

# Software Code Growth: A New Approach Based on an Analysis of Historical Actuals

Presented to  
DNI Cost Analysis Improvement Group

18 December 2007

Robert R. Jones  
Paul Hardin

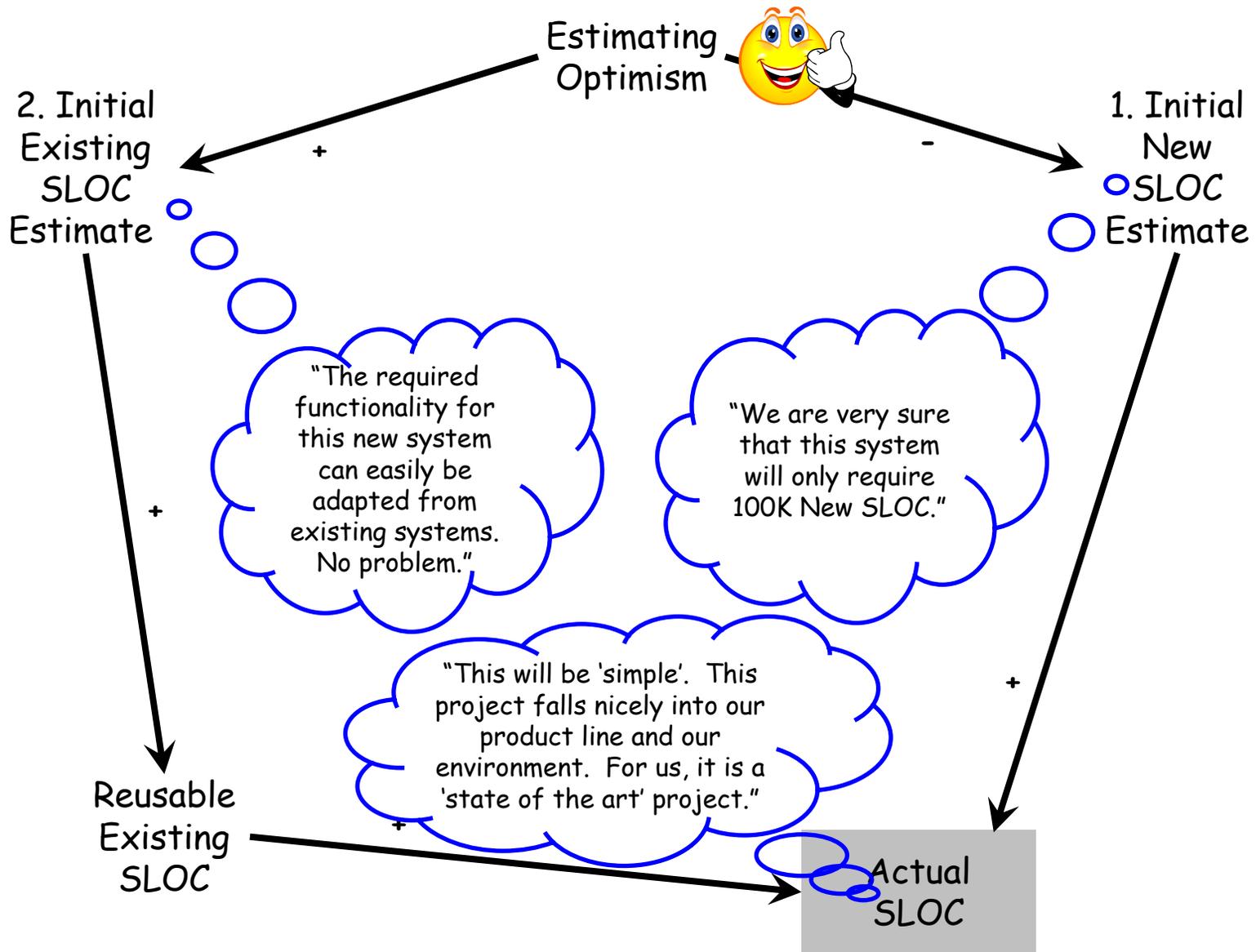
# Outline

1. Objective
2. Background
3. Analytical Hypothesis
4. Data Characteristics
5. Analysis
6. Results
7. Summary

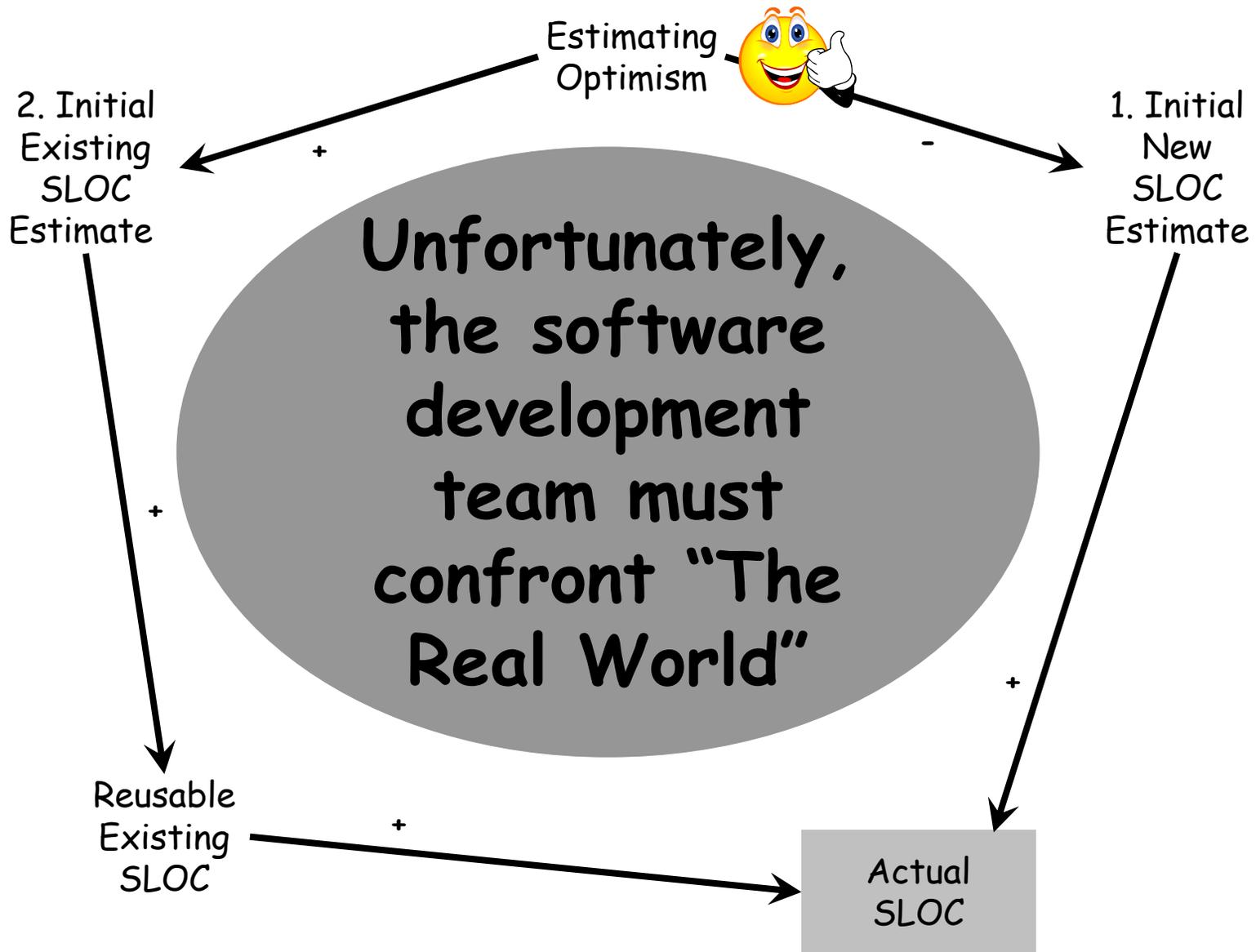
# Background

1. Research shows that the actual software source lines of code (SLOC) and the associated development effort are usually significantly greater than the initial estimates
2. Various techniques already exist to estimate SLOC growth, but our research sought to develop an improved method believing the following to be true:
  - a) SLOC developers under-estimate the amount of New SLOC relative to the total, and
  - b) they over-estimate the expected use of Existing SLOC, i.e., Modified and Unmodified (Reuse) SLOC, and
  - c) they under-estimate the software complexity

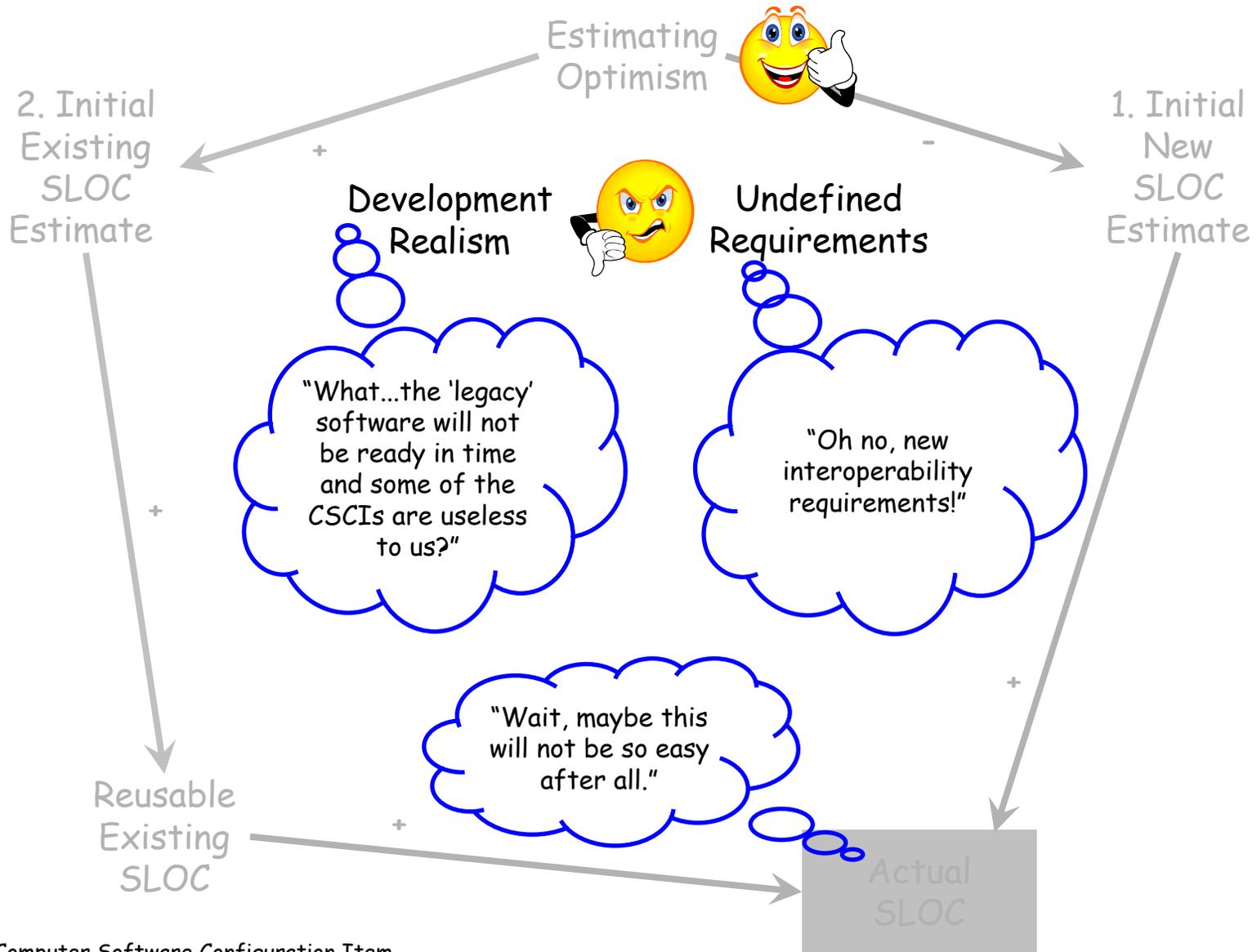
# Analytical Hypothesis, Influence Diagram



# Analytical Hypothesis, Influence Diagram

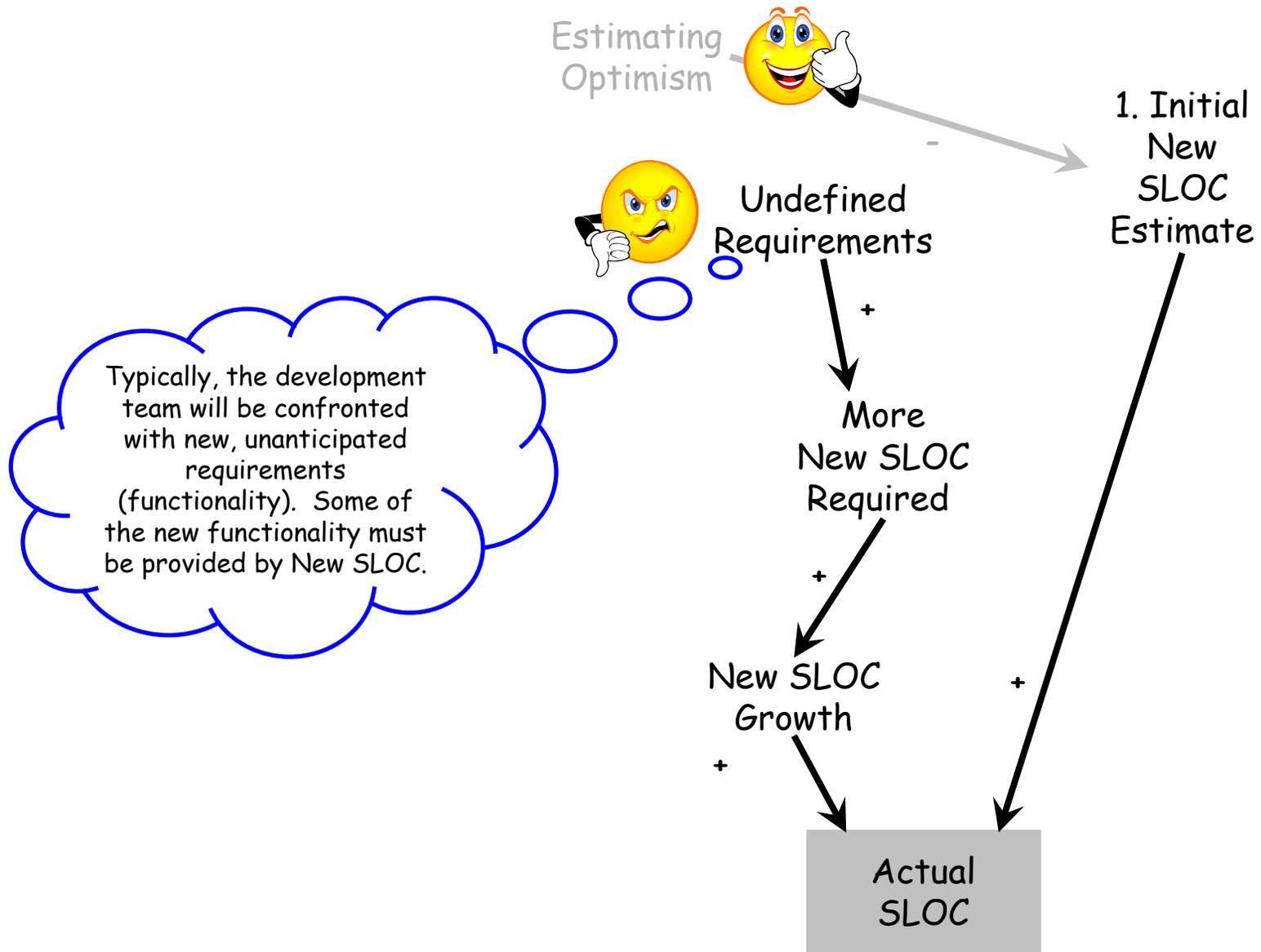


# Analytical Hypothesis, Influence Diagram

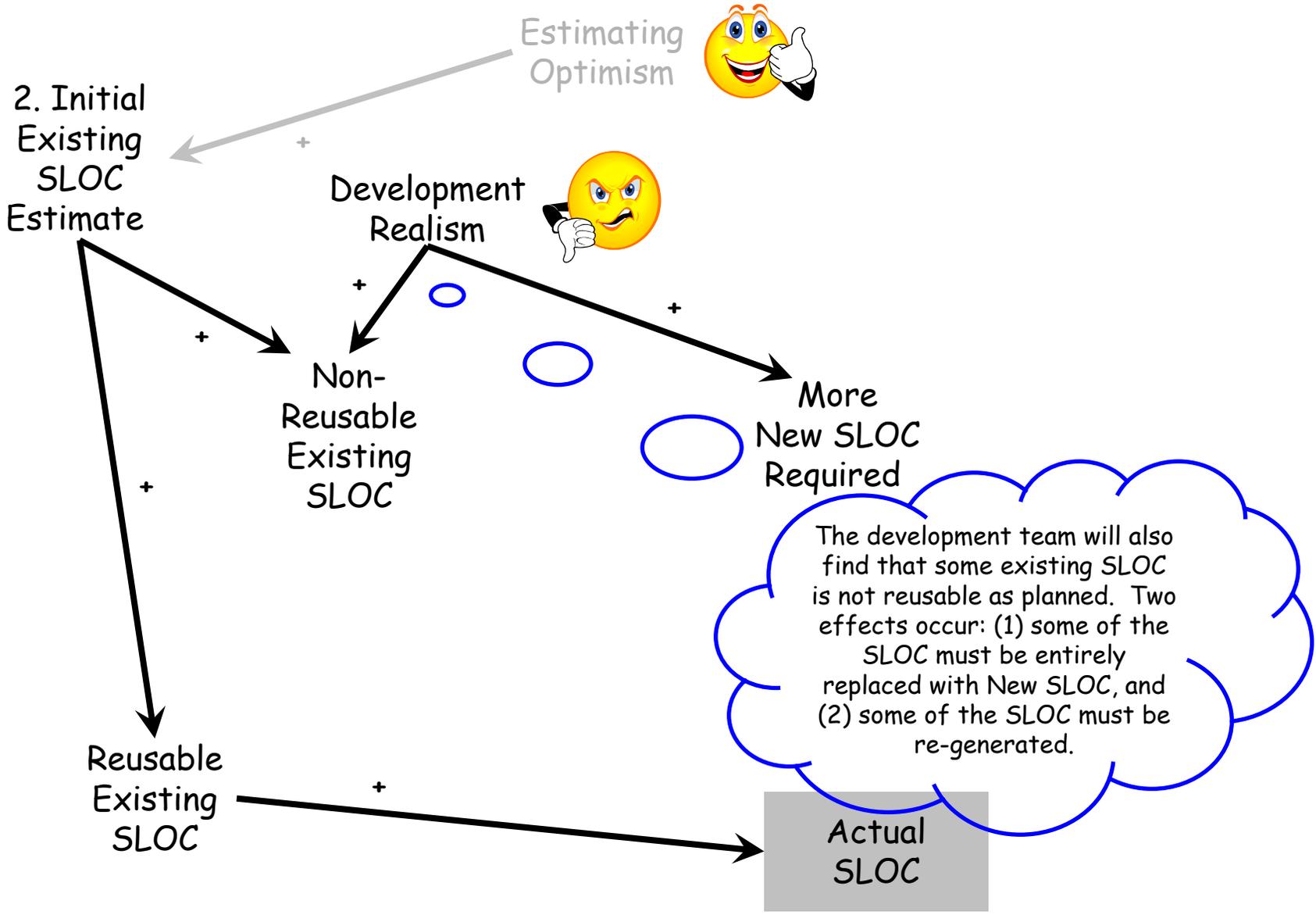


CSCI: Computer Software Configuration Item

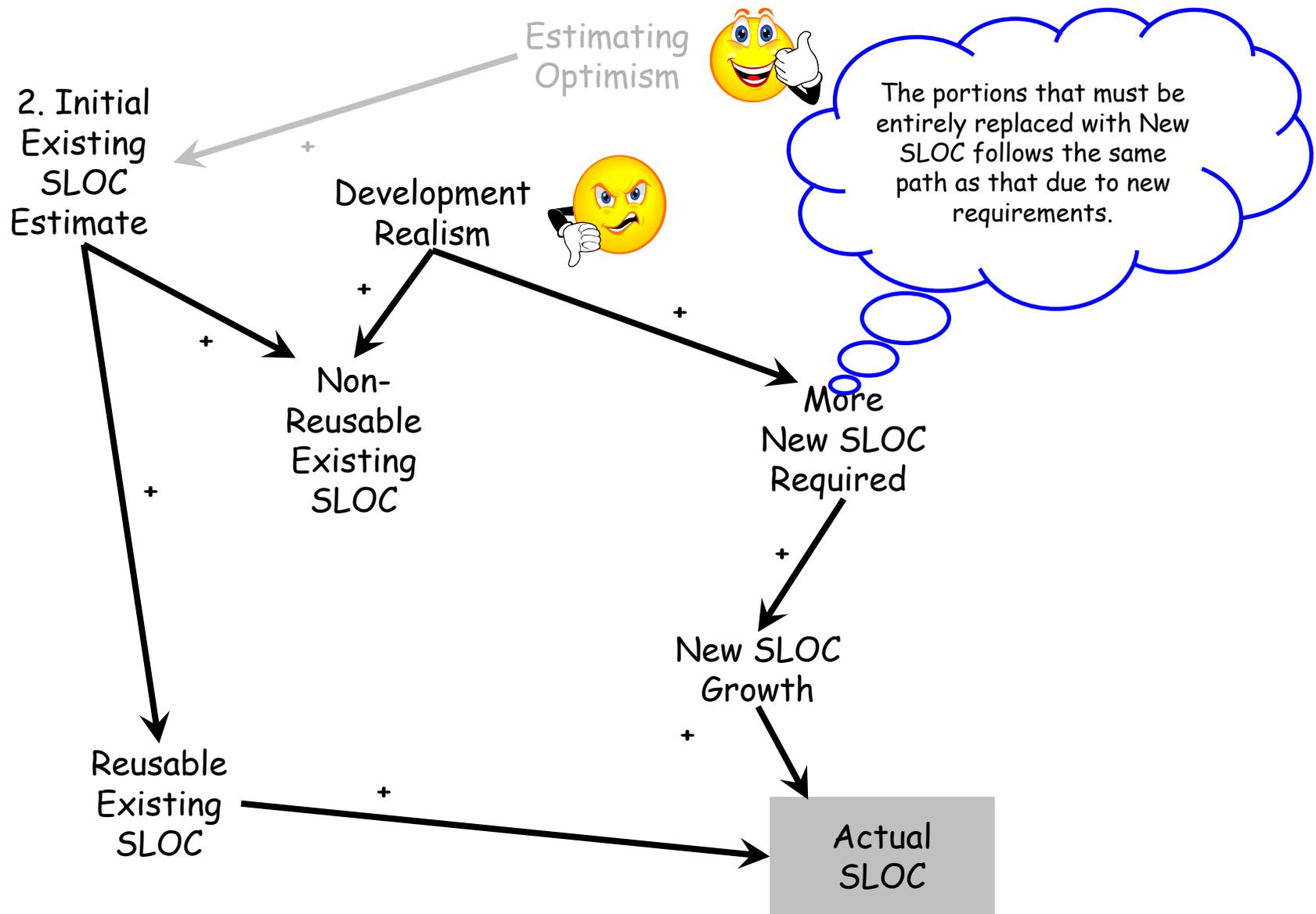
# Analytical Hypothesis, Influence Diagram



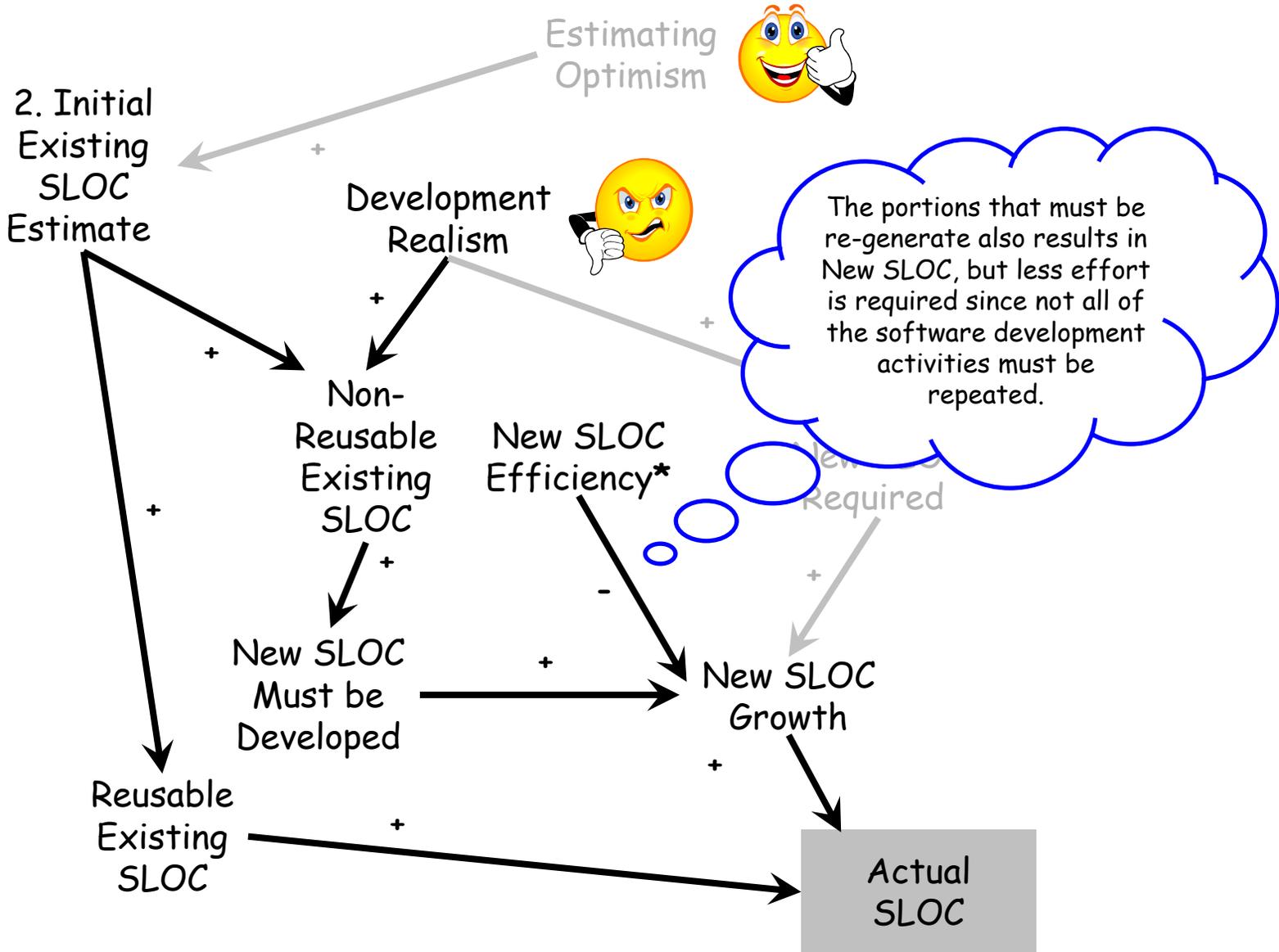
# Analytical Hypothesis, Influence Diagram



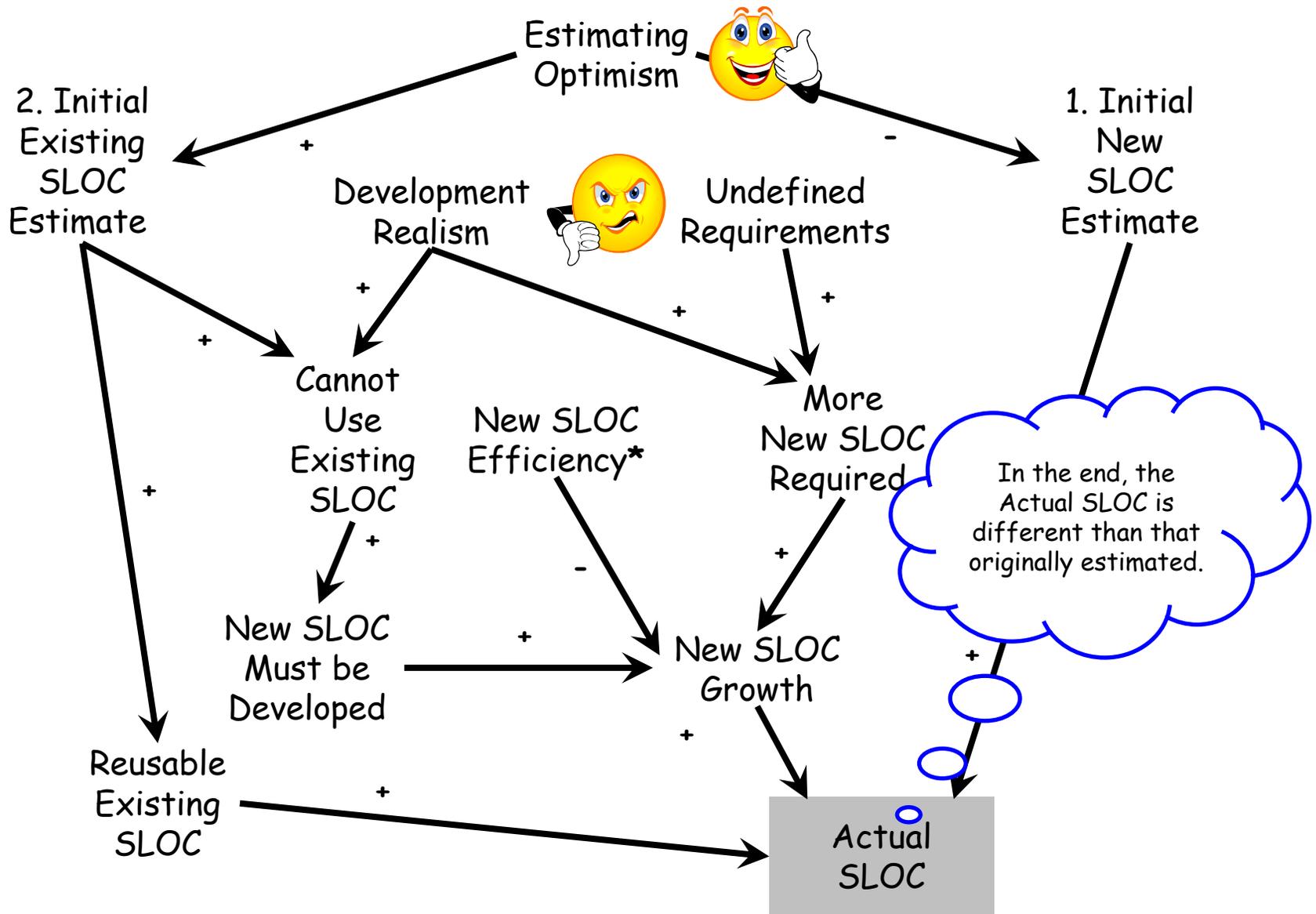
# Analytical Hypothesis, Influence Diagram



# Analytical Hypothesis, Influence Diagram



# Analytical Hypothesis, Influence Diagram



# Analytical Hypothesis, Equation

$$SLOC_{W/growth} = [(1 + a) * New_{Est} + b * (c * Existing_{Est})] + [(1 - c) * Existing_{Est}]$$

New with growth

Existing replaced by New

Existing after loss

New SLOC w/ growth due to Undefined Requirements and Development Realism.

**New<sub>Act</sub>**

New SLOC w/ growth due to Development Realism, i.e., re-generation of existing SLOC.

Re-usable, existing SLOC.

**Existing<sub>Act</sub>**

where

$a$  = New SLOC Growth Factor

$b$  = Coding Efficiency Factor (Efficiency of replacing non-reusable Existing SLOC with New SLOC)

$c$  = Existing SLOC Loss Factor (Fraction of Existing SLOC that is not reusable)

$(1-c)$  = Existing SLOC Reuse Factor (Fraction of Existing SLOC that is reusable)

$Act$  = Actual

$Est$  = Initial estimate

$$SLOC_{Est} = New_{Est} + Existing_{Est} = New_{Est} + (Modified_{Init} + Reuse_{Est})$$

New = newly developed SLOC

Modified = SLOC obtained from another project; reused with minor modifications

Unmodified = SLOC obtained from another project; reused without modifications; aka "Reuse" SLOC

# Data Characteristics

50 Programs

11 Mission types: C2 (31); Database, Diagnostic, Mission Plans, Simulation, Utilities (5); Office Automation, Software Tools, Signal Processing (6); Operating System (3); Testing (5)

3 Development processes: Waterfall (15); Incremental (12); Spiral (8); Undefined (15)

4 Complexities: Simple, Routine, Moderate, Difficult

Values in parentheses = Number of data points

Complexity of the SLOC	Number of Data Points	Product Line	Environment	State of the Art
Simple	2	Existing	Existing	Current
Routine	10	New	Existing	Current
Moderate	14	New	New	Current
Difficult	24	New	New	New

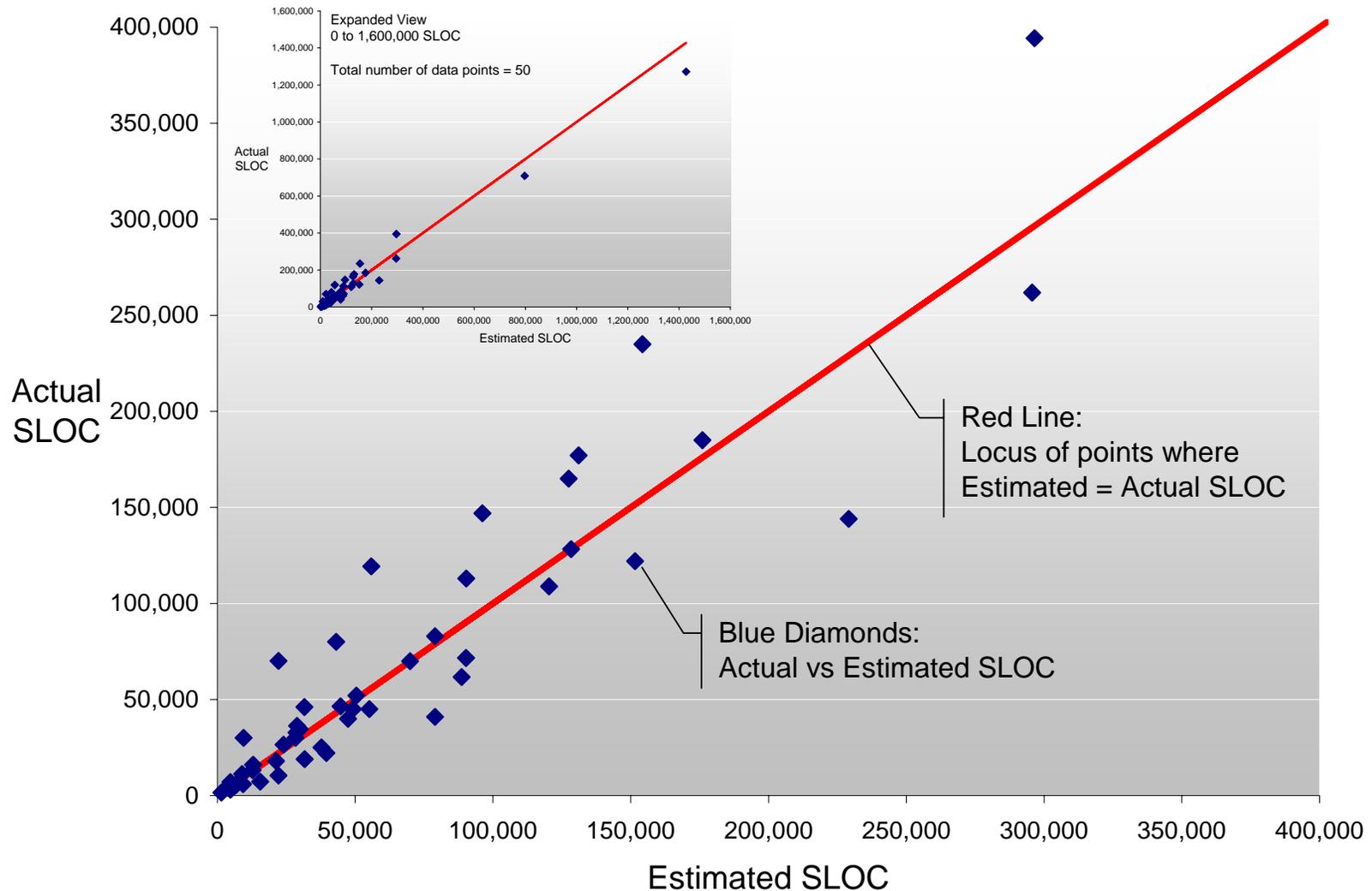
## SLOC breakout:

Category	Source Lines of Code (SLOC)							
	Initial Estimate				Final Actual			
	Minimum	Maximum	Average	Median	Minimum	Maximum	Average	Median
New SLOC	3,000	517,071	94,482	20,000	3,000	519,600	129,507	33,325
Existing SLOC	0	729,201	59,643	1,055	0	752,500	59,721	0
Total SLOC	3,000	1,246,272	154,125	30,227	3,000	1,272,200	189,234	45,034
% New SLOC	25.7%	100.0%	86.6%	90.0%	33.7%	100.0%	87.4%	100.0%

Existing SLOC = Modified SLOC + Unmodified SLOC

# Analysis

The equation parameters,  $a$ ,  $b$ , and  $c$ , are estimated using MS Excel Solver and residual-minimization techniques



# Results: Equation and Parameters

The equation parameter values are

$a = \text{New SLOC Growth Factor}$

$= 0.145$  if the Complexity of the SLOC = Simple

$= 0.290$  if the Complexity of the SLOC = Routine

$= 0.435$  if the Complexity of the SLOC = Moderate

$= 0.580$  if the Complexity of the SLOC = Difficult

$b = \text{Coding Efficiency Factor} = 0.70$

$c = \text{Existing SLOC Loss Factor} = 0.20$

$(1 - c) = \text{Existing SLOC Reuse Factor} = (1 - 0.20) = 0.80$

New SLOC growth  
due to Undefined  
Requirements and  
Development Realism.

Re-usable,  
existing SLOC.

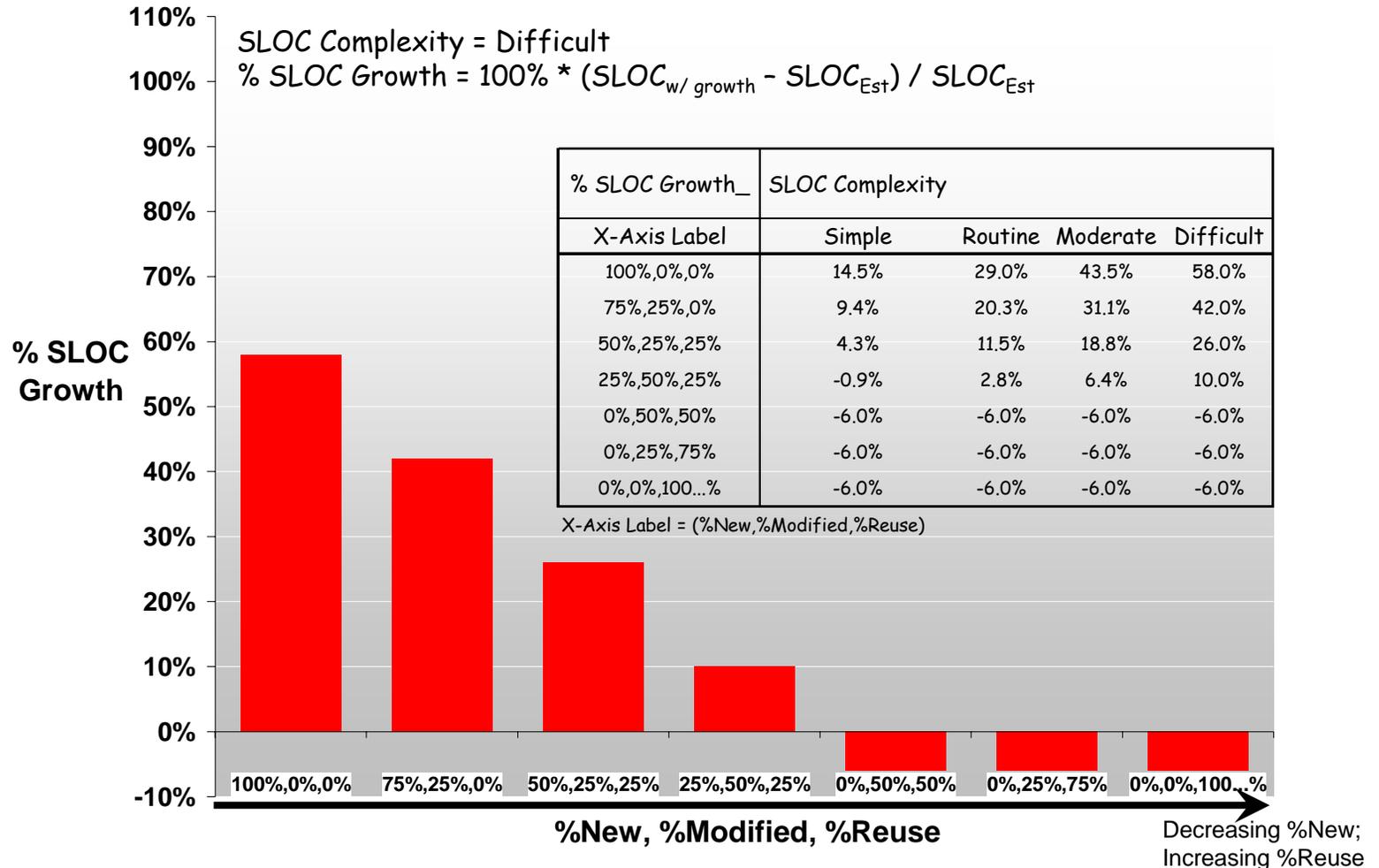
New SLOC growth  
due to Development  
Realism, i.e., re-  
generation of  
existing SLOC.

