

# SPICE in retrospect: Developing a standard for process assessment

Terence P. Rout <sup>a,\*</sup>, Khaled El Emam <sup>b</sup>, Mario Fusani <sup>c</sup>,  
Dennis Goldenson <sup>d</sup>, Ho-Won Jung <sup>e</sup>

<sup>a</sup> *Software Quality Institute, Griffith University, Queensland, Australia*

<sup>b</sup> *University of Ottawa, Ottawa, Canada*

<sup>c</sup> *ISTI-CNR, National Research Council, Pisa, Italy*

<sup>d</sup> *Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, USA*

<sup>e</sup> *College of Business Administration, Korea University, Seoul, South Korea*

Received 1 June 2006; received in revised form 10 January 2007; accepted 10 January 2007

Available online 11 February 2007

## Abstract

The SPICE Project was established in 1993 to support the development, validation and transition into use of an International Standard for software process assessment. Its efforts have resulted in the publication of a five-part Standard for Process Assessment, ISO/IEC 15504. This paper reviews the evolution of the Standard, and reflects on the parallel achievements of the SPICE Project and the standardisation effort in advancing the state of the art in process assessment and improvement.

© 2007 Elsevier Inc. All rights reserved.

*Keywords:* Process assessment; Process improvement; Standardisation

## 1. Introduction

ISO/IEC 15504 (2003) is the International Standard for Process Assessment. Its development, with the parallel empirical studies of its use by the SPICE Project (Dorling, 1993; Rout, 2003), has spanned 14 years – the initial Study Group established by JTC1/SC7 to explore the needs and requirements for the standard established in 1992 (ISO/IEC JTC1/SC7, 1992). With the publication of the final part of the Standard it is an appropriate point to review the history and impact of its development.

### 1.1. Origins

The primary impetus for the use of assessment did not come from the mainstream of the software development industry, but rather from acquirers of large, critical soft-

ware-intensive systems – notably in the defence and telecommunications sectors. Thus, the Capability Maturity Model for Software (Paulk et al., 1993) was developed by the Software Engineering Institute as a response to the needs of the US Defense Department for better techniques for the selection of contractors. Process assessment methods were also developed by a number of the major players in the telecommunications field, including British Telecom, Bell Canada/Northern Telecom and Bellcore, and applied to the management of risk in acquisition.

The increasing number of assessment approaches available and the increasing use of the technique in commercially-sensitive areas were the key motivating factors behind the development and acceptance of a proposal to develop an International Standard for software process assessment. The UK Ministry of Defence, through the Defence Research Agency, initiated a series of studies resulting in a firm proposal to develop a Standard. This was carried through to the international committee on software engineering standards, ISO/IEC JTC1/SC7, which in 1993 adopted a proposal to develop the Standard and established a Working Group (WG10) to undertake the work.

\* Corresponding author. Tel.: +61 7 3735 5046; fax: +61 7 3735 5207.

E-mail addresses: [T.Rout@griffith.edu.au](mailto:T.Rout@griffith.edu.au) (T.P. Rout), [kelemam@kelemam.ca](mailto:kelemam@kelemam.ca) (K. El Emam), [mario.fusani@isti.cnr.it](mailto:mario.fusani@isti.cnr.it) (M. Fusani), [dg@sei.cmu.edu](mailto:dg@sei.cmu.edu) (D. Goldenson), [hwjung@korea.ac.kr](mailto:hwjung@korea.ac.kr) (H.-W. Jung).

A key element of the proposal was a recommendation to adopt a development route which would produce useful output in a time-scale which was shorter than the normal standards-making route. Several factors influenced this decision; in particular, there was seen to be an urgent need for the new Standard to aid the harmonisation of approaches to assessment being developed by various organizations, projects and initiatives. In addition, organizations developing and using new or existing assessment methods had indicated a willingness to pool resources provided that a common approach can be developed quickly. As a result, the Standards working group charged with the work authorised the establishment of a dedicated Project Team, known by the acronym SPICE (Software Process Improvement and Capability dEtermination). The SPICE Project, as originally established, had three defined goals:

1. To assist the standardisation project in its preparatory stage to develop initial working drafts.
2. To undertake user trials in order to gain early experience data which will form a basis for revision of the published technical reports prior to review as full International Standards.
3. To create market awareness and take-up of the evolving Standards.

### 1.2. History

The SPICE Project developed the initial set of working draft documents between 1993 and 1995, and these were then submitted to the normal standards balloting process. There was considerable refinement of the original architecture during the ballot process, and the first version of the Standard was released as a Technical Report (Type 2) in 1998 (ISO/IEC TR 15504, 1998).

During the course of these ballots, the SPICE Trials pursued a detailed series of investigations, aimed at demonstrating the extent to which the new International Standard met its original requirements, and validating its usefulness to the software development industry. The results of the Trials provided substantial empirical evidence supporting the approach to assessment embedded in the Standard, and also identified significant ideas for improvement.

A study period was initiated following publication of the Technical Report to review the requirements with a view to revision and publication as a full International Standard; this work commenced in 2000, and the standard was published in five parts over the period 2003–2006 (ISO/IEC 15504, 2003).

## 2. Achievements

### 2.1. A discipline for process assessment

At the time that development of ISO/IEC 15504 commenced, there were many views of the technique of process

assessment – one of the justifications for the Standard. The key features recognised by the Study Group report were that the technique represented “the disciplined examination of the processes used by an organization against a set of criteria to determine the capability of those processes to perform within quality, cost and schedule goals.” The concept of evaluating the capability of an organization and its processes by comparison of performance against some form of process model was central to the concept. However, the techniques applied in performing this comparison were varied.

The technique of process assessment as practiced within the domain of software engineering has its origins in the work of Radice et al. (1985) and was further advanced through a series of methods developed at the Software Engineering Institute. A common thread throughout is the comparison of the actual performance of a process to some form of structured process model that serves as a yardstick, allowing the development of a “rating” of maturity or capability. The early work emphasised the use of interview techniques to elicit evidence from practitioners, and there was limited use of documentary forms of evidence. With increasing use of assessment results in commercially sensitive areas, there was an increasing focus on formality in the assessment process.

In the development of ISO/IEC 15504, the need to maintain consistency with other international standards led to a significant alignment to the quality audit approaches derived from certification to ISO 9001. However, there was always in the minds of the development team a clear differentiation between *audit* and *assessment*. Audits were seen as concerned primarily with conformance; assessments were seen as concentrating on two measures – conformance and effectiveness (ISO/IEC JTC1/SC7, 1992).

In ISO/IEC 15504, a disciplined approach to process assessment has been defined, characterising a coherent definition for the assessment process. The Standard prescribes minimum requirements for inputs, outputs, resources, and activities. It defines the need to establish a strategy for collection of data, and to base the ratings of process capability on objective evidence. The basic requirements for performing process assessment have had significant influence; the concept of a “conformant assessment” as the basis for demonstration of process capability has influenced work such as SCAMPI (Standard CMMI Appraisal Method for Process Improvement (SEI, 2001)), as well as several other approaches more closely aligned to the Standard.

### 2.2. Measuring process capability

In this section, the set of requirements for evaluating process capability established in ISO/IEC 15504 is presented, and the way in which they contribute to the goal of objectivity in process assessment is described. The Standard provides a comprehensive and generic approach for the model-based assessment of process capability; the

assessment proceeds through the comparison of evidence of actual performance against a model of capability for a process or set of processes.

### 2.2.1. Requirements for measuring capability

The set of requirements as a whole define a structure that establishes a well-defined process for performing process assessment. They fall into two broad classes: those associated with the performance of the assessment itself, and those concerned with the process model that forms the basis of the assessment.

The requirements for the performance of the assessment are defined in terms of:

*Objects* needed for performing the assessment, including:

- The process model that forms the basis of the evaluation.
- The objective evidence of actual performance collected in the course of the assessment.
- The categories for which a measurement is established, represented by Process Attributes (defined in the Standard).

*Actors* in the measurement framework, which include:

- Assessors.
- Sponsors.
- Other stakeholders (such as customers and suppliers: they are mentioned for describing the environment in which SPICE is used, but do not have normative duties, properties and relations).

*Operations* defined on the objects (also involving actors and other elements). These include:

- Planning of the assessment, including the definition of a required minimum set of inputs for the assessment process.
- Data collection.
- Data validation (the objective of validation is to gain confidence in the data sources, hence in rating results).
- Rating of Process Attributes.
- Assessment reporting (input, output, activity).

The requirements for the process model in turn can be placed in three distinct classes:

- Those concerned with the entities that are the subject of the assessment (processes).
- Those concerned with the measurement scale for evaluating the capability of these entities; and
- Those concerned with the assembly of these two elements to form the process model.

The *Processes* that are defined as a comparison target for a particular assessment are defined in a Process Refer-

ence Model (PRM). The Standard specifies the contents and basic structure of process reference models; in essence, these are collections of definitions of processes for a particular domain or community of interest. Each process in a PRM is described in terms of its purpose and a list of the outcomes of its implementation.

The *Measurement Scale* for the evaluation of capability is described in terms of a Measurement Framework that uses descriptions of Process Attributes achieved in implementing a process to determine a scale of levels of capability. The framework formally defines an ordinal measurement scale for process capability.

The *Model* that is used in the assessment is defined as a Process Assessment Model (PAM). A PAM is a two-dimensional representation of process capability. One dimension (the process or functional dimension) is related to processes drawn from one or more PRMs. The other dimension (the capability dimension) is related to the Measurement Framework. There may thus be references to multiple PRMs in a single PAM, and also multiple PAMs that relate to a single PRM. This feature establishes the generic nature of the overall structure of the assessment process.

Distinguishing features of the Standard are that:

1. These types of requirements are kept distinct from each other although they have defined and strict relationships to each other.
2. All types of requirements are independent of the nature of the processes to be assessed (e.g., the application domain). This may be obvious for the last two types (Measurement Scale and PAM). It is a valuable feature of SPICE that also the requirements for PRM are independent of what processes shall be like.

The chosen levels of abstraction for the various types of requirements allow for broader application field insofar PRM's are concerned, which reflects on the performance Attribute at Capability Level 1 (that is a mere process performance requirement), and on a somewhat accurate evaluation of the capability levels.

Note that the scope of an ISO/IEC 15504 assessment is no longer seen as limited to assessing the capability of processes for developing and maintaining software in an organization. The history of SPICE has gradually departed from a software oriented effort, to a more mature status that is not even limited to a broader systems engineering view, as for CMMI, but to a defined set of requirements abstract enough to be suitable for many human organized activities with defined goals, yet clear enough to allow for profitable process assessments (Coletta et al., 2005).

### 2.2.2. Pursuing objectivity in measurements

In order for an assessment to generate and transfer confidence on its results, the assessment must include significant measurement. That is, the entities to be measured (in our case, Process Attributes) must possess properties

that can be mapped to a set of values of defined type (for example: integer, ordinals, elements of a set) (Fenton and Pfleeger, 1996). SPICE complies with this theory-of-measure-related requirement by establishing a mapping between a process attribute and an ordered set of values {N (not achieved), P (Partially achieved), L (largely achieved), F (fully achieved)}.

The assessors are the principal actors bearing the responsibility of a meaningful (objective, reliable and repeatable) process attribute rating (Fabbrini et al., 2002a). To deal with the effects of subjectivity in this measurement process and reduce uncertainty with the results, SPICE provides:

- A defined set of rules for determining assessor experience and skill together with procedures for verifying their compliance.
- Requirements for (in normative documents) and examples of (in guidance documents) checkable *indicators*. These record the types of objective evidence that link to the PRM and Measurement Framework elements, and permit objective judgment of the achievement of the process purpose and attributes. Indicators can generally be distinguished as Performance Indicators (such as kinds and qualities of process outputs) or Capability Indicators (such as evidence of management practices both planned and in place); ISO/IEC 15504-5 provides several types of both classes, as examples of the types of indicators that can be established.
- Requirements for documenting the assessment process (including criteria for any decision made by the assessors and the records of their findings).

Results of analytical studies on documented assessment outputs that corroborate these features of the Standard are discussed in Section 2.4.1.

### 2.2.3. Conclusions

Since the requirements in ISO/IEC 15504 for all of the principal elements of the assessment process – the Process Reference Model, the Measurement Framework and the Assessment Model – are established at a high-level, SPICE can be adopted in a wide variety of domains. The Standard provides both mandatory elements and guidance for achieving objective, repeatable assessment results.

## 2.3. Empirical studies

### 2.3.1. SPICE trials

The purpose of this section is to present major parts of the findings of the empirical studies conducted as part of the SPICE Trials, mainly during Phase 2 of the Trials (September 1996–June 1998). Unique among software engineering standardization efforts, the developers of ISO/IEC 15504 deliberately initiated an international effort to empirically evaluate the Standard. This effort is known as the SPICE Trials (El Emam and Goldenson, 1995; MacLennan

et al., 1998). A considerable number of empirical evaluation studies have been conducted, especially during Phase 2 of the SPICE Trials. The final report of SPICE Trials has been published (SPICE Project, 2003) and the most significant studies have been summarized (Jung et al., 2001). An annotated bibliography of published papers has also been made available, as Part 3 of the final report (Hunter and Jung, 2003).

