 3 KPAs and low for Level 4 and 5 KPAs (next slide).
  - Why is this?
Frequency Table: 17 Observations

<table>
<thead>
<tr>
<th>Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost Always</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>About Half</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Between</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasionally</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarely If Ever</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does Not Apply</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do Not Know</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relationship of PMAT to Effort

- Limited data problem
  - Not enough to perform least squares error fit for the 18 KPAs.
  - Need to add other COCOMO 2.0 cost drivers.
- Correlations between KPAs and other COCOMO 2.0 cost drivers
- Are all of KPAs and COCOMO cost drivers needed to determine if a relationship exists?
Future Work:

- Collect more data on KPAs and COCOMO cost drivers.
- Process existing data into the repository.
- Consider the frequency distribution of KPA responses - your input would be welcomed.

Notes:
Data Collection Status:

- Submission via Data Collection Form
- 10 Organizations
- 100 Data Points (over half are in the repository)
- KSLOC range: 6.7 to 980
- Effort (Person Months) range: 7.5 to 7554
- Schedule (Months) range: 5 to 81

Repository Structure:

- Five relational tables
  - Project Information
  - Size
  - Scale Factors
  - Early Design Multipliers
  - Post-Architecture Multipliers
- Layered access to data
  - Physical relations (Bottom layer)
  - Logical view (Middle layer)
  - Data manipulation programs (Top layer)
Repository Structure (con’t):

Interactive Entry Form
Amadeus Import
Bulk Data Export
GOCOMO 2 Model

Top: Data Manipulation

Post-Architecture View

Early Design View

Middle: Logical Views

Application Composition View

Proj
Size
Scale Factors
Early Design
Post-Architecture

Bottom: Relational Tables

Calibration:

- Enter data into repository
  - Look for incomplete data
  - Outliers get closer inspection

\[
\text{Relative Error} = \frac{\text{Estimated Effort} - \text{Actual Effort}}{\text{Actual Effort}} > 1.0
\]

- Cost Drivers entered symbolically, e.g. H, VH
- Cost Driver offsets entered in 0.25 increments
Calibration cycle

1. Translate / Interpolate symbolic values in repository with table of Cost Driver values
2. Analyze data using least-squares error techniques
3. Adjust table of Cost Driver values
4. Go back to step 1 until values stabilize

Preliminary Results:

- **Effort**: $PM_{\text{nominal}} = A (KSLOC)^B$
  - COCOMO 81 Data: $A = 1.7$, $B = 1.24$
  - Affiliate Data: $A = 2.6$, $B = 1.06$
  - All Data: $A = 2.42$, $B = 1.12$
- Plot of Actual Effort vs. Estimated $PM_{\text{nominal}}$ for All Data.
- Plot of Relative Error distribution for All Data.
Presentation Outline:

- Updated Data Collection Form
- Data Collection Status
- Repository Structure
- Calibration

USC COCOMO Post-Architecture Software

- Data Collection Call

USC COCOMO Post-Architecture Software:

- FREE Software available for MS Windows 3.1 or Sun UNIX with Motif
- Implements COCOMO 2.0 Post-Architecture model only.
- Available from the 'cocomo' user's account on sunset.usc.edu
- Proprietary to Affiliates
- Demonstration and diskettes available on Tuesday, March 12th on the 3rd Floor of Salvatori Hall
Data Collection Call:

- Object Point Data
- Physical to logical lines of code ratio
- Data to reliably estimate SLOC from Unadjusted Function Points (UFP):
  - $SLOC = f(UFP, LL, AD)$
  
  where $LL$ is language level and $AD$ is algorithmic density
Help for Data Submission:

- **COCOMO 2.0 Points of Contact**

  For questions on USC COCOMO software, the COCOMO 2.0 Model, data definitions, or project data collection and management, contact:

  - Chris Abts or Brad Clark or Sunita Devnani .... (213) 740-6470
  - Karen Prouten ................................ ........ (213) 740-5703
  - Dr. Barry Boehm ............................ ........ (213) 740-8163
  - Center for Software Engineering FAX ............ .... (213) 740-4927
  - Internet Electronic-Mail ....................... cocomo-info@sunet.usc.edu

- **Affiliate e-mail bulletin board:**
  cocomo-affiliates@sunet.usc.edu

- **Amadeus (Product, Templates, Training):**
  - Phone: 714-725-6400
  - E-Mail: amadeus-info@amadeus.com

Notes: