

Performance Calculation and Benchmarking using the ISBSG Release 10 Data Repository

Luigi Buglione & Alain Abran



**5th Software Measurement European Forum
28-30 May 2008, Milan, Italy**

Agenda



- **Introduction**
- **Multi-Dimensional Performance Models**
 - ✓ QEST-LIME models: description
 - ✓ QEST-LIME models & Performance Targets
- **The ISBSG r10 Data Repository**
 - ✓ Available Fields
 - ✓ Selection of data samples in the ISBSG repository r10
 - ✓ Selection of the common set of measures
 - ✓ Overall Performance Levels
- **Analyzing results from performance benchmarking models**
 - ✓ Data flow for analysis of results
 - ✓ Presentation of results from ISBSG r10 data
- **Conclusions**



Introduction

State-of-the-art



- Productivity is a typical figure used for estimating effort and costs for a project...
 - ✓ **Productivity**: the amount of output produced per unit input used
 - ✓ It's a ML2 concept in most known Maturity Models (i.e. CMMI PP/PMC)
- ...but Productivity \supset Performance
 - ✓ **Performance**: the degree to which a system or a component accomplishes its designated functions within given constraints (IEEE std-610.12-1990 (R2002))
 - ✓ It's a ML4 concept in most known Maturity Models (i.e. CMMI OPP/QPM)
- Well-known and recognized sources of information are...
 - ✓ **ISO/IEC 15939:2002** (Measurement Process)
 - ✓ **CMMI MA** (Measurement & Analysis)
 - ✓ **ISO/IEC IS 15504-5:2006 ORG.5** (Measurement)



Introduction

Measurement & Historical Data



- **ISO 15939 / CMMI**

- ✓ They cover – with few differences (e.g. environment) - all the phases and activities for successfully implementing a measurement program **but** they do not provide measures, only guidance to define measures addressing specific needs

- **ISBSG r10**

- ✓ It does not provide any process, **but** provides measures and historical data from more than 4100 projects that could be useful for making ISO 15939/CMMI more effective.

Q: ...so: for practical use in industry for more efficient measurement and better decision making, can we combine these best practices?



Agenda



- **Introduction**
- **Multi-Dimensional Performance Models**
 - ✓ QEST-LIME models: description
 - ✓ QEST-LIME models & Performance Targets
- **The ISBSG r10 Data Repository**
 - ✓ Available Fields
 - ✓ Selection of data samples in the ISBSG repository r10
 - ✓ Selection of the common set of measures
 - ✓ Overall Performance Levels
- **Analyzing results from performance benchmarking models**
 - ✓ Data flow for analysis of results
 - ✓ Presentation of results from ISBSG r10 data
- **Conclusions**

Multi-dimensional Performance Models

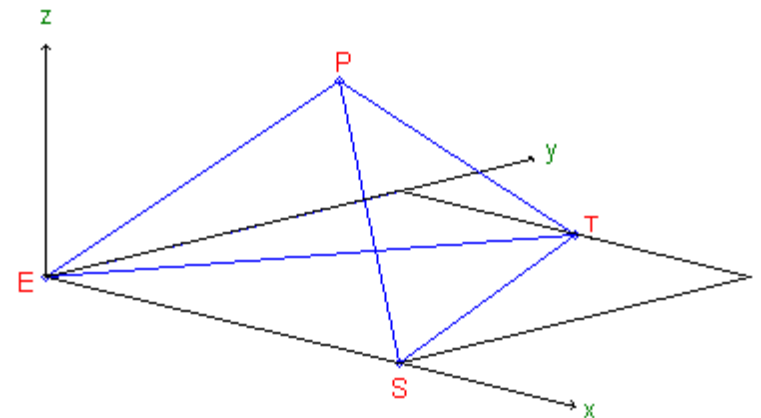


The QEST Model

- **Method:** Performance is expressed as the combination of the specific ratios selected for each of the 3 dimensions of the *quantitative* assessment (Productivity - **PR**) and the perceived product quality level of the *qualitative* assessment (Quality - **Q**)

$$\text{Performance} = \text{PR} + \text{Q}$$

- **Model:** QEST (Quality factor + Economic, Social & Technical dimensions) is a “structured shell” to be filled according to management objectives in relation to a specific project.
Such a model has the ability to handle independent sets of dimensions without predefined ratios and weights - referred to as an *open model*



Multi-dimensional Performance Models

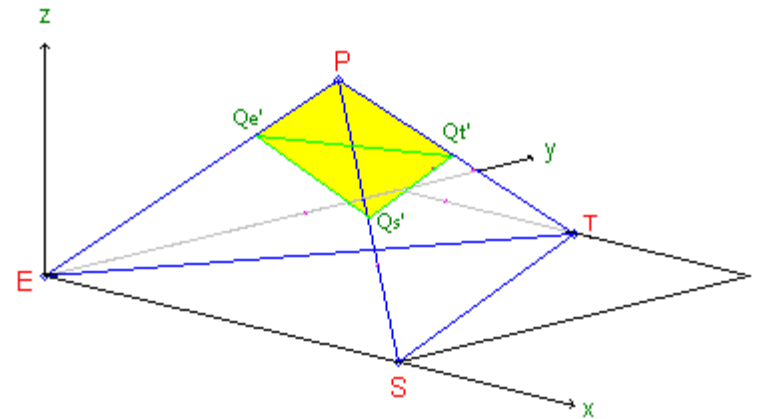
The QEST Model – Geometrical Indicators



- **Target:** measuring project performance (p) using 3 distinct viewpoints
- **Input Data:** list of weighted ratios for each dimension and quality questionnaires
- **Output Data:** an integrated normalized value of performance

It is possible to measure performance considering at least 3 distinct geometrical concepts:

- ◆ **Distance** between the tetrahedron base center of gravity and the center of the plane section along the tetrahedron height – the greater the distance from 0, the higher the performance level;
- ◆ **Area** of the sloped plane section – the smaller the area, the higher the performance level;
- ◆ **Volume** of the lowest part of the truncated tetrahedron – the greater the volume, the higher the performance level.



Multi-dimensional Performance Models

The QEST Model – Key Features



- Integrated quantitative and qualitative evaluation from 3 concurrent organisational viewpoints
- a 3D geometrical representation at a single project phase (usually after the project is completed)
- Use of *de facto* and *de jure* standards (e.g. ISO/IEC 9126 for the Quality Factor)
- Extension of the original 3D model to n possible dimensions-perspectives → QEST nD through the *simplex* as the mechanism to solve the problem from the 4th dimension on
- Performance Measurement Model to use for consolidating Balanced Scorecard (BSC) measurement outcomes



Multi-dimensional Performance Models

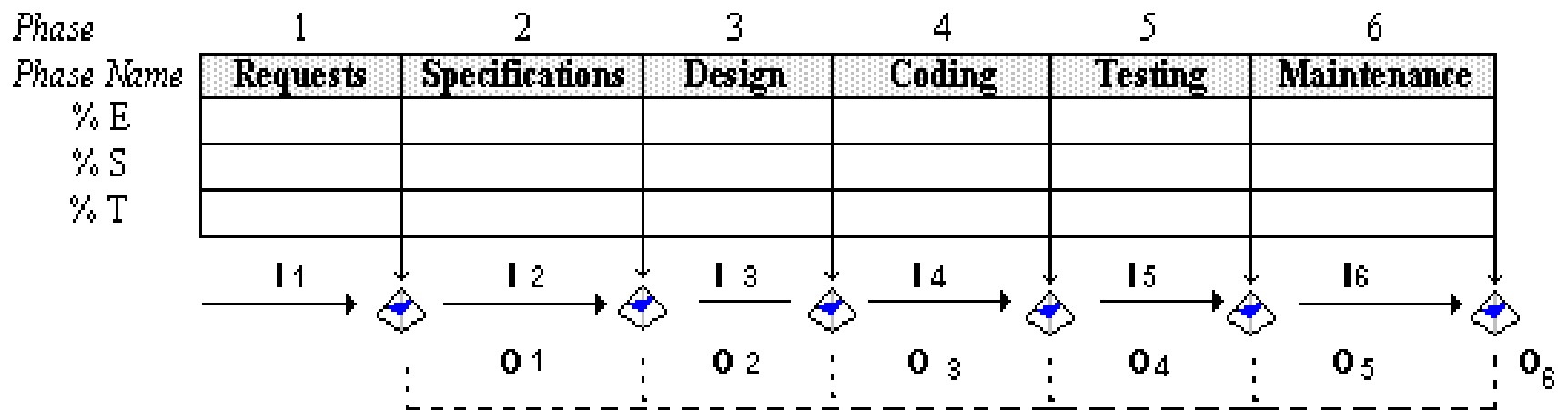


The LIME Model

LIME (**L**ifecycle **M**Easurement) model represents the extension of QEST features to a dynamic context as the SLC is.

SLC model selected: generic 6-steps Waterfall model

Logic adopted: the same than in the **ETVX** (Entry-Task-Validation-eXit) process notation

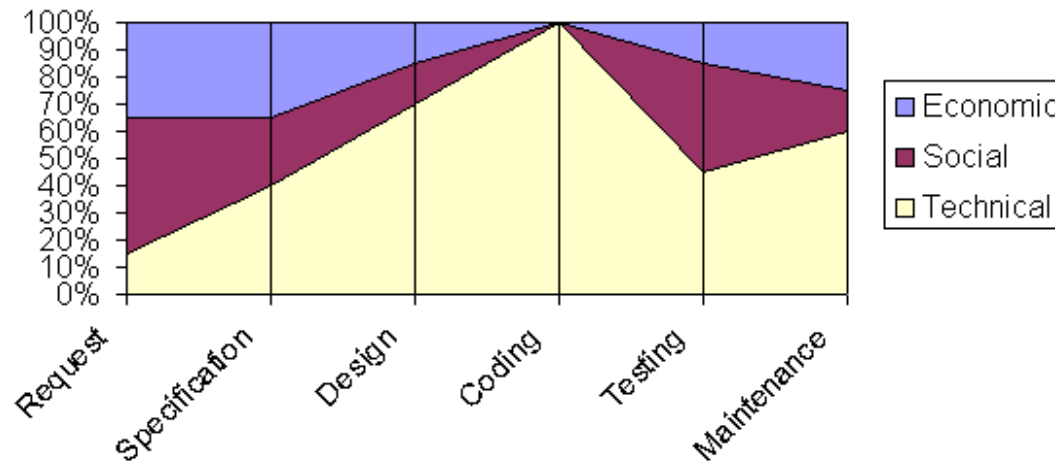


Multi-dimensional Performance Models

The LIME Model – Key Features



- 1 Flexibility of distinct relative contributions from the 3 dimensions (E, S, T) in each phase



- 2 Flexibility of distinct relative contributions of between quantitative and qualitative evaluations in each phase
- 3 Different sources for QF calculation
- 4 Flexibility in selecting measures and ratios suitable for each SLC phase

LIME was also extended also to Risk Management → **R-LIME**

Multi-dimensional Performance Models

The QEST/LIME Models & Performance Targets



- p is the performance value coming from QEST/LIME models
 - from QEST → entity: *project*
 - from LIME → entity: *SDLC phase*
- ...and it can be used for estimating next performance:

$$p_i = f(x_{1i}, x_{2i}, \dots, x_{ni}) \quad \text{For the } i\text{-th phase, from } n \text{ possible ratios}$$

$$p_{i+1} = f(p_1, p_2, \dots, p_i) \quad \text{For the } (i+1)\text{-th phase, from past phases}$$

- Once derived the $p_{(i+1)}$ values, it will be possible to use them for cost estimation (as requested in CMMi PP SP1.4-1)



Agenda



- **Introduction**
- **Multi-Dimensional Performance Models**
 - ✓ QEST-LIME models: description
 - ✓ QEST-LIME models & Performance Targets
- **The ISBSG r10 Data Repository**
 - ✓ Available Fields
 - ✓ Selection of data samples in the ISBSG repository r10
 - ✓ Selection of the common set of measures
 - ✓ Overall Performance Levels
- **Analyzing results from performance benchmarking models**
 - ✓ Data flow for analysis of results
 - ✓ Presentation of results from ISBSG r10 data
- **Conclusions**



The ISBSG r10 Data Repository

Available Fields

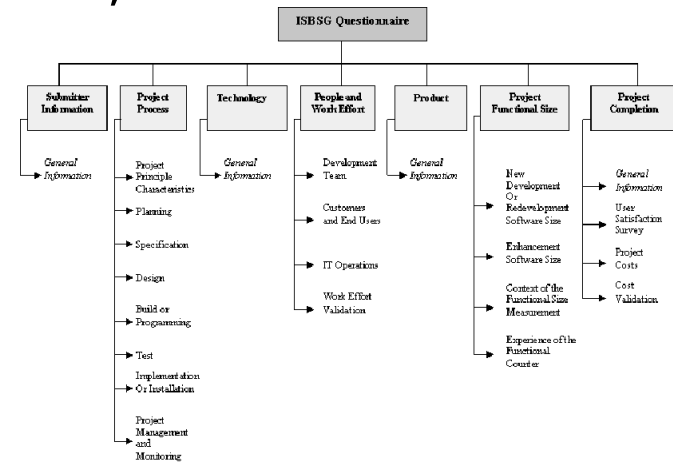


- **ISBSG (www.isbsg.org)**

- ✓ Non-profit organization created in 1994,
- ✓ Goal: to develop the profession of software measurement by establishing a common vocabulary and understanding of terms
- ✓ Data repository (quite) yearly produced (current release: r10)
- ✓ Organizational and Technical Data gathered
- ✓ Questionnaire with 7 sections, 131 questions, 7 sections

- **Advantages using ISBSG r10**

- ✓ Helps in faster implementations of the software measurement process
- ✓ Data from more than 4100 projects, both development and enhancement ones



The ISBSG r10 Data Repository

Selection of Data Samples: quantitative fields



Group	Measure	Description	Id.	#
Sizing	Functional Size	<i>Unadjusted Function Points</i>	UFP	m01
Effort	Normalized Work Effort	<i>Full life-cycle effort for all teams reported</i>	NWE	m02
Productivity	Normalized PDR (<i>ufp</i>)	<i>Normalized Productivity Delivery Rate in hours per functional size unit</i>	NPDR	m03
Schedule	Project Elapsed Time	<i>Total elapsed time for the project in calendar-months</i>	PET	m04
Quality	Total Defects Delivered	<i>Total number of defects reported in the first month of use of the software</i>	TDD	m05
Effort Attributes	Max Team Size	<i>The maximum number of people who worked at any time on the project (peak team size)</i>	MTS	m06
	Average Team Size	<i>The average number of people who worked on the project</i>	ATS	m07
Size Attributes (IFPUG/NESMA/Mark II)	Input Count	<i>The unadjusted size of external inputs</i>	EI	m08
	Output Count	<i>The unadjusted size of external outputs</i>	EO	m09
	Enquiry Count	<i>The unadjusted size of external inquiries</i>	EQ	m10
	File Count	<i>The unadjusted size of internal logical files</i>	ILF	m11
	Interface Count	<i>The unadjusted size of external interface files</i>	EIF	m12
Size Attributes (all methods)	Added	<i>The unadjusted size of additions</i>	UFPA	m13
	Changed	<i>The unadjusted size of changes</i>	UFPC	m14
	Deleted	<i>The unadjusted size of deletions</i>	UFPD	m15
Size (other than FSM)	Lines of Code	<i>The number of the source lines of code (SLOC) produced by the project</i>	LOC	m16



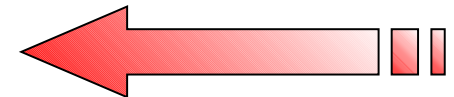
The ISBSG r10 Data Repository

Selection of Data Samples: IFPUG Projects



Step	Attribute	Filter	Projects Excluded	Remaining Projects
1	Functional Sizing Approach	= IFPUG	825	3281
2	Data Quality Rating (DQR)	= {A B}	177	3104
3	UFP Rating	= {A B}	313	2791
4	Total Defects Delivered	= {non-blanks}	2276	515
5	IFPUG BFC (EI, EO, EQ, ILF, EIF)	= {non-blanks}	269	246
6	Project Elapsed Time	= {non-blanks}	9	237
7	Max Team Size	= {non-blanks}	76	161
8	Average Team Size	= {non-blanks}	122	39
9	Development Type	= {New Development}		18
		= {Enhancement}		16
		= {Re-development}		5

<i>Dev. Type / FSM method</i>	IFPUG
New Development	A (18 projects)
Enhancement	B (16 projects)



The ISBSG r10 Data Repository

Selection of the common set of measures



Ratio	Id.	Description	E	S	T
m1/m2	UFP/NWE	Productivity	x		x
m5/m1	TDD/UFP	Defects/FP			x
m6/m7	MTS/ATS	Max team size / avg team	x	x	
m2/m4	NWE/PET	Schedule Compression Index	x		
m5/m4	TDD/PET	Defects/Prj Elaps.Time		x	x
m1/m4	UFP/PET	Duration Delivery Rate	x		x
(m11+m12)/m1	(ILF+EIF)/UFP*100	% file / UFP [only for IFPUG projects]	x		
(m08+m09+m10)/m1	(EI+EO+EQ)/UFP*100	% transactions / UFP [only for IFPUG projects]	x		

P012 (NewDev)	RATIO	NAME	E	S	T	Rmin	Rmax	Abs Value	R Value
m1/m2	UFP/NWE	Productivity	x		x	0.0	0.7	0.0	0.0
m5/m1	TDD/UFP	Defects/FP			x	0.0	0.1	0.03	0.3
m6/m7	MTS/ATS	Max team size / avg team	x	x		1.00	2.2	1.6	0.5
m2/m4	NWE/PET	Schedule Compression Index	x			50.7	14938.0	1712.0	0.1
m5/m4	TDD/PET	Defects/Prj Elaps.Time		x	x	0.00	7.9	1.2	0.2
m1/m4	UFP/PET	Duration Delivery Rate	x		x	14.0	1326.0	43.3	0.0
(m11+m12)/m1	(ILF+EIF)/UFP*100	% file / UFP	x			0.2	0.6	0.5	0.8
(m08+m09+,10)/m1	(EI+EO+EQ)/UFP*100	%transactions / UFP	x			0.4	0.8	0.5	0.2

The ISBSG r10 Data Repository

Selection of the common set of measures: NewDev (proj P012)



		Weight (1)	R values (2)	Final Values (3=1*2)	Acceptability Threshold (4)	Delta Values (5=3-4)
E				0.14	0.33	-0.18
<i>e1</i>	Productivity	0.45	0.02	0.01	0.03	-0.02
<i>e2</i>	Max team size / avg team	0.05	0.47	0.02	0.01	0.01
<i>e3</i>	Schedule compress. Index	0.15	0.11	0.02	0.01	0.01
<i>e4</i>	Duration Delivery Rate	0.20	0.02	0.00	0.01	-0.00
<i>e5</i>	% file / UFP	0.10	0.80	0.08	0.03	0.04
<i>e6</i>	%transactions / UFP	0.05	0.20	0.01	0.03	-0.02
		1.00				
S				0.21	0.25	-0.03
<i>s1</i>	Max team size / avg team	0.20	0.47	0.09	0.07	0.03
<i>s2</i>	Defects/Prj Elaps.Time	0.80	0.15	0.12	0.06	0.06
		1.00				
T				0.12	0.42	-0.30
<i>t1</i>	Productivity	0.35	0.02	0.00	0.03	-0.02
<i>t2</i>	Defects/FP	0.25	0.32	0.08	0.02	0.06
<i>t3</i>	Defects/Prj Elaps.Time	0.20	0.15	0.03	0.02	0.01
<i>t4</i>	Duration Delivery Rate	0.20	0.02	0.00	0.01	-0.00
		1.00				

$$P_e = 0.14; P_s = 0.21; P_t = 0.123$$

p Formula:

$$p = 1 - \prod_{i=1}^n (1 - p_i) = 0.41$$



The ISBSG r10 Data Repository

Overall Performance Levels (IFPUG, NewDev)



	P_e	P_s	P_t				
	E	S	t	$1-p_e$	$1-p_s$	$1-p_t$	P
P001	0.27	0.01	0.17	0.73	0.99	0.83	0.40
P002	0.12	0.13	0.10	0.88	0.87	0.90	0.31
P003	0.14	0.18	0.24	0.86	0.82	0.76	0.46
P004	0.14	0.07	0.02	0.86	0.93	0.98	0.21
P005	0.22	0.10	0.23	0.78	0.90	0.77	0.46
P006	0.12	0.07	0.03	0.88	0.93	0.97	0.20
P007	0.16	0.88	0.50	0.84	0.12	0.50	0.95
P008	0.23	0.14	0.18	0.77	0.86	0.82	0.46
P009	0.12	0.03	0.05	0.88	0.97	0.95	0.18
P010	0.48	0.14	0.27	0.52	0.86	0.73	0.68
P011	0.12	0.10	0.06	0.88	0.90	0.94	0.27
P012	0.15	0.21	0.12	0.85	0.79	0.88	0.41
P013	0.12	0.12	0.05	0.88	0.88	0.95	0.27
P014	0.12	0.11	0.14	0.88	0.89	0.86	0.33
P015	0.16	0.28	0.09	0.84	0.72	0.91	0.45
P016	0.15	0.23	0.04	0.85	0.77	0.96	0.37
P017	0.55	0.16	0.36	0.45	0.84	0.64	0.76
P018	0.27	0.07	0.19	0.73	0.93	0.81	0.44
						Max	0.95
						Median	0.41
						Average	0.42
						Min	0.181

The ISBSG r10 Data Repository

Overall Performance Levels (IFPUG, Enh)



	P_e	P_s	P_t				
	E	s	t	$1-p_e$	$1-p_s$	$1-p_t$	P
P001	0.16	0.24	0.15	0.84	0.76	0.85	0.45
P002	0.07	0.06	0.17	0.93	0.94	0.83	0.27
P003	0.18	0.13	0.10	0.82	0.87	0.90	0.36
P004	0.11	0.07	0.14	0.89	0.93	0.86	0.29
P005	0.04	0.00	0.02	0.96	1.00	0.98	0.06
P006	0.07	0.00	0.05	0.93	1.00	0.95	0.12
P007	0.15	0.23	0.09	0.85	0.77	0.91	0.41
P008	0.21	0.43	0.36	0.79	0.57	0.64	0.71
P009	0.09	0.00	0.04	0.91	1.00	0.96	0.13
P010	0.73	0.87	0.82	0.27	0.13	0.18	0.99
P011	0.06	0.00	0.03	0.94	1.00	0.97	0.08
P012	0.08	0.03	0.07	0.92	0.97	0.93	0.17
P013	0.04	0.00	0.01	0.96	1.00	0.99	0.04
P014	0.12	0.18	0.20	0.88	0.82	0.80	0.43
P015	0.16	0.10	0.10	0.84	0.90	0.90	0.31
P016	0.34	0.00	0.27	0.66	1.00	0.73	0.51
						Max	0.76
						Median	0.44
						Average	0.95
						Min	0.41

Agenda



- **Introduction**
- **Multi-Dimensional Performance Models**
 - ✓ QEST-LIME models: description
 - ✓ QEST-LIME models & Performance Targets
- **The ISBSG r10 Data Repository**
 - ✓ Available Fields
 - ✓ Selection of data samples in the ISBSG repository r10
 - ✓ Selection of the common set of measures
 - ✓ Overall Performance Levels
- **Analyzing results from performance benchmarking models**
 - ✓ Data flow for analysis of results
 - ✓ Presentation of results from ISBSG r10 data
- **Conclusions**



Analyzing Results

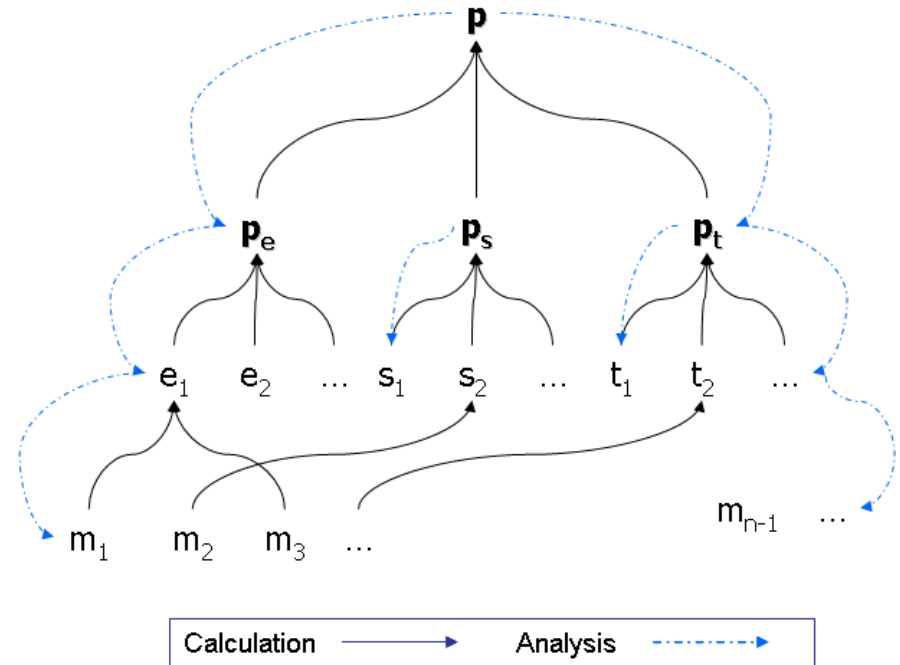
Data Flow for Analysis of Results with QEST model



- Hp: in the NewDev dataset, $P002=0.26$ and $P003=0.42$
- In order to compare the two projects' performance, we have to follow the "Analysis" path (top-down)
- In this case, analyzing in detail projects' data and ratios, some high-level hypothesis could be:

