

NASA/IVV Presentation

Experimental Studies on the Benefits of Independent Verification and Validation

Ioana Rus and Marvin V. Zelkowitz
Fraunhofer Center for Experimental Software Engineering
College Park, Maryland 20742

NASA Subcontract No. 93-393B-FUSA
October, 1999

1

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Introduction

- As a result of the 1986 Challenger accident, NASA instructed to develop an IV&V program
- In 1999, a center initiative was instituted to look at the economic benefit of this IV&V process
- Participants from the Fraunhofer Center for Experimental Software Engineering, Maryland, West Virginia University and Portland State University
- **Project goal:** Determine the economic payoff to NASA for instituting an IV&V process
- **Initial Fraunhofer Center - Maryland goal: Assess current practice in IV&V evaluation methodologies**

2

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Independent Verification and Validation

- An organization *independent* from the developers study the artifacts of software production.
- This requires:
 - **Technical independence.** Members of the IV&V team may not be personnel involved in the development of the software.
 - **Managerial independence.** The responsibility for IV&V belongs to an organization outside the contractor and program organizations that develop the software.
 - **Financial independence.** Control of the IV&V budget is retained in an organization outside the contractor and program organization that develop the software.
- IV&V often tests the code after the development is completed



IV&V activities

Algorithm analysis	Analytic modeling	Back-to-back testing
Boundary value analysis	Code reading	Control flow analysis
Coverage analysis	Critical time analysis	Database analysis
Dataflow analysis	Desk checking	Error seeding
Event tree analysis	Functional testing	Inspections
Interface analysis	Mutation testing	Performance testing
Proof of correctness	Prototyping	Regression testing
Requirements analysis	Requirements traceability	Reviews
Sensitivity analysis	Simulation	Size analysis
Slicing	Software failure mode anal.	Stress testing
Structural testing	Symbolic execution	Test certification
Walkthroughs		

–From [Wallace, 96]



Initial concept to IV&V initiative

- Study literature to determine effectiveness of various IV&V techniques
- Study NASA projects to learn which software projects used which IV&V techniques
- Develop a return on investment (ROI) model
- Use the improvement found for each technique as parameters for the ROI model
- Use ROI model to predict costs on future NASA projects

- As an example, the following uses Net Present Value (NPV ...)



Example: Net Present Value

- Net Present Value has following equation:

$$NPV = \sum_{t=1}^n \frac{F_t}{(1+k)^t} - I$$

where:

- F_t is the net cash flow
- k is the marginal cost of capital (i.e., the interest rate for capital)
- n is the project's expected life
- I is the initial cost of the project
- $NPV > 0$ is what is desired

- We need to evaluate the parameters F_t and I .
 - I is a measure of the cost of performing IV&V now
 - F_t is the future benefit from avoiding an activity later (e.g., less testing)

However, the value of these parameters depend upon the set of techniques employed, and the benefits depend upon previous studies into these technologies.



Changes to IV&V concept

Two things intervened to change the original focus:

- NASA uses a concept for IV&V different from the one classically used in the software engineering literature.
 - No value judgment on this.
 - Only that the process is different
- The research literature in IV&V improvement is not as rich as we would like or need to do a complete analysis.
 - Much about this later



NASA IV&V Concept (From 1996 RPF)

- Phase independent analyses
 - Maintain a high-risk functions list
 - Perform traceability analysis on high-risk components
 - Maintain an issues tracking to monitor progress
- Phase dependent analyses
 - Review documentation in each phase as needed
 - Review process documentation
 - Perform a software requirements analysis
 - Perform an interface requirements analysis
 - Perform a software design analysis.
 - Perform a software code analysis.
 - Validate the verification and validation process used by the developer.
 - Ensure that the developed software is maintainable.
 - Perform an audit at each major program milestone.



IV&V at NASA

- Projects are sustaining engineering
 - Not new projects
 - New releases (operational increments (OI) of shuttle program)
 - Emphasis is mission safety
- Funding insufficient to validate every component
 - Choose high risk components
 - Audit process to determine that component has been validated; not separately tested
 - IV&V becomes institutional memory between OIs
 - IV&V works with developer during requirements phases. Technical independence of IV&V team somewhat compromised
- **IV&V at NASA is a risk reduction process**
- **Does not appear to be a comparable IV&V activity outside of NASA**



IV&V Literature Search

- Current report document extensive, but not yet complete
- 146 references found
- Major sources were:
 - IEEE Transactions on Software Engineering
 - IEEE Computer
 - IEEE Software
 - International Conferences on Software Engineering
 - Other sources as appropriate
- Emphasized last 20 years, but older references also included

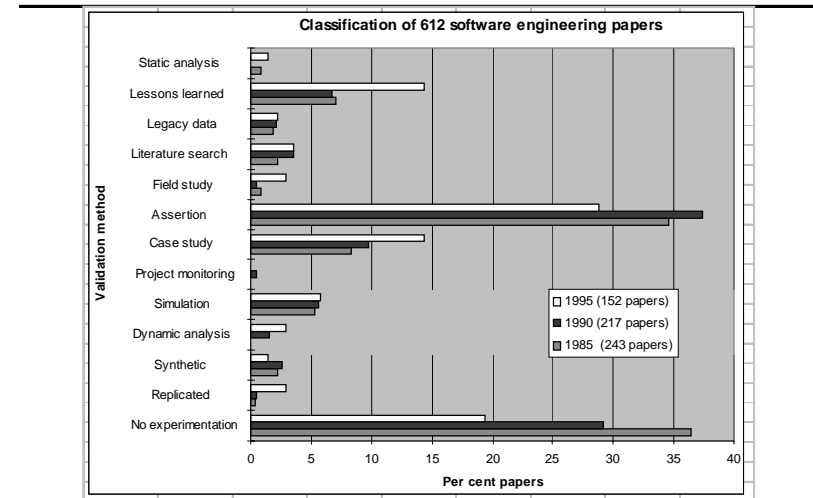


Overall Classification of Papers

- Domain: Software engineering papers that present an economic (e.g., quantitative) evaluation of the effectiveness of that technique
 - Ignored papers that simply defined a technique and explained its virtues
 - Ignored techniques that we did not think were relevant to IV&V
- Relevance of software engineering papers
 - Vast majority of papers had no evaluation at all
 - A minority had data on the use of that technique, but not on its relation to any other technique
 - A still smaller minority had data comparing that technique to others **(This is the category we were searching for.)**
- Lack of papers with validation consistent with studies by Tichy [95] and Zelkowitz [97].
 - **50% of published papers contain no validation.**

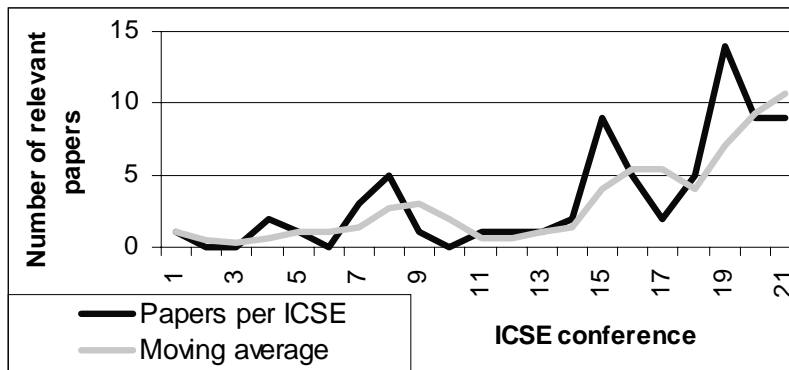


Zelkowitz-Wallace survey (May '98 Computer)



Relevance of ICSE papers

Early ICSEs had few relevant papers. Situation similar in other journals.



But situation since 1992 is improving.



Environments

Environment	Count
Industrial	83
University	24
Unknown	41

- Most papers contain data relevant to industrial world
- Most of the unknown papers are probably university studies



Techniques Measured

Audits	Model checking
Capture/recapture	Models (e.g., cost, defect, schedule, complexity)
Cleanroom	Process (e.g., CMM, PSP, ISO 9000, Experience Factory, QIP)
COCOMO	Programming environments
Data mining	Prototyping
Debugging	Reading
Defect tracking	Requirements testing
Design models	Reuse
Fault trees	SCR
Inspections and reviews	Simulation
IV&V general	Slicing
Management issues (e.g., personnel, environments)	Static analysis
Measurement	Testing (e.g., mutation, structural, functional, reading, integration)
Middleware	Value chaining



Frequency of papers

Inspections and reviews	21	Static analysis	3
Models (e.g., cost, defect, schedule, complexity)	20	Cleanroom	2
Testing (e.g., mutation, structural, functional, reading, integration)	17	Debugging	2
IV&V general	16	Fault trees	2
Defect tracking	10	Model checking	2
Management issues (e.g., personnel, environments)	10	Reading	2
Process (e.g., CMM, PSP, ISO 9000, Experience Factory, QIP)	9	SCR	2
Requirements testing	8	Value chaining	2
Measurement	5	Audits	1
Data mining	4	Capture/recapture	1
Design models	4	Middleware	1
COCOMO	3	Prototyping	1
Programming environments	3	Simulation	1
Reuse	3	Slicing	1



Defect detection

Inspections and reviews	[PORTER1994][JALOTE1998][FAGAN1976][CONRAD1999] [PORTER1997A][BUSH1999][VOTTA1993][SEAMAN1997]
Models (e.g., cost, defect, schedule, complex)	[RAMAMOORTHY1984][BRIAND1993][BRIAND1993]
Testing (e.g., mutation, structural, functional, reading, integration)	[KOREL1996][FRANK1997][DALAL1999] [ROTHERMEL1998][GRAVES1998][WONG1995][PIWOWARSKI1993]
IV&V general	[CARD1987][ARTHUR1999]
Defect tracking	[SCHNEIDEWIND1998][WEERAHANDI1994] [CHILLAREGE1991][MIYAZAKI1987][SHOOMAN1975]
Mgmt issues (e.g., personnel, environments)	[DEMARCO1985]
Process (CMM, PSP, ISO 9000, Ex. Fact, QIP)	[DIAZ1997][HALEY1996]
Requirements testing	[WATANABE1997][HAMMER1997] [DUMOUSQUET1999]
Measurement	[ELBALUM1995]
Data mining	[BHANDARHI1993][BHANDARHI1993]
Design models/COCOMO	
Programming environments/Reuse	
Static analysis	[TANAKA1998][CARD1985]
Cleanroom	[LINGER1993]
Debugging/Fault trees/Model checking	
Reading	[BASILI1994]
SCR/value chaining/Audits	
Capture/recapture	[VANDERWEIL1993]
Middleware/Prototyping/Simulation/Slicing	

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

17

10/20/99

Productivity

Inspections and reviews	[JOHNSON1997][PERPICH1997][GRADY1994][PORTER1997B] [OGASAWARA1996][BRIAND1998][BUSH1999][PORTER1995] [PORTER1997A]
Models (e.g., cost, defect, schedule, complex)	[RAMAMOORTHY1984][RUBIN1985][REIFER1985][FERRENTINO1981] [PILLA1997][STENSURD1995][CHAN1996]
Testing (e.g., mutation, structural, functional, reading, integration)	[OFFUTT1993][HUTCHINS1994][MEMON1999][ROTHERMEL1998A] [WONG1995][PIWOWARSKI1993][DUNIE TZ1997][DUESTERWALD1996]
IV&V general	[ARTHUR1999][CARD1987][BELFORD1979][BOEHM1979][CHEN1981] [LEWIS1992]
Defect tracking	[SHOOMAN1975]
Mgmt issues (e.g., personnel, env)	[BROOKS1975][VOSBURGH1984][DEMARCO1985][UEMORA1984]
Process (CMM, PSP, ISO9000, Ex.Fact, QIP)	[DIAZ1997][HALEY1996][HAYES1995][JOHNSON1994]
Requirements testing	[JAGADEESAN1997][LEDERER1998][SHEPPERD1997]
Measurement	[TANAKA1995]
Data mining	[GILL1991]
Design models	[DAUSBOURG1998][TAMAI1993][ZWEBEN1995][HOWES1984]
COCOMO	[BOEHM1981][BOEHM1984]
Programming environments	[HATTORI1985]
Reuse	[TRACZ1994][GAFFNEY1992]
Static analysis	[CARD1985]
Cleanroom	[LINGER1993]
Debugging	
Fault trees	[LEVESON1983]
Model checking	[ALUR1997][HOLZMANN1999]
Reading	
SCR	[CHMURA1988]
Value chaining	[BOEHM1987][BOEHM1985]
Audits/Capture recapture/Middleware	
Prototyping	[BOEHM1984]
Simulation	
Slicing	[NISHIMATSU1999]

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

18

10/20/99

Reliability

Inspections and reviews	
Models (e.g., cost, defect, schedule, complexity)	[RAMAMOORTHY1984][FINK1995] [SHIMA1997][MIYOSHI1993][DUVAL1998] [HATTON1994]
Testing (e.g., mutation, structural, functional, reading, integration)	[PODGURSKI1997]
IV&V general	[BOEHM1979]
Defect tracking	[SCHNEIDEWIND1999] [SCHNEIDEWIND1998]
Mgmt issues (e.g., personnel, environments)	
Process (e.g., CMM, PSP, ISO 9000, Experience Factory, QIP)	[MCGARRY1995]
Requirements testing	
Measurement	[OGASAWARA1996][POORE1988]
Data mining/design models/COCOMO	
Programming env/Reuse/Static analysis	
Cleanroom	[SHERER1996]
Debugging	[OHBA1989]
Fault trees/Model checking/Reading/SCR	
Value chaining/Audits/Capture recapture	
Middleware/Prototyping/Simulation/Slicing	



Scheduling

Inspections and reviews	[PORTER1997][CONRAD1999][SEAMAN1997]
Models (e.g., cost, defect, schedule, complexity)	[TAKADA1994]
Testing (e.g., mutation, structural, functional, reading, integration)	
IV&V general	
Defect tracking	[ATKINSON1995]
Management issues (e.g., personnel, environments)	[BROOKS1975][TSUDA1992]
Process (CMM, PSP, ISO 9000, Ex. Fact., QIP)	
Requirements testing/Msmt/Data mining/Design models	
COCOMO	[BOEHM1981]
Programming env/Reuse/Static anal/Cleanroom	
Debugging/Fault trees/Model checking/Reading	
SCR	[NORCIO1986]
Value chaining	
Audits	[APRIL1998]
Capture/recapture/Middleware/Prototyping/ Simulation/Slicing	



Internal consistency measures

Inspections and reviews	[STEIN1997]
Models (cost, defect, schedule, complex)	[WANG1998]
Testing (e.g., mutation, structural, functional, reading, integration)	[NAUMOVICH1999][PAVLOPOULOU1999]
IV&V general	[NEAL1999]
Defect tracking	[DALAL1993][CHILLAREGE1992]
Management issues (e.g., personnel, environments)	[BASIL1979][BOEHM1976] [ZELKOWITZ1996][RIDDLE1985]
Process (CMM,PSP,ISO9000,Ex Fact,QIP)	[BASIL1994][KITSON1993][PAULK1993][PAULK1995]
Requirements testing	[REESE1980][WILSON1997]
Measurement	[LUVILA1995]
Data mining	[COOK1998]
Design models/COCOMO	
Programming environments	[HOUGHTON1985][EICKELMANN1996]
Reuse	[LEE1997]
Static analysis	[AVRUNIN1997]
Cleanroom	
Debugging	[DUCASSE1999]
Fault trees	[BOEHM1991]
Model checking	
Reading	[LAITENBERGER1999]
SCR/Value chaining/Audits/Capture-recapture	
Middleware	[MCGREGOR1995]
Prototyping	
Simulation	[RASK1993]
Slicing	



Program size measures

- A few papers dealt with converting from effort or defects to program size estimates:
 - [BELADY1976]
 - [WALSTON1977]
 - [EBRAHIMI1999]
 - [KITCHENHAM1993]
 - [LEHMAN1995]



Remaining literature search work

- A few additional citations need to be checked -- particularly earlier references in the IEEE Transactions of Software Engineering
- This should be completed within a month
- But, we don't expect the few additional references to materially change any of the results here



Results from literature search

- The number of papers that contain quantifiable data on software engineering technologies is quite small.
- Most of the papers address techniques "in the small."
- Testing techniques and inspections seem to be a favorite topic of investigation among university researchers.
- The NASA model of IV&V does not fit the usual approach discussed in the literature.
- It is outside of the scope of the current study, but initial investigations into the data that is kept on the shuttle program indicates that all relevant data may not be readily available for developing a quantitative model of IV&V as part of the next phase of the project



What data is needed to compute an ROI?

- **Who needs data affects what needs to be collected**
- **Who desires this data? Who is the customer?**
 - NASA management - Has the Fairmont facility been effective?
 - Shuttle management - Have shuttle \$\$\$ been spent wisely?
 - Developer management - How to maximize profit?
- **For what purpose is data needed?**
 - Characterization - How has budget been spent in past?
 - Evaluation - Has money been spent wisely?
 - Prediction - What will the spending profile be in the future?
- **Each has an impact on what data is needed**
 - GQM (Goal-Question-Metric paradigm) way to propose this clearly

25

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Goal Question Metric Paradigm (GQM)

- **A mechanism for defining and interpreting operational, measurable goals**
 - Developed by Basili and Rombach at Maryland in late 1980s
- **It uses four parameters:**
 - a model of an **object of study**,
a process, product, or any other experience model (e.g., **IV&V activities**)
 - a model of one or more **focuses**,
models that view the object of study for particular characteristics (e.g.,
effects on cost)
 - a **point of view**,
the perspective of the person needing the information (e.g., **management**)
 - a **purpose**,
e.g., how the results will be used (e.g., **evaluation**)

to generate a GQM model relative to a particular environment

26

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Using GQM

- **Steps in a GQM analysis:**
 1. **Formulate goals**
 2. **Develop and refine subgoals and questions**
 3. **Establish appropriate metrics**
 4. **Plan investigation layout and analysis methods**
 5. **Design and test data collection scheme**
 6. **Perform investigation concurrently with data validation**
 7. **Analyze data**
- **Risks:**
 - **Items in blue under our control**
 - **Critical items in red (5, 6) not under our control**



Characterization of IV&V activities

- **Focus:**
 - **Analyze** NASA shuttle IV&V activities **for the purpose of** characterizing **with respect to their** effect on costs of each operational increment **from the point of view** of NASA management
- **Questions:**
 - How are resources being spent for shuttle software?
- **Data needed:**
 - Total costs of each OI of developed shuttle software
 - Total costs of IV&V activities
- **Use:**
 - Can characterize expenditures for IV&V activities
- **Evaluation:**
 - This data is probably available



Evaluation of IV&V activities

- **Focus:**
 - **Analyze** NASA shuttle IV&V activities **for the purpose of** evaluating **with respect to their** effect on costs of each operational increment **from the point of view** of NASA management
- **Questions:**
 - How much does each IV&V activity cost?
 - Is each IV&V activity effective in finding defects?
- **Data needed:**
 - Total costs and IV&V costs of each OI of developed shuttle software
 - IV&V activities applied to each increment
 - Defects found for each IV&V activity
- **Use:**
 - Can evaluate expenditures for IV&V activities
- **Evaluation:**
 - Not clear if costs can be associated with specific IV&V activities

29

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Prediction of IV&V activities

- **Focus:**
 - **Analyze** NASA shuttle IV&V activities **for the purpose of** predicting **with respect to their** effect on costs of each operational increment **from the point of view** of NASA management
- **Questions:**
 - What IV&V activities will be most effective for future OIs?
- **Data needed:**
 - Developer costs and IV&V costs of each activity for each OI
 - Defects found by developer and by IV&V for each OI
 - Effectiveness of each technique in finding defects or reducing costs (from literature or NASA data)
- **Use:**
 - Can develop a predictive model of ROI
- **Evaluation:**
 - This data probably is not available

30

10/20/99

Fraunhofer USA



Center for Experimental
Software Engineering,
Maryland

Risks

- Developer
 - Views process as auditing process to find excess funding
 - Where can budget be cut?
- IV&V contractor
 - Will recompute contract in 2 years.
 - Data may be proprietary
- Both groups view cooperation as a downside risk.
 - What is upside benefits to both?



What data is collected?

- IV&V Data
 - ITRs - Issues tracking reports. Discrepancies occurring during requirements analysis and carried through rest of OI
 - DR - Discrepancy reports, when occur during IV&V process
 - Effort data -- not sure what is really collected here
- Developer data
 - Effort data, but not sure if available and at what granularity
 - DRs - if during integration testing and later
 - Error reports during code development are not reported. Not sure if they exist.



Project

- **Due:** December 1, 1999
 - **Requirement:** Develop a measurement program for the NASA IVV center to evaluate its economic benefit to NASA
 - Develop GQMs for this process
 - Develop a series of questions from these GQMs
 - Develop a series of data items to collect
 - What changes should be made to current NASA data collection to help in the future?
 - What is the probable impact on this proposal?
 - Be sure to include GQMs that allow for a win-win situation where the developers have an incentive to cooperate with the research program.
 - **Length:** As appropriate. This should be a well written report, not a series of questions to be answered.
 - **Submission:** Email postscript, pdf, or word document. If not possible, turn in hard copy.
-