The original trials plan organized the SPICE Trials into three broad phases as follows:

*Phase 1* took place in 1995 and its goals were to validate the design decisions inherent in the initial document set as well as to test the usability of the core product documents (SPICE version 1 – the complete set of Working Draft documents (SPICE Project, 1995)).

*Phase 2* took place between September 1996 and June 1998 and was based on the PDTR (Preliminary Draft Technical Report) version of the emerging ISO/IEC 15504 Standard. In addition to evaluating the complete document set and design decisions, its objectives include providing guidance for applying the emerging Standard most effectively. Phase 2 of the SPICE Trials evaluates the ISO/IEC PDTR 15504 documents (SPICE version 2).

A third phase of the Trials was intended to commence in July 1998, continuing until the full publication of the International Standard. A lack of resources, and the increasing use of the Technical Report outside the scope of the trials, resulted in significant problems in the collection of relevant data, and in October 2000 this phase was abandoned; the Phase 2 report thus became the final report from the Trials (SPICE Project, 2003). Empirical studies continued, with significant involvement of members of the SPICE Trials.

### 2.3.2. Data collection

Phase 2 of the SPICE Trials used the regional structure defined for the project as a whole, which divides the world into five zones, serviced by Regional Trials Centers (RTCs), located in Canada (including Latin America), Europe (including South Africa), North Asia Pacific (centered on Japan and including Korea), South Asia Pacific (centered on Australia and including Singapore) and the USA. At the country or state level, Local Trials Coordinators (LTCs) liaised with the assessors and Organization Units (OUs) to ensure assessors' qualifications, to make the questionnaires available, to answer queries about the questionnaires, and to ensure the timely collection of data. There were 26 such coordinators worldwide during the second phase of the SPICE Trials (SPICE Project, 2003).

The dataset submitted to the International Trials Coordinator (ITC) for each trial included the ratings data from each assessment and answers to a set of questionnaires that followed each assessment. Lead assessors and OUs completed the questionnaires related to the assessment, the OU, the project, etc. During the Phase 2 Trials, there were 70 assessments of 44 organizations, 169 projects, and 691 process instances from the five regions: Europe (24 trials), South Asia Pacific (34 trials), North Asia Pacific (10 trials),

USA (1 trial), and Canada/Mexico (1 trial) (Hunter, 1998; El Emam, 1999; El Emam and Birk, 2000a). Since more than one assessment occurred in some OUs, the number of OUs was less than the number of assessments.

The most common primary business sectors in which the organizations were involved were Defence, IT Products and Services and Software development. The assessments involved both large and small organizations and covered all the processes in the reference model. The median number of process instances per assessment was 6.5. The most costly activity during an assessment is the collection of evidence (47% of effort) and the least costly is the final presentation (3% of effort). The assessors involved had a broad range of experience with many having participated in phase 1 of the trials, used the SW-CMM, and/or been involved in ISO 9001 audits (mainly on the context of the TickIT scheme). Almost all of the competent assessors (93%) were received assessment training in the context of ISO/IEC 15504, in most cases.

The Phase 3 Trials did not publish its final report for evaluating the TR version, but some studies have been published in journals and proceedings. Most of studies in the Phase 3 Trials based on TR version were re-evaluation of issues of the Phase 2 Trials. The next section of this paper presents results of the Phases 2 and 3 together.

#### 2.4. Validating the standard

Through a set of Trials, the SPICE Project validated ISO/IEC 15504. This section briefly summarizes some important empirical studies based on TR as well as PDTR. The Phase 2 Trials Reports (Hunter and Jung, 2003; Jung et al., 2001; SPICE Project, 2003) provide additional details.

##### 2.4.1. Reliability of process capability dimension

Since process assessment involves a subjective measurement procedure, the reliability of this procedure is vital in order to have confidence in the assessment results. Reliability is defined as the extent to which the same measurement procedure yields the same results on repeated trials (Carmines and Zeller, 1979). Lack of reliability is caused by random measurement error. An important aspect of reliability is called *internal consistency*. Internal consistency is affected by ambiguities in wording and inconsistencies in the interpretation of wording by respondents. The most popular measure of internal consistency is Cronbach's alpha (Cronbach, 1951).