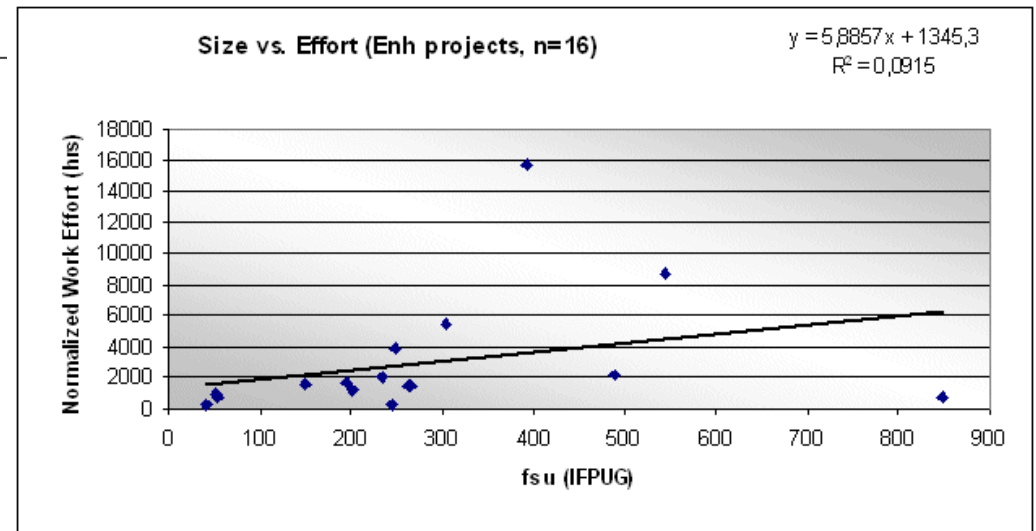
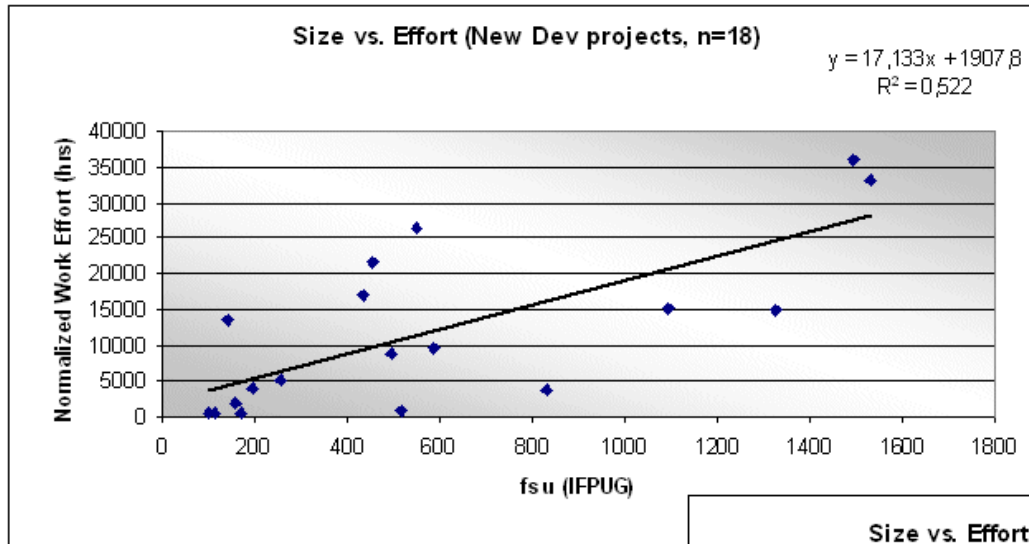
✓ different project productivity levels (e_1 and t_1 ratios) → due to a people issue or to a not homogeneous distribution between F vs NF effort → P002 seems to have more environments to which be connected than P003

✓ same elapsed time, but fewer people working on the project, generating a higher probability for defects and pressure to deliver on time for P003 (half of the people, just 1 calendar-month more)



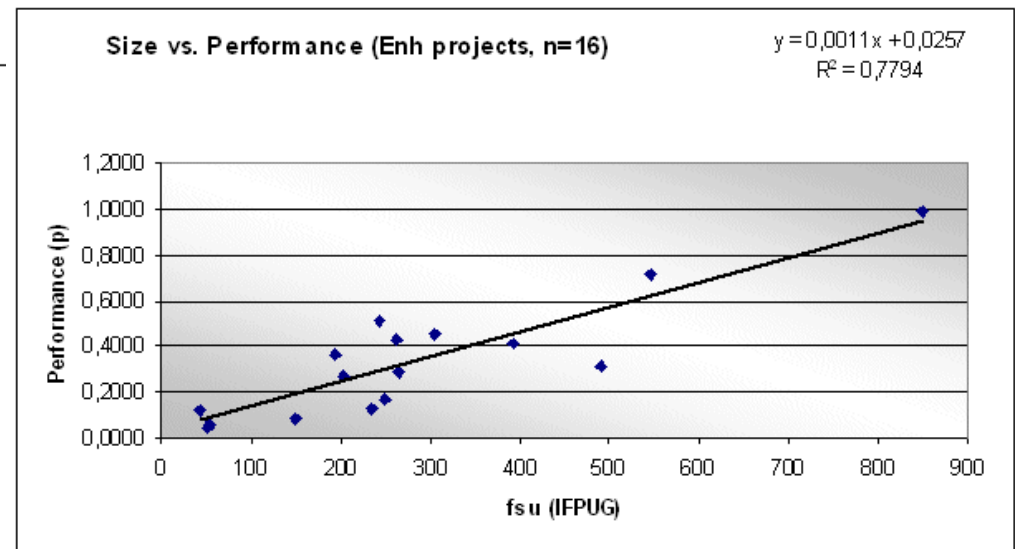
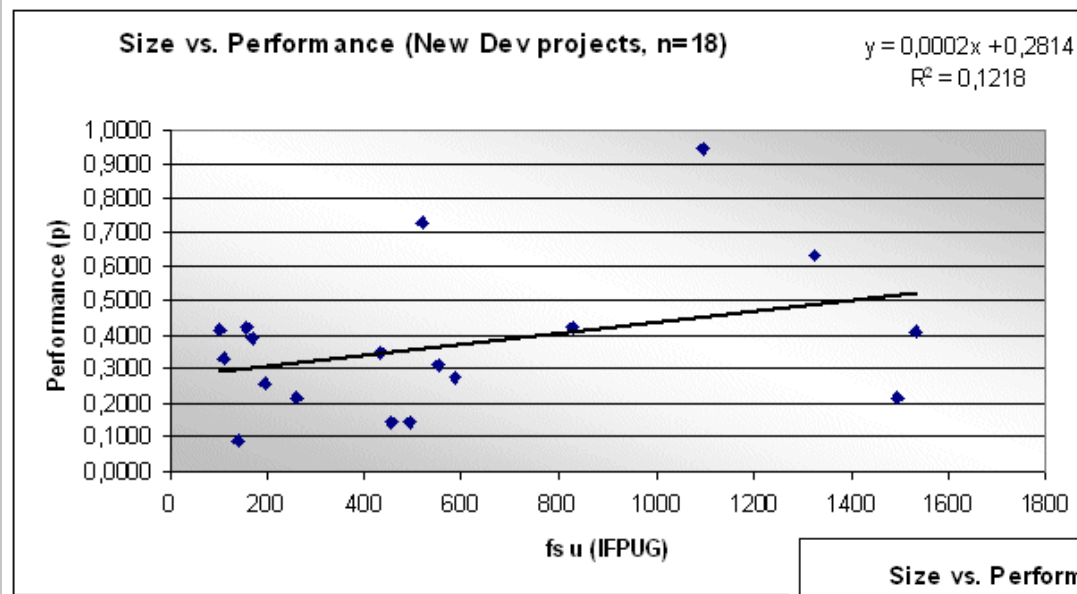
Analyzing Results

