Size: The Forgotten Measure

SEPG North America

March 15, 2012

Albuquerque, N.M.

Joe Schofield

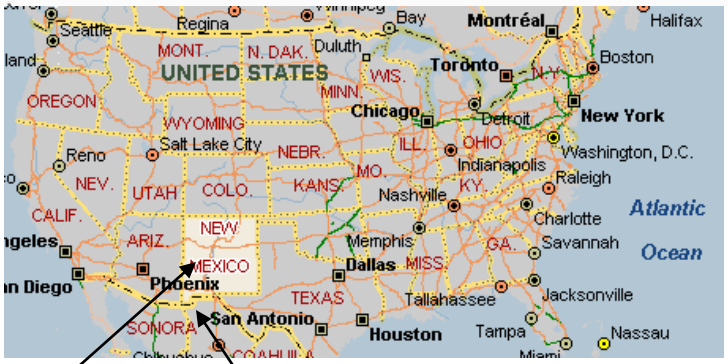
CFPS, CSMS, CSQA, LSS BB

SEI-Certified CMMI Instructor

President, International Function Point Users Group

GEOGRAPHICAL CONTEXT DIAGRAM

State joke



New Mexico Mexico

State flag & statue

State animal

State billboard

State motto

State mineral

ABSTRACT / TOPICAL CONTENT

Provide a Session Abstract for Publication between 50 and 100 words: Scorecards and dashboards that focus on cost and schedule, while minimizing product size may hinder clear project reporting. Measures that offer little opportunity for benchmarking and cross-project comparison are limiting the value organizations obtain from their measurement systems. Use case points, story points, function points—what's the point, if they aren't providing decision-makers with objective data needed for improvement. Specific practices within the CMMI-DEV mitigate the impact of "soft" estimates on project outcomes. A practical approach to life cycle management for software sizing is rendered along with 5 C's of useful measurements: complete, correct, consistent, current, and connected.

Length: 40 Minutes

Which one conference track best suits your presentation?: Getting Started

Learner Outcome 1: Bringing schedule and cost into balance with size is necessary for estimating and measurement

Learner Outcome 2: Story points, use case points, function points – what's the point? Understanding the strengths and limitations of these product measures

Learner Outcome 3: Why your line of code measures are less reliable than you think – a statistical analysis of PSP(SM) course programs

Setting the context for our content . . .

SIZE IS FOUND 55 TIMES IN THE "DEV MODEL"

Within the Project Planning process area . . .

SP 1.2 Establish Estimates of Work Product and Task Attributes

Establish and maintain estimates of work product and task attributes.

Size is the primary input to many models used to estimate effort, cost, and schedule. Models can also be based on other attributes such as service level, connectivity, complexity, availability, and structure.

"*Size*" is a base measure in the CMMI®.

Size is also found (in context) 13 times in the Services model; 20 more in the Acquisition model.

EXAMPLES OF ESTIMATING ATTRIBUTES INCLUDE:

- Number and complexity of requirements
- **Number and complexity of interfaces**
- Volume of data
- **Number of functions**
- **Function points**
- Source lines of code
- Number of classes and objects
- **Number of database tables**
- **Number of fields in data tables**
- Architecture elements
- Experience of project participants
- Amount of code to be reused versus created
- Team velocity and complexity
- **Number of pages**
- **Number of inputs and outputs**
- Number of technical risk items
- **Number of database tables**
- **Number of fields in data tables**
- Architecture elements

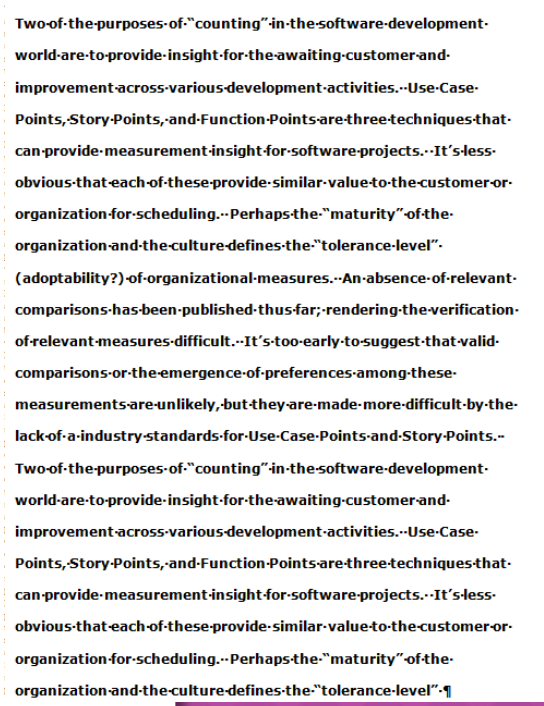
- Experience of project participants
- Amount of code to be reused versus created
- Number of logic gates for integrated circuits
- Number of parts (e.g., printed circuit boards, components, mechanical parts)
- Physical constraints (e.g., weight, volume)
- Geographic dispersal of project members
- Proximity of customers, end users, and suppliers
- How agreeable or difficult the customer is
- Quality and "cleanliness" of the existing code base

- **Ten of these attributes are also attributes of Function Point Analysis**
- **"Agreeable or difficult customer"**
- **Two sets of duplicate entries**

PAGE *SIZE* VARIATIONS

	Margins	Font	Font Size	Spacing	Bolding	Char. Per page	% Content Loss
Initial settings	.3 top & bottom; .4 sides	Times New Roman	10	Single	none	7584	0
	1”					5450	28
		Verdana				5686	25
			12			5177	32
				Double		4353	43
					ON	7185	5
Initial settings	1”	Verdana	12	Double	ON	1403	83

- Read “% Content Loss” (last column) as *variation*!
- Cumulative difference of one page to almost six
- Consider still larger font, font size, spacing, charts, diagrams, pictures, etc.
- **Impact on PMC SP1.1** – Monitor actual values of project planning parameters against the project plan.


Two·of·the·purposes·of·"counting"·in·the·software·development·world·are·to·provide·insight·for·the·awaiting·customer·and·improvement·across·various·development·activities.··Use·Case·Points,·Story·Points,·and·Function·Points·are·three·techniques·that·can·provide·measurement·insight·for·software·projects.··It's·less·obvious·that·each·of·these·provide·similar·value·to·the·customer·or·organization·for·scheduling.··Perhaps·the·"maturity"·of·the·organization·and·the·culture·defines·the·"tolerance·level"·(adoptability?)·of·organizational·measures.··An·absence·of·relevant·comparisons·has·been·published·thus·far;·rendering·the·verification·of·relevant·measures·difficult.··It's·too·early·to·suggest·that·valid·comparisons·or·the·emergence·of·preferences·among·these·measurements·are·unlikely,·but·they·are·made·more·difficult·by·the·lack·of·a·industry·standards·for·Use·Case·Points·and·Story·Points.·

Two·of·the·purposes·of·"counting"·in·the·software·development·world·are·to·provide·insight·for·the·awaiting·customer·and·improvement·across·various·development·activities.··Use·Case·Points,·Story·Points,·and·Function·Points·are·three·techniques·that·can·provide·measurement·insight·for·software·projects.··It's·less·obvious·that·each·of·these·provide·similar·value·to·the·customer·or·organization·for·scheduling.··Perhaps·the·"maturity"·of·the·organization·and·the·culture·defines·the·"tolerance·level"·¶

LOC SIZE LIMITATIONS

All of these for the same set of requirements, same LOC counting, similar education levels, student-selected language based on their familiarity, same instructor / product owner from Personal Software Process classes

Course	Attendee	*P1	P2	P3	P4	P5	P6	P7	P8	P9
1	Attendee 1	89	34	67	40	102	235	23	38	168
1	Attendee 3	82	23	33	48	61	34	33	27	52
1	Attendee 4	177	119	67	85	136	276	165	112	233
1	Attendee 5	76	48	305	244	61	121	66	77	127
1	Attendee 7	46	33	17	37	60	95	129	46	186
3	Attendee 5	22	40	100	58	68	131	58	58	102
3	Attendee 6	46	20	30	42	73	82	51	72	82
2	Attendee 7	95	155	147	94	54	191	174	102	218
	Min	22	20	17	37	54	34	23	27	52
	Max	177	155	305	244	136	276	174	112	233
	% Variation	805	775	1794	659	252	812	757	415	448
	Mean	79	59	96	81	77	146	87	67	146
	Std. Dev.	47	50	95	69	28	82	60	30	66

2	Attendee 5	193	137	48	102	107				
1	Attendee 2	77	163	168	123	134				
3	Attendee 1	73	37	36	95	101				
3	Attendee 2	74	97	143	153	279	146	176	80	305
3	Attendee 4	114	71	108	80	219	189	142	95	163
	Min	73	37	36	80	101	138	51	66	134
	Max	193	163	168	153	279	207	238	178	305
	% Variation	64	441	33	100	46	73	36	69	82
	Mean	106	101	101	111	168	169	145	97	184
	Std. Dev.	51	50	58	28	78	29	69	47	71

	P3	P4	P5	P6	P7	P8	P9
	128	103	227	186	306	155	61	283
	143	114	13	110	63	113	84	85
	106	36	34	53	51	54	61	125
	38	51	61	134	99	43	58	126
	311	271	289	142	122	190	383	219
	179	56	150	202	185	155	118	144
	30	15	30	61	116	69	43	147
	96	102	197	64	158	85	87	126
Attendee 9	64	63	36	169	56	23	99	73	83
Attendee 10	101	116	108	49	66	103	71	51	73
Min	35	30	15	13	53	23	43	43	73
Max	221	311	271	289	202	306	190	383	283
% Variation	631	1037	1807	2223	381	1330	442	891	388
Mean	113	121	89	122	107	123	103	102	141
Std. Dev.	56	81	73	97	56	81	49	101	65

Java - variance % range 252 - 1794

C++ - variance % range ~381 - 2223

Visual Basic - variance % range 33 - 441

Source: The Statistically Unreliable Nature of Lines of Code; CrossTalk; April, 2005

EXPERIENCE OF PROJECT PARTICIPANTS NOT RELEVANT

Not a "weakest link" induced dilemma

Attendee 3 had 4 of the "largest programs"

Attendee 3 also had the shortest 7th program! (see P7)

Seven attendees had a shortest program!

	*P1	P2	P3	P4	P5	P6	P7	P8	P9
Attendee 1	33	40	30	108	65	176	79	107	284
Attendee 2	51	52	24	72	109	166	87	145	270
Attendee 3	76	56	30	115	175	158	27	104	128
Attendee 4	60	32	31	108	94	155	72	94	235
Attendee 5	22	51	25	50	75	105	47	21	102
Attendee 6	65	27	80	45	95	141	91	60	209
Attendee 7	22	51	25	50	75	105	47	21	102
Attendee 8	65	27	80	45	95	141	91	60	209
Min	22	27	24	45	65	105	27	21	102
Max	76	56	80	115	175	176	91	145	284
% *Variation*	345	207	333	256	269	168	337	690	278
Mean	49	45	41	74	98	143	68	77	192
Std. Dev.	21	12	24	31	34	26	24	44	73

Four other attendees had the "largest program" in other assignments

Perhaps most surprising, five attendees had both a longest and shortest program in their set of assignments.

Attendee 7 had four of the shortest programs and none of the largest. ("best"?)

Attendee 4 had neither a longest or a shortest program. ("most mediocre"?)

SIZE – WHEN DOES IT MATTER?

- Largest one day swings on Wall Street
- World's largest mansion
- World's largest cruise ship
- World's largest city
- World's largest lottery payout
- World's largest aircraft
- World's largest baby birth weight
- ~~World's largest software~~

S, M, L, XL, XXL

Large, Jumbo, Grand

Supersized

SIZE-RELATED PROJECT AILMENTS

Standish Chaos Report
Challenged projects suffer from:

1. Lack of User Input
2. Incomplete Requirements and Specifications
3. Changing Requirements and Specifications
4. Lack of Executive Support
5. Technology Incompetence (DTRA, XML?)
6. Lack of Resources
7. Unrealistic Expectations
8. Unclear Objectives
9. Unrealistic Time Frames
10. New Technology

Impaired (cancelled) projects suffer from:

1. Incomplete Requirements
2. Lack of User Involvement
3. Lack of Resources
4. Unrealistic Expectations
5. Lack of Executive Support
6. Changing Requirements and Specifications
7. Lack of Planning
8. Didn't Need it Any Longer
9. Lack of IT Management
10. Technology Illiteracy

IEEE Spectrum, Robert N. Charette, September, 2005
Why Software Fails

1. Unrealistic or unarticulated project goals
2. Inaccurate estimates of needed resources
3. Badly defined system requirements
4. Poor reporting of the project's status
5. Unmanaged risk
6. Poor communication among customer, developers, and users
7. Use of immature technology
8. Inability to handle the project's complexity
9. Sloppy development practices
10. Poor project management

Size measures tell us:

Three of Wall Street's largest one day gains occurred in a 32-day stretch from October 13, 2008 to November 13, 2008 (including the 28th)

Five of Wall Street's 10 largest one day losses occurred between September 29th and December 1st of 2008.

Source: Wall Street Journal online; photo from wikipedia

SIZE INFLUENCE IN M & A

Measurement Information Categories	*Example Base Measures*	*Example Derived Measures*
Schedule and progress	Estimated and actual start and end dates by task	Milestone performance Percentage of project on time Schedule estimation accuracy
Size and effort	Estimated and actual effort and size	Productivity
Effort and cost	Estimated and actual cost	Cost performance Cost variance
Size and stability	Requirements count	Requirements volatility Size estimation accuracy
	Function point count	Estimated vs. actual function points
	Lines of code count	Amount of new, modified, and reused code
Quality	Number of defects inserted and detected by lifecycle phase Product size	Defect containment by lifecycle phase Defect density
Cost	Number of defects inserted and detected by lifecycle phase Effort hours to correct defects Labor rates	Rework costs
Information Assurance	Number of system vulnerabilities identified and number of system vulnerabilities mitigated	Percentage of system vulnerabilities mitigated

From the CMMI-DEV model: Table MA.1: Example Measurement Relationships (last three columns only)

Size measures tell us:

The world's largest mansion has 400,000 square feet of living space.

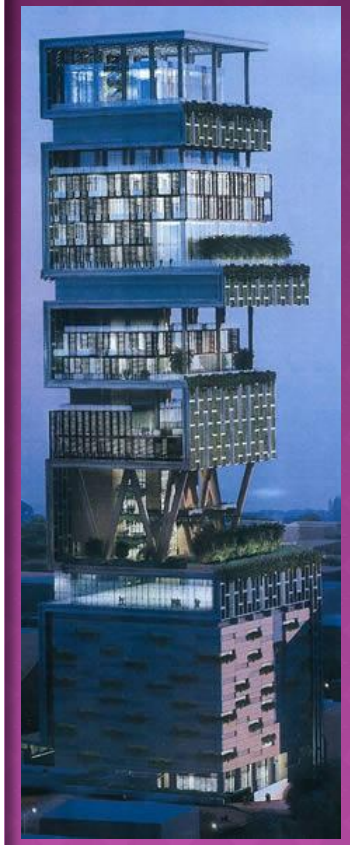


Photo from Wikipedia

USE CASE POINTS, STORY POINTS, FUNCTION POINTS — *WHAT'S THE POINT*?

Characteristic	Function Points	Use Case Points	Story Points
Useful at the project level for estimating or planning	With historical FP data	With historical UCP data	With historical SP data
ISO / Standards based	ISO 20926	no	no
Captures customer view	Expected	Expected	Definitely
Useful for benchmarking outside the company	Could be	Could be	Less so
Easy to calculate	Less so	More so	Yes
Easy to validate for repeatability / consistency	More so	More so	Less so
Objectivity	More so	More so	Less so (team / team member variability)
Technologically independent	Yes	Yes	Maybe
Functional measurement to customer	Yes	Yes	Not exclusively (may include refactoring, design, and other work)

Size measures tell us:

The world's largest cruise ship - Allure of the Seas - is 225,282 Gross Tons.

Photo from Wikipedia

FP ANALYSIS SUPPORTS OBJECTIVE SIZING AS . . .

- As an ISO standard (ISO/IEC 20926 SOFTWARE ENGINEERING) Function Point Analysis (FPA) provides a basis for repeatable and consistent sizing
- Supported by IFPUG and its membership community, FPA remains viable as new technologies and approaches to software development evolve (case studies, books, conferences, workshops, certifications, and, the "standard")
- Functional sizing is not influenced by programming language, in-house or COTS development
- Functional sizing is not impacted by development approach: outsourcing, in-sourcing, iterative, incremental, scrum, or agility
- Functional sizing can be approximated at the first sighting of customer requirements, estimated with a design, and counted upon delivery
- FPA can be used to track requirements volatility over the life of a project (FPs added, changed, deleted) to size *requirements creep*

Size measures tell us:

The world's largest city, Tokyo, has some 13 million people, 35 million directly around it; Tokyo also is home to the most Global 500 companies - 47.

Photo from Wikipedia

SOFTWARE *NON*-FUNCTIONAL ASSESSMENT PROCESS (SNAP)

Targeted at:

- building better benchmarks
- improving software estimation
- quantifying technical strategies
- communicating non-functional issues between stakeholders

A methodology to:

- measure the non-functional size of a software product supporting quality and productivity analysis,
- estimate cost and resources required for software development and maintenance,
- determine the non-functional size of a purchased application package by assessing all the portions and categories included in the package,
- help consumers to determine the benefit of an application package driven by requirements.

Released in 2011, SNAP complements IFPUG's Function Point counting practices:

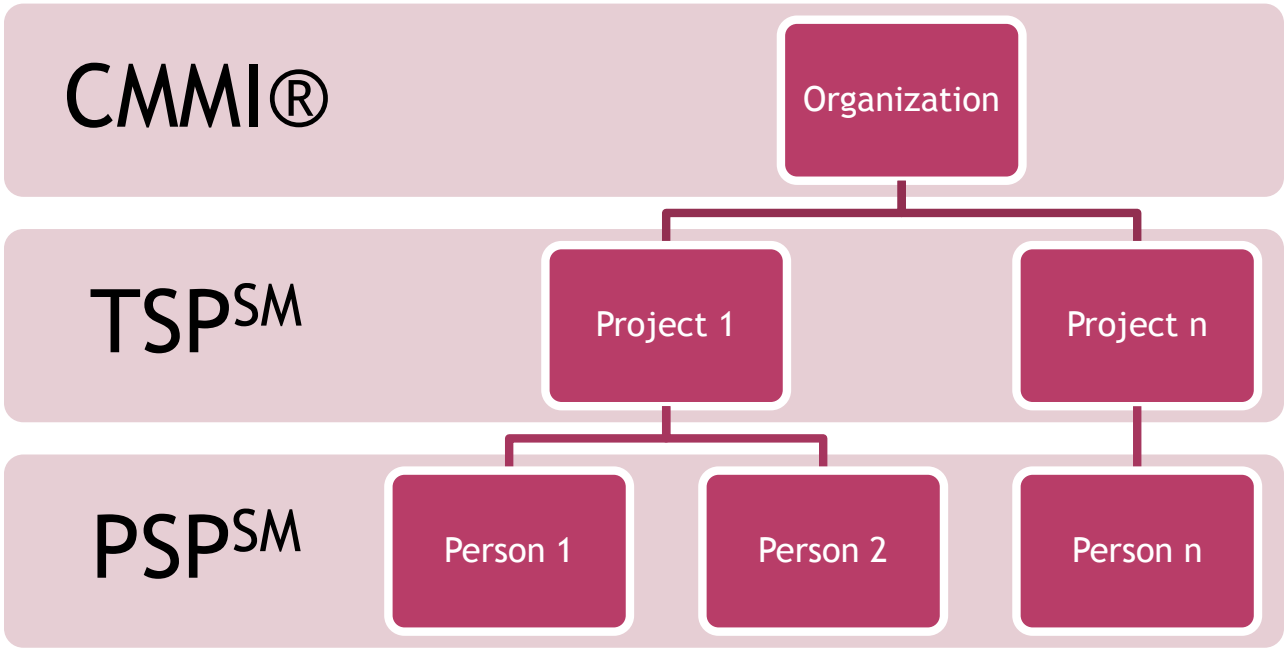
Data Operation

Interface Design

Technical Environment

Architecture

BUILDING MEASUREMENT CAPABILITY



Size measures tell us:

The world's largest lottery for a single ticket winner was a $365M Powerball ticket; Nebraska, 2006; split by eight folks.

SOURCES OF MEASUREMENT RESISTANCE

Early reticence to measurement for organizations and teams is often characterized by comments like:

- Measurement is hard; (it's even harder if postponed or ignored).
- We make only limited decisions based on measurement data, thus the need to collect it seems specious.
- How do *my* numbers compare to others? We may need to adjust them.
- Our numbers could be used against us.
- Our numbers could be better; for now, they're good ballpark estimates.

Size measures tell us:

The world's largest aircraft, the An-225 weighed 1,410,000 lbs. The vehicle has six engines and 32 wheels.