The capability measure of ISO/IEC 15504 was found to have as two dimensions in both PDTR and TR versions. The two dimensions based on the PDTR version are named as "Process Implementation" with the attributes from levels 1 to 3 (Cronbach's alpha, 0.88) and "Quantitative Process Management" with the attributes from levels 4 to 5 (Cronbach's alpha, 0.87), respectively (El Emam, 1998). For a Korean dataset based on assessments of TR version, the Cronbach alpha coefficient up to Level 3 had a high

value of 0.93 (Jung, 2003b). In both versions, the Cronbach alpha coefficient of each capability dimension showed a high enough internal consistency to be useful in practice. In addition, the 4-category scales of measuring process attributes, 'F', 'L', 'P', 'N', cannot be improved in terms of internal consistency by reducing it to a 3- or a 2-category scale (Jung, 2003b).

The ISO/IEC 15504 capability dimension in PDTR has also been evaluated by using a Guttman scaling method (Jung, 2005b). This study evaluated empirically whether the ordering of set of PAs, as measures of capability, is consistent with the Standard. For this purpose, the study estimates the Coefficient of Reproducibility (CR) statistic that measures the extent to which the observed ratings are identical to the pattern inferred by the Standard. Analyses based on ratings of 689 process instances show that generally PA order of capability levels is consistent with that inferred by the Standard.

Interrater agreement, called the external reliability, is used to show the extent to which two assessors or teams of assessors agree when making independent judgments about the same software engineering processes. Assuming nominal scales, Cohen's Kappa (Cohen, 1960) is the most popular index to describe the strength of agreement using a single summary index. A series of interrater agreement studies conducted as part of the international SPICE trials does show reasonably high levels of interrater agreement. Agreement was found to be almost always higher than "moderate agreement (0.60)." If it is accepted that moderate agreement is a minimum for practical usage, then these results are encouraging for users of ISO/IEC 15504 (El Emam et al., 1996; El Emam and Marshall, 1998; El Emam, 1999). More recent work by Jung (2003a) provides further discussion about paradoxes in the interpretation of the Kappa coefficient that is used in the SPICE studies.

##### 2.4.2. Predictive validity

A basic premise of ISO/IEC 15504 is that higher capability is associated with better project performance and product quality. Furthermore, improving higher capability is expected to subsequently improve both performance and quality. Testing this premise is an evaluation of the predictive validity of the assessment measurement procedure. This section is based on two studies conducted by El Emam and Birk (El Emam and Birk, 2000a,b) on predictive validity of process capability measures.

El Emam and Birk (2000a,b) evaluated the predictive validity of four processes shown in Table 1. They used a two-stage analysis procedure to evaluate the predictive validity. The first stage determines whether the association between "Process Implementation" (up to level 3) of the development processes and each of the performance measures is "clinically significant" using the Pearson correlation coefficient. This procedure indicates a magnitude that is sufficiently large. If it does, then the statistical significance of the association was exploited by utilizing an OLS (Ordinary Least Squares) regression. The OLS model is separately

Table 1  
Summary of the predictive validity study

| OU size             | Performance measure                       | Process(es)  |
|---------------------|---|--|
| Small organizations | Ability to meet budget commitments        | Develop software design (ENG.3)  |
|                     | Ability to meet schedule commitments      |  |
|                     | Ability to achieve customer satisfaction  |  |
|                     | Ability to satisfy specified requirements |  |
|                     | Staff productivity                        |  |
| Large organizations | Staff morale/job satisfaction             | Develop software design (ENG.3)<br>Implement software design (ENG.4)<br>Develop software design (ENG.3)<br>Develop software design (ENG.3)<br>Develop software design (ENG.3)<br>Develop software requirements (ENG.2)<br>Integrate and test software (ENG.5)<br>Develop software design (ENG.3) |
|                     | Ability to meet budget commitments        |  |
|                     | Ability to meet schedule commitments      |  |
|                     | Ability to achieve customer satisfaction  |  |
|                     | Ability to satisfy specified requirements |  |
|                     | Staff productivity                        |  |
|                     | Staff morale/job satisfaction             |  |

applied to each of the performance measure with univariate analysis and a Bonferroni adjusted alpha level when performing significance testing (Rice, 1995), where overall alpha level set to be 0.1. Table 1 shows the findings from the predictive validity evaluations, where small organization implies less than or equal to 50 IT staff (Sanders, 1998).