Presentation of results: FunctSize Vs Effort



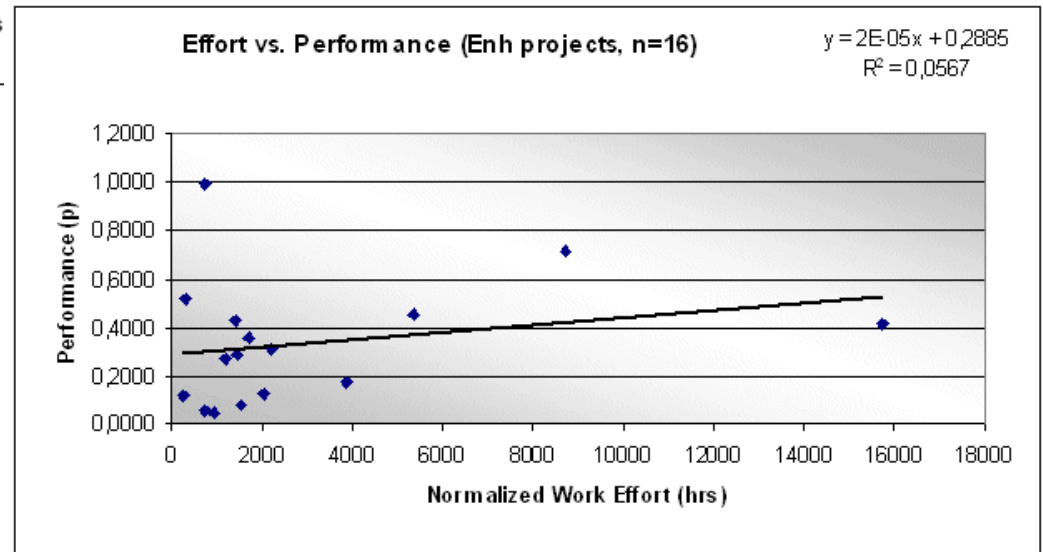
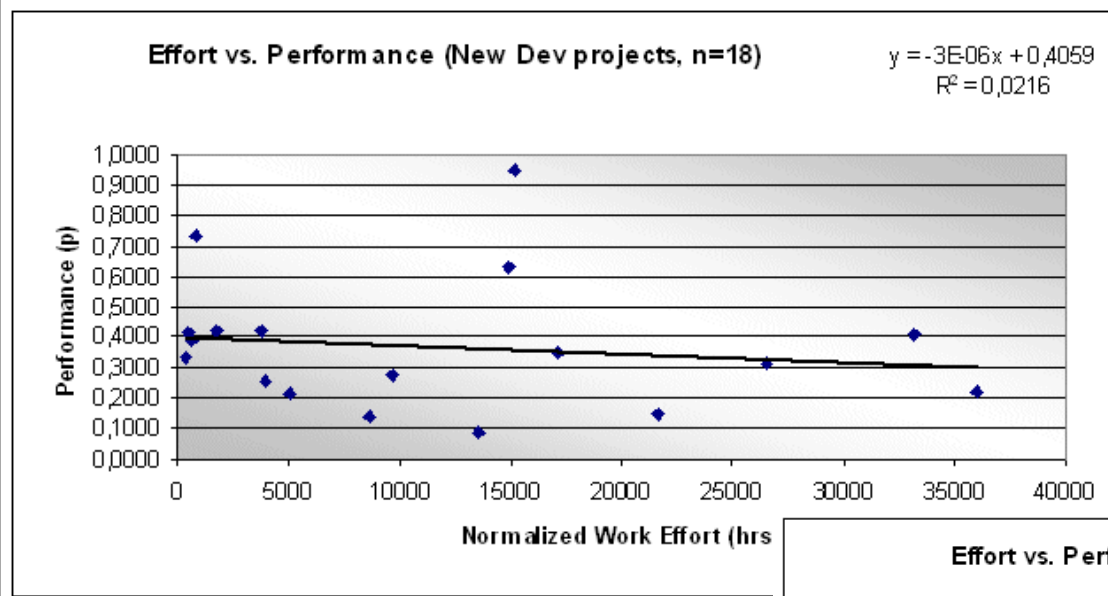
Analyzing Results

Presentation of results: FunctSize Vs Performance



Analyzing Results

Presentation of results: Effort Vs Performance



Analyzing Results

Presentation of results: Equations/R² by type



- The best relationships seems to be:
 - ✓ FunctSize vs Effort (NewDev)
 - ✓ FunctSize vs Performance (Enh)

Relationship	FSMM	Dev. type	N	R ²	Rank	Equation
<i>Size vs. Effort</i>	IFPUG	New Dev	18	0.52	☺	y=17.133x+1907.8
	IFPUG	Enhance.	16	0.09	☹	y=5.8857x+1345.3
<i>Size vs. Performance</i>	IFPUG	New Dev	18	0.12	☹	y=0.0002x+0.2814
	IFPUG	Enhance.	16	0.78	☺	y=0.011x+0.0257
<i>Effort vs. Performance</i>	IFPUG	New Dev	18	0.02	☹	y=-3E-06x+0.4059
	IFPUG	Enhance.	16	0.06	☹	y=2E-05x+0.2885



