SIMPROS 2002

IV Simpósio Internacional de Melhoria de Processo de Software

Anais

Anais do SIMPROS 2002

ASSESSING SOFTWARE PROCESSES: THE RATING DILEMMA BETWEEN MEASUREMENT AND HUMAN JUDGMENT

Fabrizio Fabbrini, Mario Fusani, Giuseppe Lami
IEI – CNR, Pisa, Italy

Determining to what extent actual software development activity reaches its purposes is a major concern of all the people involved, no matter what their responsibility level is. Various reference models for software processes and their capability levels have been proposed in technical literature, as well as a set of requirements and guides for assessing the capability of the processes against the models. This talk is aimed at discussing possible methods for carrying out an important phase, the *rating* phase, of an assessment conformant to one of the most popular and promising reference standard, ISO/IEC 15504 (SPICE). The focus is on the techniques to be used to evaluate the software processes under assessment at defined capability levels, on the basis of collected evidence. The ISO/IEC 15504 - TR2 document set (1998), in its Part 5 [1], provides guidance (with examples) of what is a *compatible* assessment model, among all the possible models acceptable according to the standard. Although this model contains a description of indicators and criteria to be adopted in an assessment process, a good deal of flexibility and relative freedom in interpreting the model still remains for the assessor. This regards, among other things, actions to be planned and executed for rating the Process Attributes associated to any capability level. Different *rating methods* can be applied which would not be in contrast with the compatible model presented in the various releases of the document set. It is not possible to describe the complete variety of such methods, which can be referred to as *valid* methods. On the other hand, it would be unpractical, nor general, if just a single method were introduced: in fact, giving just one method would reduce further interpretation choices to a minimum, thus the user would get maximum guidance. However, hardly a given method can be applied to different scenarios that may occur in real assessments.

Then, what is the way to keep generality without loosing practicality?

The solution adopted here is to give the *requirements* to be satisfied for a rating method to be a valid one.

The requirements are such that any valid rating method, i. e. any rating method which satisfies the requirements, also satisfies the criteria for a compatible assessment model described in the guidance documents associated with the standard, but there is no reason why some of these requirements may be non-applicable in particular assessments scenarios. Some of the requirements also introduce trade-offs in their achievement, thus different requirement prioritisation may occur in different assessment contexts.

The presentation is organised as follows:

I. Introductory concepts are presented and discussed, which are thought of as the basis for generating any requirement for a rating method.

II. On such grounds, a set of requirements for a rating method (such as repeatability, cost effectiveness, functionality, a sub-set of quality requirements, etc.) is proposed and discussed.

III. Some characteristics of data collection are described, which are likely to affect the achievement of the requirements.

IV. Various schemes for a rating method are presented and discussed. This is the central part of the talk, in which the problem is tackled of how to deal with quantitative, calculable data to perform objective rating (or *measurement* according to some measurement theory [2]) without spoiling that intuitive, precious talent typical of well-experienced assessors. The standard document set provides little help in such a crucial step of the assessment process: So, various possible solutions, ranging from the mere processing of measured indicators up to the unaided assessors' judgment are checked against the proposed set of requirements. This is done by identifying how each rating solution affects the degree of achievement for each requirement. A criterion can be derived to identify the risks that the rating process, and then the assessment process, do not reach their stated goals.

V. Criteria are given to determine the applicability of Process Attribute indicators (such as those suggested by the standard document set) depending on the assessment context and assessment constraints. It is shown how this aspect plays a role in the rating method itself.

The talk presents an interim status of a research about assessing techniques and should not be taken as a generator of complete rating methods, but rather as a guidance for the assessors to choose the method that best satisfies a set of privileged requirements for rating.

References

[1] ISO/IEC TR 15504 (1998) - Part 5: An assessment model and indicator guidance

[2] Fenton NE and Pfleeger SL, 'Software Metrics: A Rigorous and Practical Approach', PWS, 1998

Fabrizio Fabbrini obtained his degree in Computer Science from the University of Pisa, Italy, in 1974. Since 1975 he has served as a scientific researcher at the Institute for Information Processing (IEI) of the Italian National Research Council (CNR), where now he is Senior Researcher and coordinates the Software Laboratory of the Center for Software Certification. Fabrizio Fabbrini's present activity is focused on Software Quality, and more precisely on the development of methodologies and standards for the assessment and the evaluation of software products and processes, with particular

attention to Software Engineering Standards and Software Certification. Software Process Assessment & Improvement, Software Verification & Validation, Computer Security & Data Privacy represent the main fields of application of such research activities.

Mario Fusani obtained his degree in Electrical engineering from the University of Pisa, Italy, in 1971. Since 1973 he has served as a scientific researcher at the Institute for Information Processing (IEI) of the National Research Council (CNR), where now he is Senior Research. His present activity is focused on Software Quality, including the development of methodologies and standards for the assessment and the evaluation of software products and processes. Since 1999 he has been the Scientific Coordinator of the Center for Software Certification of the Italian National Research Council.

Giuseppe Lami graduated in Computer Science in 1994 at the University of Pisa (Italy). He participated in the "Leonardo da Vinci – Software Quality Evaluator" European project (1997-1999) and in the "Software Product Evaluation and Certification" project founded by the European Space Agency (1998-1999). Since year 2000, he has been working at the Istituto di Elaborazione dell'Informazione (Institute for Information Processing) of the Italian National Council for Research. His research interests are: software process assessment (the SPICE ISO15504 model), Requirements Engineering (natural language requirements quality evaluation), Software Testing (coverage methods). Since 1998 has given lectures in Fundamentals of Computer Science at the Faculty of Telecommunication Engineering of the University of Pisa. Since 2001 he has been a Phd student in Informatics Engineering at the University of Pisa.

ASSESSING SOFTWARE PROCESSES: THE RATING DILEMMA BETWEEN MEASUREMENT AND HUMAN JUDGMENT

Fabrizio Fabbrini, Mario Fusani, Giuseppe Lami
ISTI – CNR, Pisa, Italy
fabbrini@iei.pi.cnr.it, fusani@iei.pi.cnr.it, glami@iei.pi.cnr.it

1

Process attribute rating

- Rating
- Requirements for rating
- Strict measurement vs. experience and intuition
- Risks
- Conclusions

2

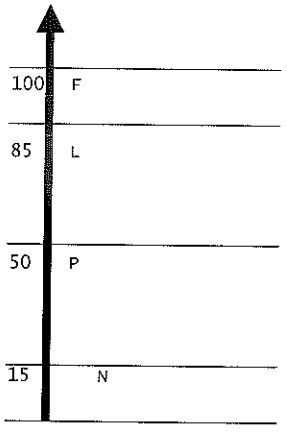
Rating

- capability determination through attribute measurement according to a defined method

Some measurements (examples)		
Object or entity	Attributes	Scales
Desk	*Length, etc.*	*Ratio*
Process	*Capability, Proc. Attribute*	*Ordinal*

3

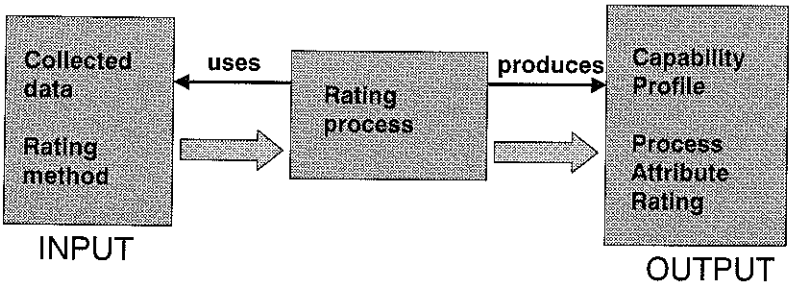
Rating output domain



- Scale: {N,P,L,F}
- Scale: {0, 1, …, 100}
- "mapping"

4

Rating process



5

Rating for ISO/IEC 15504

- In spite of:
  - ISO/IEC 15504 requirements
  - Conformance (process reference model (e.g.: 12207) and applicable model (e.g.: TR2 Part 5))

Many acceptable solutions are still possible for defining a Rating method

6

Rating: what method?

- To define requirements for rating
  - Source and criteria: Standard, Quality, Applicability
- To choose a Rating method that minimizes risks of requirement missing
  - Various solutions to be expected

7

Rating: requirements

1	**Conformity to Part 2 and Part 3** (ISO/IEC TR 15504 (1998))
2	**Conformity to Part 5 interpretation of the reference model (use concepts and features of Part 5)**
3	**Quality** **Usability** **Efficiency** **Quality of the results** **Objectivity** **Repeatability** **Meaningfulness of the measures**
4	**Functionality**
5	**Completeness**

8

Rating: requirements (cont.)

6	Consistency
7	Effective Redundancy
8	Providing measurable Indicators
9	Cost Effectiveness

9

Elements used by a Rating process

ELEMENT	ROLE
purpose of the software process under assessment outcomes from successful process performance capability attributes	Known by: MODEL
ideal status of Indicators (*as defined in a compatible model*)	Known by: MODEL
actual status of Indicators (*as defined in a compatible model*)	Known by: COLLECTED EVIDENCE

10

Elements used by a Rating process(cont.)

ELEMENT	ROLE
Context (particular situation of the assessment or of the Organizational unit to be assessed)	May affect indicators and checking criteria
Assessor's judgment	Rating instrument
Procedures (include checklists and model vs. actual checking criteria)	Rating instrument

11

Indicators

- Work Products (input, output)
- Base Practices (BP)
  - Testify process performance
- Management Practices (MP)
  - Associated to Process Attributes at each capability level
- Interviews
- Supporting processes

12

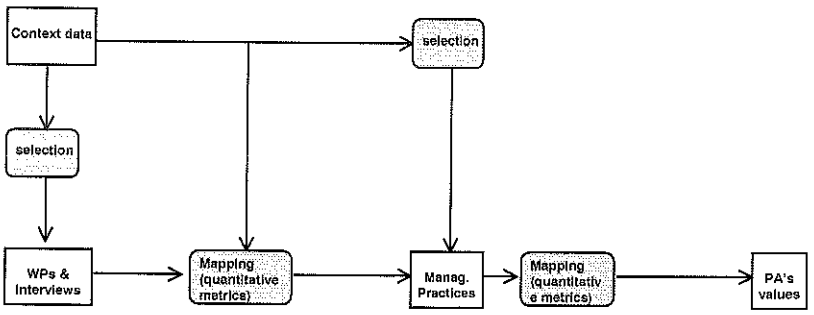
How to use Indicators

- Ideal vs. actual (measures)
  - Existence (binary), completeness (ratio), quality profile (ordinal)
- Mapping between Work Products and other Indicators
- Mapping between interview results and Indicators
  - Consistency
- Priority
- Context dependence
- Measurement (with normalization and integration)
- Validation