In Table 1, the verisimilitude of the predictive validity premise for small organizations was supported with weak evidence. This may be an indicator that the process capability measure is not appropriate for small organizations, or that the capabilities stipulated in ISO/IEC 15504 do not necessarily improve project performance in small organizations.

#### 2.4.3. Evaluation of the exemplar assessment model (Part 5)

The ISO/IEC 15504 document set (ISO/IEC 15504, 2003) contains an exemplar assessment model (Part 5). One motivation for developing this model was to make it easier for organizations to use the Standard immediately. El Emam and Jung (2001) intensively evaluate a variety of issues such as use of the exemplar model; usefulness and ease of use of the exemplar model; meaningfulness of the rating aggregation scheme; usability of the rating scale; usefulness of indicators; understanding of the process and capability dimensions.

Nearly all of the participants in the trials used the exemplar assessment model (Part 5: PDTR version) as a source of indicators (95.5%). Approximately 82% of the respondents have used Part 5 intensively. This implies that Part 5 was intensively used in assessments. In general, lead assessors found Part 5 both useful and easy to use. Furthermore, they were satisfied with the level of detail of the exemplar model (87.2%). However, a minority expressed some concern that they could have produced accurate judgments with less detailed evidence. For the capability dimension almost all of the assessors were confident about their understanding up to level 3 attributes. However, the confidence level dropped for levels 4 and 5 in the perceived consistency and repeatability of their judgments. Subsequently, the Standard was revised according to the recommendation.

#### 2.4.4. Assessment effort

The studies of assessor effort based on Phase 2 of the SPICE Trials show that consensus and interrater agreement are two of the most important factors to reduce assessment cost. The consensus problem and the extent to which assessors reach high interrater agreement is partially related to issues discussed in the exemplar assessment model. Improvement of the exemplar assessment model can expect to decrease the concerns about the high cost of assessments (Jung et al., 2001; SPICE Project, 2003).

In addition, since high interrater agreement can reduce the consolidation effort in assessments of only “Organizational” type processes, it is suggested that WG10 formally or informally introduce a classification of processes into “Organizational” and “Project” type processes. Then, assessors can pay most attention to the ratings of “Organization” type processes in order to reduce the cost of the assessment, and also that future research should focus on improving the reliability of rating this type of process (SPICE Project, 2003).

#### 2.4.5. Comparison with ISO 9001

A comparison study between ISO/IEC 15504 and ISO 9001 was conducted to provide empirical answers to the following questions relating to ISO 9001 and ISO/IEC 15504:

- At what ISO/IEC 15504 capability level would one expect an ISO 9001 certified organization’s processes to be?
- Is there any significant difference in the SPICE capability levels achieved by the processes of ISO 9001 certified organizations and those of non ISO 9001 certified organizations?

The capability level for each of the 29 software processes in ISO 9001 certified OUs is higher than that in non ISO 9001 certified OUs except for two processes with small sample size. The (sample) average capability level for each SPICE process lies between 1 and 2.3 in ISO 9001 certified OUs (Jung and Hunter, 2001).

Since ISO 9001 (2000) uses a process-based approach and includes some changes to the requirements of ISO 9001 (1997), ISO/IEC 15504 and the latest version of ISO 9001 are much more consistent than before. Therefore, it is expected that this consistency between the two Standards would tend to make ISO 9001 certified OUs achieve higher capability levels in terms of ISO/IEC 15504 than presented here (Jung and Hunter, 2001).

### 3. Looking to the future

#### 3.1. Organizational maturity

At the commencement of the SPICE Project, there was no clear distinction between *organizational maturity* and *process capability*. The Study Group report (ISO/IEC JTC1/SC7, 1992) established as a requirement that the Standard should “support output as process profiles which allow views at different levels of detail.”

A significant number of available approaches to process assessment provide results on the basis of an overall scale of organizational maturity, where each point on the scale represents gaining specific capabilities across a set of processes. While initial “maturity models” fell within the domain of software engineering, the concept has been extended to other domains (see, for example, the CMMI (Chrissis et al., 2003)), and most recently to the discipline of project management.

With the evolution of concepts of process capability, it has become accepted (see, for example, CMMI (Chrissis et al., 2003)) that levels of organizational maturity can be considered in terms of “process profiles”, where a specified Maturity Level equates to specified achievement of Capability Levels across a pre-defined set of Processes.