Analyzing Results

Presentation of results: Ranking (NewDev)



#	ISBSG Id.	Fsu	NWE	Prod (hrs)	Prod (m/d)	P	Ranking – Productivity	Ranking – Performance
P001	11465	112	406	0.28	2.21	0.40	2	10
P002	11718	194	3934	0.05	0.39	0.31	12	13
P003	12080	157	1754	0.09	0.72	0.46	6	4
P004	12310	456	21684	0.02	0.17	0.21	16	16
P005	14257	100	440	0.23	1.82	0.46	4	5
P006	14863	494	8706	0.06	0.45	0.20	10	17
P007	19665	1095	15165	0.07	0.58	0.95	8	1
P008	20677	828	3742	0.22	1.77	0.46	5	6
P009	21908	140	13528	0.01	0.08	0.18	18	18
P010	22338	1326	14938	0.09	0.71	0.68	7	3
P011	24768	258	5048	0.05	0.41	0.27	11	15
P012	26294	433	17120	0.03	0.20	0.41	15	9
P013	28180	1493	36046	0.04	0.33	0.27	14	14
P014	28589	585	9653	0.06	0.48	0.33	9	12
P015	28819	1533	33140	0.05	0.37	0.45	13	7
P016	30455	550	26520	0.02	0.17	0.36	17	11
P017	31784	517	784	0.66	5.28	0.76	1	2
P018	32133	171	640	0.27	2.14	0.44	3	8



Analyzing Results

Presentation of results: Ranking (Enh)



#	ISBSG Id.	Fsu	NWE	Prod (hrs)	Prod (m/d)	P	Ranking – Productivity	Ranking – Performance
P001	10481	304	5377	0.06	0.45	0.45	14	4
P002	10783	202	1171	0.17	1.38	0.27	7	10
P003	11474	195	1717	0.11	0.91	0.36	9	7
P004	12583	265	1464	0.18	1.45	0.29	5	9
P005	15274	53	703	0.08	0.60	0.06	11	15
P006	18873	42	242	0.17	1.39	0.12	6	13
P007	19114	392	15729	0.02	0.20	0.41	16	6
P008	21580	545	8740	0.06	0.50	0.71	13	2
P009	21686	235	2024	0.12	0.93	0.13	8	12
P010	22168	848	730	1.16	9.29	0.99	1	1
P011	24093	149	1532	0.10	0.78	0.08	10	14
P012	25714	249	3877	0.06	0.51	0.17	12	11
P013	26189	51	949	0.05	0.43	0.04	15	16
P014	29491	262	1417	0.18	1.48	0.43	4	5
P015	29644	490	2192	0.22	1.79	0.31	3	8
P016	31360	244	300	0.81	6.51	0.51	2	3



Analyzing Results

Relative Productivity & Performance Rankings



Dataset		A<B	B<A	A=B
A (new dev)	Abs	9	6	3
	%	50%	33%	17%
B (enhancements)	Abs	10	5	1
	%	63%	31%	6%

Legend: **A**=rank(productivity); **B**=rank(performance)

- Being performance a high-maturity concept (that includes productivity), the design of an improvement path requires a multi-perspective analysis (E/S/T) with a different balancing of weights across the selected measures
- Periodical analysis could help in rethinking the way projects are monitored and controlled
- Not necessarily a high-productivity project is successful, because it must achieve several business goals at the same time, not only that one



Agenda



- **Introduction**
- **Multi-Dimensional Performance Models**
 - ✓ QEST-LIME models: description
 - ✓ QEST-LIME models & Performance Targets
- **The ISBSG r10 Data Repository**
 - ✓ Available Fields
 - ✓ Selection of data samples in the ISBSG repository r10
 - ✓ Selection of the common set of measures
 - ✓ Overall Performance Levels
- **Analyzing results from performance benchmarking models**
 - ✓ Data flow for analysis of results
 - ✓ Presentation of results from ISBSG r10 data
- **Conclusions**



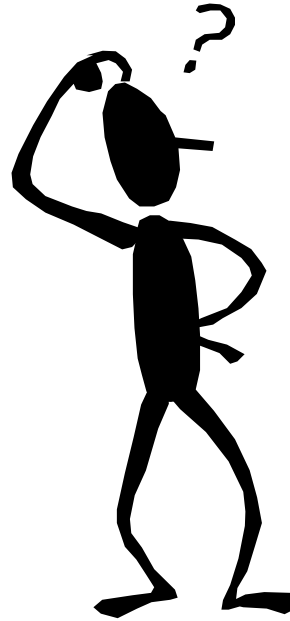
Conclusions



- Traditional cost estimation models in Software Engineering are mainly based on the concept of Productivity, while the usage of Performance could provide benefits, being a more mature and comprehensive concept.
- Performance can be calculated in several ways; here the QEST model was used jointly with ISBSG r10 data using two data samples (IFPUG, NewDev – Enh projects)
- 2 data samples from IFPUG projects were used (NewDev – Enh) for calculating p values.
- The ranking between the productivity & performance values revealed that performance leads to a broader coverage of business goals, with more than a single perspective at a time.
- The analysis of multi-dimensional performance values can bring several benefits
- New analysis and investigations will be performed using ISBSG data on:
 - ❖ the impact of relationships of various variables on the performance results themselves
 - ❖ The same analysis by size ranges



Q & A



Grazie! Thank you!



Luigi.buglione@eng.it, alain.abran@etsmtl.ca