= Input to assessor judgment

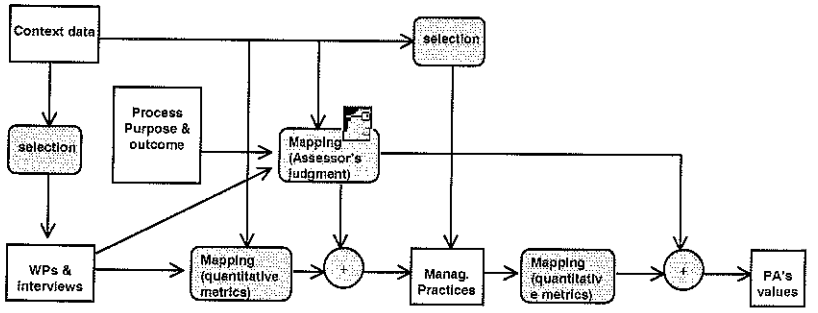
13

Rating methods: analytical



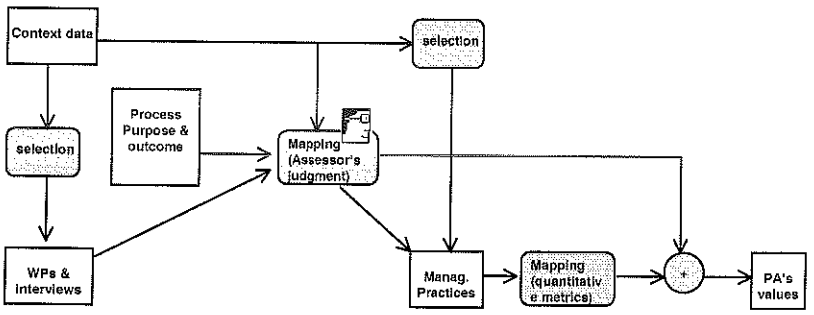
14

Rating methods : composite

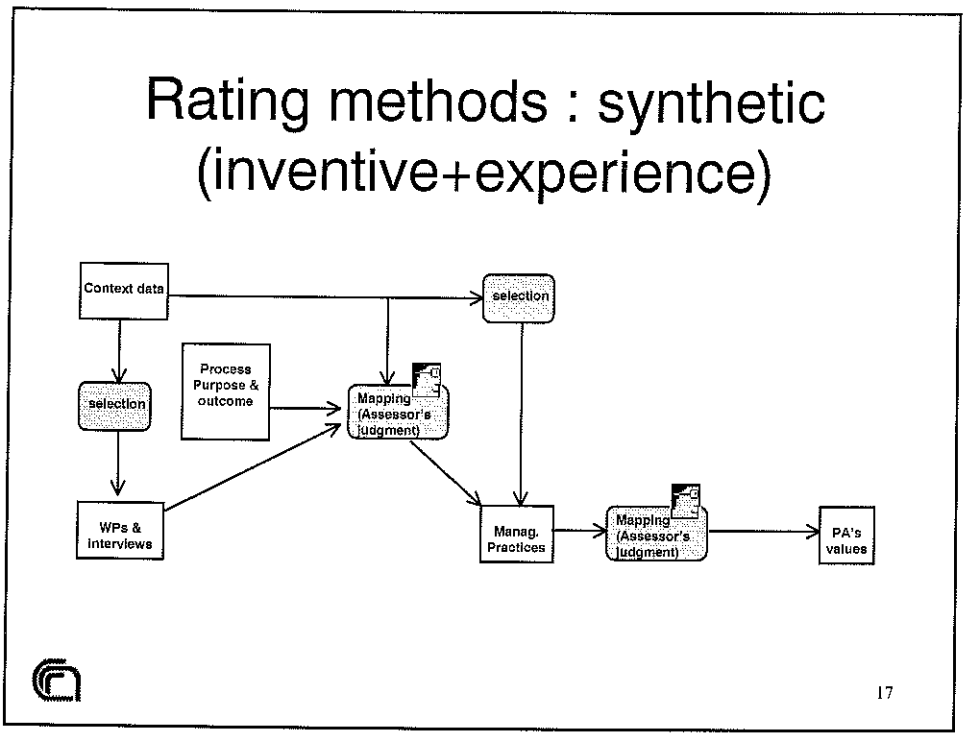


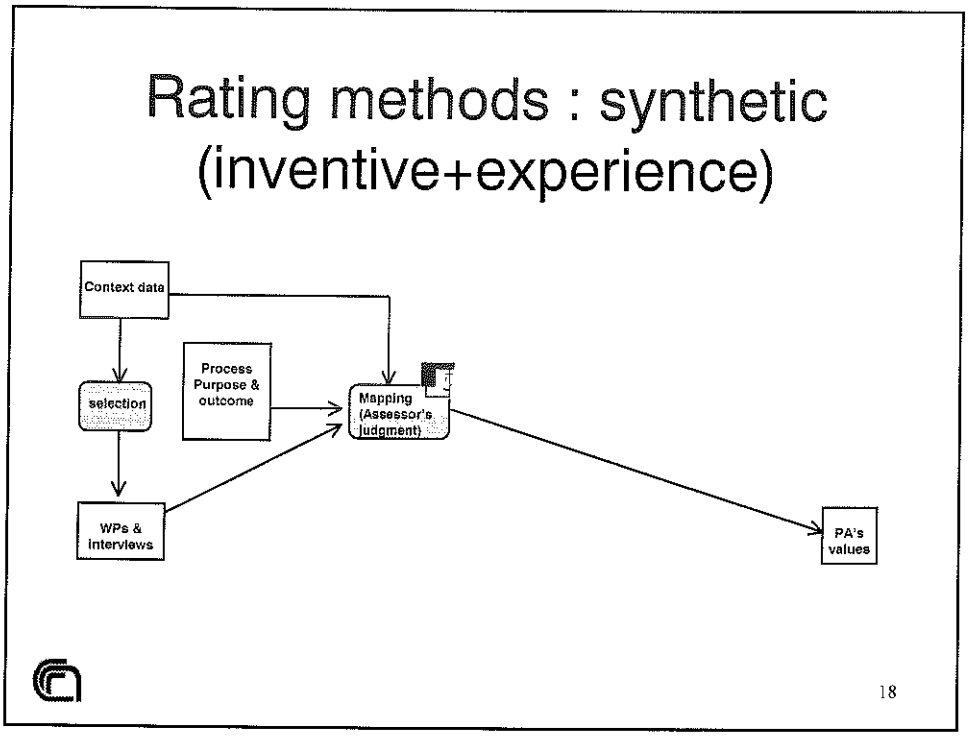
15

Rating methods : composite



16


Rating methods : synthetic (inventive+experience)
Context data
selection
Process Purpose & outcome
selection
Mapping (Assessor's judgment)
WPs & interviews
Manag. Practices
Mapping (Assessor's judgment)
PA's values
17


Rating methods : synthetic (inventive+experience)
Context data
selection
Process Purpose & outcome
Mapping (Assessor's judgment)
WPs & interviews
PA's values
18

Risks of NOT satisfying requirements

	Methods		
Requirements for Rating	Analytical	Composite	Synthetic
Repeatability	*Low risk* 	*Medium risk* 	*High risk* 
Cost effectiveness	*Medium risk* 	*Medium risk* 	*Low risk* 
Meaningful results	*High risk* 	*Medium risk* 	*Low risk with expert assessor* 

19

Conclusions

- Assessors play a central role
- Measurements are just an aid
- Understand Psychology of Evaluation and use into assessment
- Always keep requirements in mind
- Always keep assessment risk in mind

20