Language Factors in SEER for Software

The View From 2019
Summary

SEER for Software uses a set of language factors for a large list of programming languages and development frameworks. The recent Agile Planner release includes a major update to these language factors. This presentation will provide an overview of the language factors and how they are developed and updated and trends emerging from the analysis.
How SEER-SEM Uses Language Factors

Language factors are key metrics in generating a SEER-SEM estimate, especially when using functional size metrics.

How Language Factors are Developed

Analysis and derivation of language factors follows a process that is specific to SEER-SEM.

What We Have Observed

Trends observed in this update follow general trends of what we have seen in the past.
New languages and new data require that we review the factors built into SEER-SEM. New languages are added at each release along with minor changes to existing languages. Major updates lead to major review of the language factors and a table for each major version is included with the application.
Functional Size Metrics

Key ideas about FSM

- **Measures Functionality**
  - The Function Point is a normalized measure of software functionality and can be used to measure the functional requirements.

- **Technology Independent**
  - FSM measures what is being delivered, not how. Therefore it can be used to compare projects across different platforms and languages.

- **ISO Standard Metrics**
  - ISO/IEC has five recognized functional size metrics. There are many variations, but the 5 standard ones are widely used.

---

<table>
<thead>
<tr>
<th>SIZE METRIC</th>
<th>Size Metric Description</th>
<th>Fast Function Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>External Inputs (EI)</td>
<td>20 20 30</td>
</tr>
<tr>
<td></td>
<td>External Outputs (EO)</td>
<td>50 55 60</td>
</tr>
<tr>
<td></td>
<td>External Inquiries</td>
<td></td>
</tr>
<tr>
<td></td>
<td>External Interface</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internal Logical Fix</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SIZE METRIC</th>
<th>Size Metric Description</th>
<th>Mark II Function Point Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Input Data Elements (IT)</td>
<td>100 110 120</td>
</tr>
<tr>
<td></td>
<td>Entity References (ER)</td>
<td>18 20 22</td>
</tr>
<tr>
<td></td>
<td>Output Data Elements (OT)</td>
<td>30 30 36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SIZE METRIC</th>
<th>Size Metric Description</th>
<th>Cosmic Function Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CFP - Entries</td>
<td>20 20 22</td>
</tr>
<tr>
<td></td>
<td>CFP - Exits</td>
<td>20 20 24</td>
</tr>
<tr>
<td></td>
<td>CFP - Reads</td>
<td>100 100 100</td>
</tr>
<tr>
<td></td>
<td>CFP - Writes</td>
<td>80 80 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SIZE METRIC</th>
<th>Size Metric Description</th>
<th>NESMA Indicative</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Internal Data (ILF)</td>
<td>3 4 5</td>
</tr>
<tr>
<td></td>
<td>External Data (EIF)</td>
<td>10 12 12</td>
</tr>
</tbody>
</table>

---

**ISO/IEC**

5 recognized functional size metrics are widely used.
Relating Size to Estimated Effort and Schedule

\[ \text{Effort} = D^p \left( \frac{S_e}{C_{te}} \right)^q \]

\[ \text{Schedule} = D^r \left( \frac{S_e}{C_{te}} \right)^s \]

\( S_e \) is the effective size

Formulas shown are a simplification, but do represent the basic form. \( S_e \) is normalized effective size. \( D \) represents the staffing complexity. \( C_{te} \) is the effective technology which is the quantification of the many parameters in SEER-SEM that describe people, process, tools and other factors impactive the overall productivity in the development process.
Use of Language in Calculations

Function Implementation Mechanism is SEER-SEM for Language

Function Implementation Mechanism is the language input in SEER-SEM. When using functional base size metrics, this is a key input for estimating effort. For Lines of Code, this gives an indication of relative volume, thus indirectly impacting effort.

Language Factors \( L_x \) are used to compute a normalized effective size from a functional size metric. AdjFactor represent an adjustments for platform, application and phase at estimate. That normalized size is in turn used to estimate effort.

\[ S_e = L_x \times (\text{AdjFactor} \times UFP)^r \]
The Process of Updating Factors

Recipe for Making Changes

01 Gather Data
Galorath maintains a private repository of nearly 20K projects.

02 Model Using Current Factors
Aided by scripting, generate estimates for all data points.

03 Measure Errors
Generate estimate ratios, compute errors and PRED figures.

04 Compute New Factors
Updated language factors can be computed based on analysis of estimate ratios.

05 Review Changes
Reasonableness checks, judgement review.

06 Implement the Table
Compile figures for use in estimating.
Data Roundup

Historical Project Data in SEER-HD format is an important part of the process. This data has size, effort schedule metrics along with descriptive information such as language, platform and application. Other metrics relating to staffing and defects are also included.

Model Data Points

Modeling in SEER-SEM is accomplished through use of automated scripting. This will be used to make before & after comparisons.
Measure Errors

Estimate Ratios are computed for each observation. Aggregate factors for each language grouping such as PRED(25%) are also computed.

Compute New Factors

Proposed New Language Factors for each language are computed based on the median estimate ratio.

\[ \hat{L}_x = L_x (\text{Eff Adj}) \frac{1}{q} \]
Math Behind New Language Factor

How Estimate Ratios Can Be Applied

Estimated Effort

$$E = D^p \left( \frac{S_e}{C_{te}} \right)^q$$

$$S_e = L_x \times (AdjFactor \times UFP)^r$$

$$\hat{E} = D^p \left( \frac{\hat{S}_e}{C_{te}} \right)^q$$

$$\hat{S}_e = \hat{L}_x \times (AdjFactor \times UFP)^r$$

Estimate Ratio Compares The Before & After

$$EffAdj = \frac{\hat{E}}{E} = \left( \frac{\hat{S}_e}{S_e} \right)^q = \left( \frac{\hat{L}_x}{L_x} \right)^q$$

New Language Factor

$$\hat{L}_x = L_x (EffAdj)^{1/q}$$

Most constants cancel out, making the new **Language Factor** a function of the estimate ratio and the effort entropy.
05

Review Changes
Deep Dive into the results to understand the change. This involved looking different cuts of the data including super and subsets. Lots of pivot tables filtering on data set sub groups. Visual checks against scatterplot data were also examined. When results are compiled, a team review of each language change was conducted.

06

Implement the Table
SEM 8.3 contains the new table, which is automatically used for estimation.
Estimate Continuity
Options for Using Language Factors

New Estimates
SEM 8.3 Language Factors will be used when you create a new estimate.

Legacy Estimates
Earlier Language Factors will be used when you open a legacy estimate. You can change it to use the new table, but it won't happen automatically.

You Can Choose
Language Table Selector is found in the project parameters. You can choose which table to use for any estimate.

We always encourage customers to use the latest version of the factors and knowledge bases offered. However, there are cases and situations where maintaining continuity for an estimate is important. The system is set up to use the new factors for new estimates, but the user has control over what is used. In some cases, customers have tailored the language factor list.
Performance Improvement With Language Updates

SEER-SEM captures the latest trends in relative productivity based on programming language

- Several new languages and many more updated in a comprehensive update and review based on over 16,000 records.

- New languages include AngularJS, Cognos, Groovy, Informatica, Jquery, MSBI, Oracle, Oracle Data Integrator, R, SAP Business Objects, TIBCO and Xquery

- Data-derived updates were judged for their ability to improve statistical measures of estimate performance.

- Updates were reviewed to understand the nature of the change, and to ensure that single sources were not introducing specific bias.
3GL and 4GLs

A trend of convergence?
Looking at the Popular Languages

Looking at 10 most popular languages as indicated by the Tiobe Community Index.
The Study Continues

What we are watching

Low Code Platforms
Low Code Development Platforms offer potential productivity boosts. We have seen some examples of this (Outsystems) and are watching adoption.

Frameworks/Language
Frameworks offer efficiency and play a role in productivity. Frameworks combined with different languages are being looked at.

Stable Workhorses
Minimal Change Detected for Cobol, Ada despite new data.

Maintenance
IS Software Support impacted by language to the same degree as it is for development.
The SEER for Software Team

Who Did This

Steve Acelar
Manager Software/IT Products

Alton Ng
Application Analyst

Karen McRitchie
VP of Product Development

Lee Fischman
Senior Director of Product Development

Svetlana Johnson
Senior Systems Engineer