Significant debate has arisen over the appropriate scope and composition of appropriate process profiles for a declared level of “organizational maturity”, as well as the necessary level of rigour and scope for assessment of this attribute. The availability of a defined scale and a mechanism for defining appropriate conditions applying to a maturity scale would help to resolve these concerns. These pressures have led to a new attempt to define a standard covering the issue of organizational maturity. Through the development and empirical investigation of process assessment, a better understanding has been achieved of the relationships between sets of process capabilities (often described at the project level) and organizational maturity; a consequent opportunity has been identified to enhance ISO/IEC 15504 to address the assessment of organizational maturity.

#### 3.2. Scope of models

The initial scope of work for ISO/IEC 15504 was limited to the processes of the software life cycle. In this it reflected the scope of the Software CMM, and also embodied the

then scope of the standardisation community within JTC1/SC7. Over the course of its development, however, the applicability of the approach has been seen to be appropriate for other domains; and this has been accompanied by recognition of the synergy between software and systems engineering, with an accompanying increase in the scope of standardisation efforts.

The major challenge for the Standard is validation that the requirements can be used to establish conformance not only of “purpose-built” assessment models, but also of models constructed by other communities of interest that do not reflect the structures implicit in the Standard. A key challenge in relation to this is the establishment of formal conformance of the CMMI model to the requirements of ISO/IEC 15504. Considerable progress has been made in this regard; a detailed mapping of the Process Areas of CMMI to the processes defined in ISO/IEC 12207 has been developed (Tuffley and Rout, 2004), and a correspondence established between the Generic Practices of CMMI and the Measurement Framework of ISO/IEC 15504-2 (Rout and Tuffley, 2005).

Since the publication of ISO/IEC 15504-2, with the explicit requirements for external Process Reference Models and the development of associated Process Assessment Models, there has been significant interest in the development of domain-specific PRMs and PAMs. Models have been developed that reflect the interests of the Aerospace domain (SPICE for SPACE) (Rodriguez-Dapena, 2003); the Automotive sector (Automotive SPICE) (Automotive SIG, 2005; Fabbrini et al., 2002b) while additional models have been initiated in the field of Medical Device software (McCaffery et al., 2004). Within the standards community, JTC1/SC7 has taken on (through its WG10) the development of an exemplar Process Assessment Model for Systems Engineering.

### 4. Impact

The publication of ISO/IEC 15504 is an appropriate time to review the impact of its development, and the parallel undertaking of the SPICE Project. There is little doubt that the innovative approaches embodied in this development have advanced the state of knowledge and practice, not only in its subject area, but more widely in the fields of standardisation and empirical software engineering.

#### 4.1. Understanding processes

It is important, in communicating information about processes, to have a mechanism that ensures that the entities under discussion are clearly identified; so that, when we are talking about “oranges”, all of the entities are “oranges”, and there are no “mandarins”. This issue became an important concern in the development of the SPICE framework; one of the key contributions of the work has been a resolution that offers a new and fruitful approach to the definition of processes.

A consideration of existing approaches to process definition leads to the following conclusions:

- A process is decomposable into other elements (actions, activities, steps, ...).
- The elements of a process form some type of sequence – are interdependent in their results.
- The elements are assembled in order to achieve a specific end-result.

From these concepts, it can be seen that there is not a unique set of elements for a specific purpose – that is, there is more than one set of actions that will achieve a desired result. Also, it is possible for elements of a process themselves to be decomposable – that is, a process can be comprised of other processes.

The further contribution of the SPICE project has been to make explicit the variety of ways of specifying processes.

- If we want to describe how a process should be performed in order to achieve the desired result, we would specify a set of specific actions (practices) and the relationships between them.
- The descriptions of actions, and of their inter-relationships, can be more or less prescriptive. A more prescriptive approach – giving detailed instructions of what is to be done – will be suited for use within an organization, especially where the context of use is consistent. A less prescriptive approach will be more suited to application across organizations and industry sectors; here the descriptions of actions (practices) try to convey what is to be done, rather than how it is to be performed. This approach is common in standards and descriptive process models, such as ISO/IEC 12207 or CMMI.
- If we wish to specify what a process is meant to achieve, a different approach is needed. One which has proved successful is to specify the process in terms of its purpose, and the outcomes of its implementation. The outcome statements describe tangible evidence of achievement of the process purpose. The approach has been used successfully in ISO/IEC 15504 and has been adopted by JTC1/SC7 as a basis for process specification.