Equation (1) below estimates the median value for SLOC with growth included

$$SLOC_{w/growth} = [(1 + a) * New_{Est} + 0.70 * (0.20 * Existing_{Est})] + [0.80 * Existing_{Est}]$$

# Results: Exploration of Equation (1)

Here, we systematically explored Equation (1) by varying the % New, % Modified, and % Reuse; all with Complexity = Difficult



# Results: Interpretation of Equation (1)

## 1. New SLOC

- a. Growth is always positive
- b. Growth has two components
  1. New SLOC growth
  2. Replacement of Existing SLOC with New SLOC
- c. Growth is dependent on the Complexity of the SLOC, i.e., Simple, Routine, Moderate, or Difficult
- d. Growth ranges from 14.5% to 58.0%

## 2. Existing SLOC

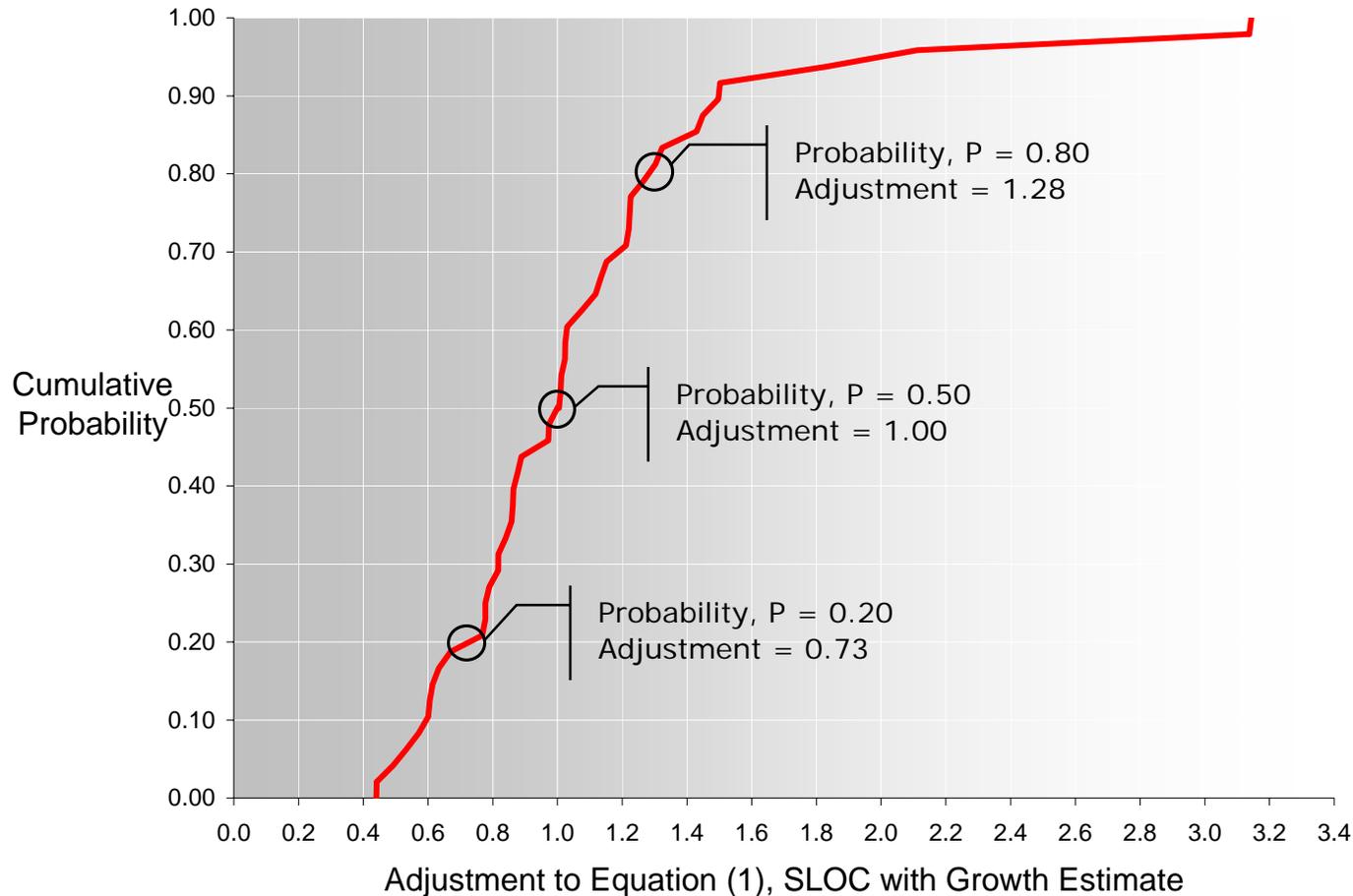
- a. Reusability is 20% less than expected
- b. Non-reusable Existing SLOC is replaced by New SLOC using 30% fewer lines of code

## 3. Total SLOC: Growth range is -6% to 58%

# Results: Risk Analysis

Via analysis of the residuals, risk-based adjustments can be made to the SLOC with growth estimate

For example:  $P = 0.80$ , SLOC with growth  $\leq 1.28 * \text{Equation (1)}$



# Results: Effects of Equation (1) on ESLOC

Here, we systematically explored Equation (1) by varying the % New, % Modified, and % Reuse, all with Complexity = Difficult, and we calculated ESLOC using the following Weighting Factors:

**New=1.00** (2x Modified, 8.9x Reuse); **Modified=0.50** (4.4x Reuse); **Reuse=0.1125**

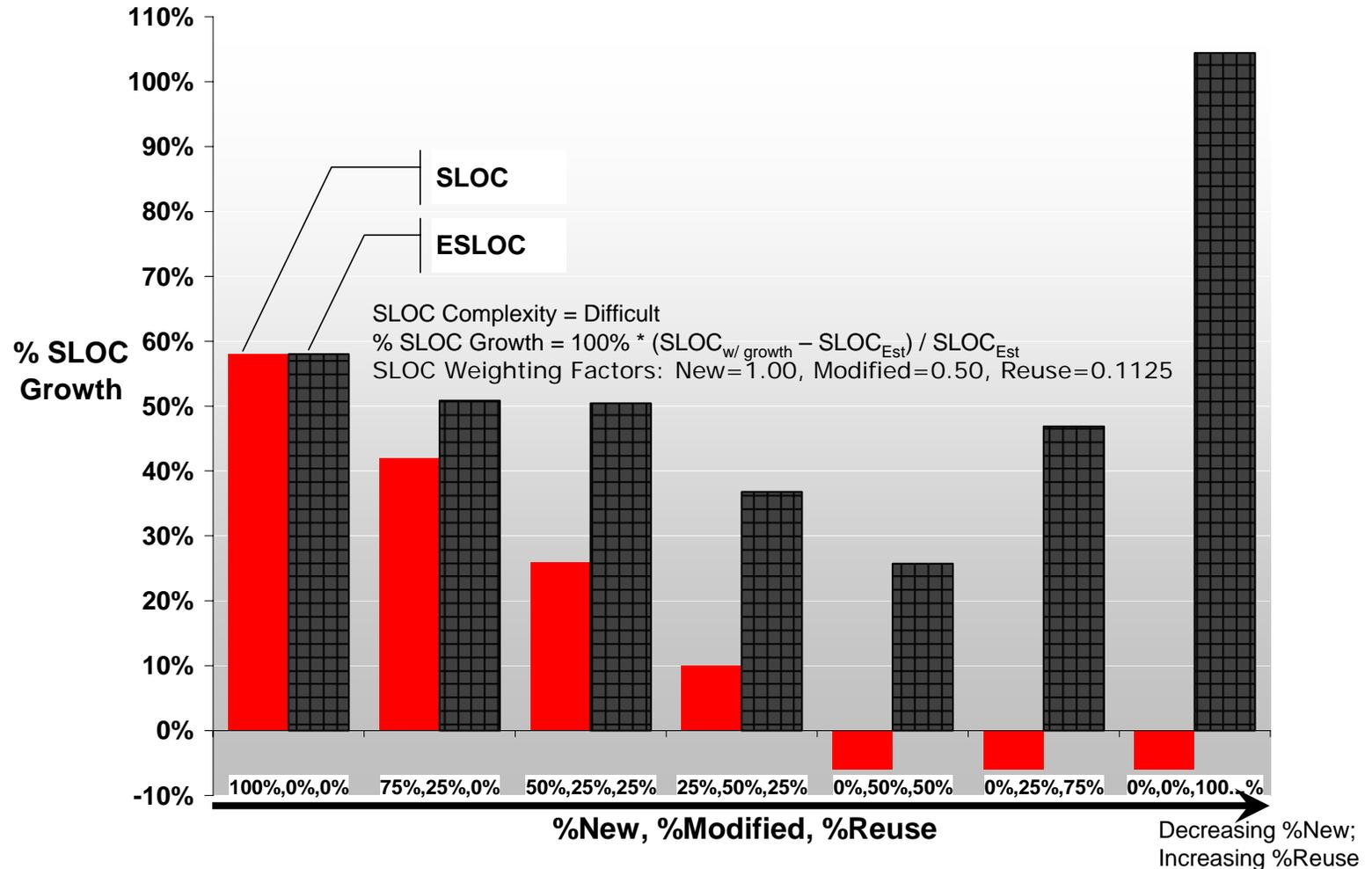
		SLOC Complexity			
X-Axis Label	Data	Simple	Routine	Moderate	Difficult
100%,0%,0%	% SLOC Growth_	14.5%	29.0%	43.5%	58.0%
	% ESLOC Growth_	14.5%	29.0%	43.5%	58.0%
75%,25%,0%	% SLOC Growth_	9.4%	20.3%	31.1%	42.0%
	% ESLOC Growth_	13.6%	26.0%	38.4%	50.9%
50%,25%,25%	% SLOC Growth_	4.3%	11.5%	18.8%	26.0%
	% ESLOC Growth_	17.1%	28.2%	39.3%	50.4%
25%,50%,25%	% SLOC Growth_	-0.9%	2.8%	6.4%	10.0%
	% ESLOC Growth_	16.2%	23.1%	29.9%	36.8%
0%,50%,50%	% SLOC Growth_	-6.0%	-6.0%	-6.0%	-6.0%
	% ESLOC Growth_	25.7%	25.7%	25.7%	25.7%
0%,25%,75%	% SLOC Growth_	-6.0%	-6.0%	-6.0%	-6.0%
	% ESLOC Growth_	46.9%	46.9%	46.9%	46.9%
0%,0%,100...%	% SLOC Growth_	-6.0%	-6.0%	-6.0%	-6.0%
	% ESLOC Growth_	104.4%	104.4%	104.4%	104.4%

X-Axis Label = (%New,%Modified,%Reuse)

# Results: Effects of Equation (1) on ESLOC

ESLOC growth, i.e., effort growth, will always be positive because

1. New SLOC positive growth, and New SLOC is required to replace Existing SLOC
2. SLOC to ESLOC Weighting Factors:  $1.00 > 0.50 \gg 0.1125$



# Summary

## 1. Analysis conclusions

- a) SLOC growth is driven by (a) the Complexity of the SLOC (usually under-estimated), and (b) the distribution of New, Modified, and Reuse SLOC
- b) New SLOC is usually under-estimated
- c) Re-usability of Existing SLOC is usually over-estimated
- d) SLOC growth can be positive or negative
- e) ESLOC and effort growth are always positive

## 2. SLOC with growth and % SLOC growth can be estimated using Equation (1)

$$SLOC_{w/growth} = [(1 + a) * New_{Est} + 0.70 * (0.20 * Existing_{Est})] + [0.80 * Existing_{Est}]$$

$$\% SLOC Growth = 100\% * (SLOC_{w/growth} - SLOC_{Est}) / SLOC_{Est}$$

## 3. Probabilities can be assigned to the SLOC with growth estimate

## 4. Using weighting factors, a SLOC with growth estimate can easily be transformed into % ESLOC or % effort growth

- a) Growth will always be positive
- b) Initial estimates with 100% Reuse SLOC are the highest-risk (growth) option