Photos from wikipedia

5C's OF SIZING MEASURES

Complete
- Captures all of the product delivered to the customer
- Doesn't capture "hows" but rather the whats

Correct
- Captures the measure
- Doesn't allow for manipulation of base measures

Consistent
- Captures clearly defined measures similarly for all
- Doesn't facilitate "local" massaging before entry

Current
- (Repository) differentiates between recent and ancient values
- Doesn't incorporate irrelevant measures for predictive models

Connected
- Measures are linked to organizational objectives; project measures to the organization's measures
- Doesn't allow measurement providers to create (and interpret) their own measures in place of organizational measures

Size measures tell us:

The world's largest biggest baby at birth according to Guinness was 23.12 lbs.; born in Canada in 1879.

Photo courtesy of my (1st) granddaughter

MEASUREMENT LIFECYCLE

Before planning begins, establish organizational capability, intent, and decision needs

Educate

At first sign of requirements or scope determination

Approximate

When requirements are firm and team is assigned; when requirements change beyond volatility threshold

Estimate

Throughout the life of the project - this is monitoring and controlling; actual vs. estimated over the *phase* or *iteration* of the project

Actuate

Upon (release) completion; actual vs. estimated over the *life* of the project

Calibrate

Glossary

Measurement Repository

Estimation and performance models

Related Process Areas:

- M & A
- OPD
- PP
- PMC
- OPP
- OPM
- QPM

An example of "approximate" based on "educate"

WAIT, IT COULD GET WORSE

When people make interventions . . . in most of the cases the intervention intended to improve the process actually causes outcomes to get worse before they get better, if they get better at all.

Underestimation in the "When It Gets Worse Before it Gets Better" Phenomenon in Process Improvement; Advanced Concurrent Engineering, 2011, Part 1, 3-10, DOI: 10.1007/978-0-85729-799-0_1
Ricardo Valerdi and Braulio Fernandes

The cognitive approach has identified a number of erroneous beliefs held by gamblers, which cause them to over-estimate their chances of winning.

Decision-making during gambling: an integration of cognitive and psychobiological approaches; Luke Clark

Research has shown that the confidence individuals express in their judgments generally exceeds the accuracy of those judgments on difficult tasks.

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 48, 100-130 (1991); Influences on the Appropriateness of Confidence in Judgment: Practice, Effort, Information, and Decision-Making; PAUL W. PAESE, University of Missouri-St. Louis; JANET A. SNIEZEK

Software and measurement professionals have demonstrated that estimation results—which weren't very close originally—often grow worse when subjected to competition and pressure to succeed.

Measurements, Biases, Judgments: Understanding Variations for Reliable Estimates; Keynote Address - Sao Paulo, Brazil; Brazilian Software Measurement & Analysis Conference; November 11, 2011; Joe Schofield

"I used to be better when I wasn't very good at all (or I didn't know how bad it was)."

CLOSING THOUGHTS

A Google search of world's largest *software* returns a list of world's largest software *companies* – 346M hits as of 2/2012.

Given our limited understanding of size, are schedule and budget variances taking more than their share of criticism for project overruns?

It's preferable to "miss the mark" than not to have a model that predicts where the mark should be.

Function points are mentioned twice in the CMMI-DEV®. Sometimes shunned for being complex, they offer an objective, repeatable, and benchmark-worthy value.

Size may have multi-constellational application (SVC, ACQ).

REMINDER: a (wrong) "correction" could worsen performance

AND THANK YOU!

Hopefully you have come to realize that *size* - it really does matter!

ADDITIONAL READINGS

FPA – Function Point Analysis; http:www.ifpug.org, http://en.wikipedia.org/wiki/Function_point

GQM – Goal Question Metric; *en.wikipedia.org/wiki/GQM*

ISO / IEC 20926:2009 - http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35582

PSM – Practical Software and Systems Measurement; http://www.psmsc.com/

SNAP - http://ifpug.org/about/ITperformance.htm

The Statistically Unreliable Nature of Lines of Code; CrossTalk, April, 2005

A Discipline for Software Engineering; Watts Humphrey; Addison-Wesley; 1995

Counting Lines of Code: Virtually Worthless for Estimating and Software Sizing, IT Metrics and Productivity Journal; December, 2009

Is There a Weakest Link After All?, IT Metrics and Productivity Journal; December, 2009

Certified Function Point Specialist Examination Guide; Garmus, et. al.; 2010; ISBN 978-1-4200-7637-0

Measures need to reflect intent, not merely "compliance."