#### 4.2. Developing standards

The process of standards development – particularly at the international level – has been shown to be time consuming, resource intensive, and on occasion to result in products that satisfy few of the needs of the user community.

The development of standards – particularly at the international level – is always a process of compromise, in which political concerns and technical ideals need to be balanced (ISO/IEC JTC1/SC7, 1992). This reality was recognised in the study group report that initiated the SPICE Project, with the recognition of requirements such as:

“The process assessment standard shall:

- be culturally independent;
- be supportive of and consistent with other ISO/JTC1/SC7 standards and projects; and
- be supportive of and consistent with the ISO 9000 series of standards.”

More generally, there is a recognised requirement for all Standards that they should encourage freedom of commerce between nations, and should not be capable of use in erecting trade barriers. Where a Standard – such as the Process Assessment Standard – is intended for use in commercial contractual situations, these political and commercial considerations can substantially impact on the final design of the Standard.

Where feedback from potential users of the developing Standard is not available, concerns over such political issues may outweigh the technical concerns, resulting in Standards in which the technical matters have been so compromised to relate to real or imagined potential commercial implications that their usefulness to the user community is significantly limited.

#### 4.3. Empirical software engineering

The SPICE trials have had influence on empirical software engineering well beyond the development of ISO/IEC 15504. With the exception of a few notable case studies (El Emam and Goldenson, 2000), relatively little quantitative empirical research about process assessment and its role in model-based process improvement existed prior to the SPICE trials. The trials provided opportunities for such research that were not available previously.

As yet, however, no other ISO/IEC Working Group has included empirical trials as a proactive basis for developing a software engineering standard. The SPICE trials remain unique.

Clearly, the development of ISO/IEC 15504 was influenced by many factors. Yet, the SPICE Trials did affect the final outcome; specifically, the results of Phase 1 affected decisions taken in the development of ISO/IEC TR 15504-2, while some of the Phase 2 findings were critical to the review of the Measurement Framework in the process of revision to a full International Standard. The repeatability studies in particular helped make the case for assessment team size requirements. Confidence in the usability and validity of the Measurement Framework was provided by the trials results. In addition, experiential knowledge gained through participating in the trial assessments helped shape the thinking of many Working Group members, and the performance results provided confidence in the suitability of the entire Standard.

The most successful standardisation efforts are those that are based on pre-existing material developed by user groups within the community of interest – the efforts of Working Groups 11 (aligned with the CDIF Group) and



12 (aligned with various Function Point and Software Measurement groups) within ISO/IEC JTC1/SC7 are good examples of this.

Where no widely acknowledged user group exists, the development of standards occurs to a greater or lesser extent, in a closed community of experts whose primary interest is the development of the standard, with its subsequent use being a lesser concern. There is predominance in the software engineering standards community of representation from acquirers – bodies with an interest in seeing that developers use defined processes and adhere to common standards. The primary users of the standards are less well represented, and the extent of their involvement in the development is variable.

There is a need for better inclusion of user feedback into standards development. The existing process for standards development is oriented primarily towards document review as the principal mechanism for feedback, with no ready mechanism for inclusion of experiential input from users. The use of user trials within SPICE has shown the benefits that can be derived from such feedback, and the project should establish a precedent for better user contribution to the standardisation process.

#### 4.4. Adoption

It is one thing to develop a standard, another to evaluate its usage. Data available to evaluate the usage of ISO/IEC 15504 comes from a number of sources. El Emam and Garro (2000) estimated the number of SPICE assessments conducted using ISO/IEC 15504 during Phase 2 of the SPICE Trials; a capture-recapture method estimated around 1250 SPICE based trial assessments up to 2000. A Special Working Group of SC7 (SWG5) collected information on the number of “Google Hits” for each of the Standards developed by the Committee; a search for “ISO/IEC 15504” resulted in 4960 hits in June 2003; as at October 2005, a similar search yielded 49,200 hits. The substantial increase can be attributed partly to the expansion of the Internet itself; however, it also confirms growing interest in and use of the Standard. Scholarly research into the domain of the Standard is also considerable; A search of published literature using Google Scholar yielded 1,040 articles in October 2005.

Because the performance of assessments using ISO/IEC 15504 is not specifically linked to any specification scheme for certification of assessment results (unlike CMMI, for example), the number of assessments performed cannot be determined with any accuracy. Experiences of the authors indicate widespread adoption of the SPICE approach, especially in Europe. Adoption in other countries has also been monitored; there has been extensive use of the Standard in Korea (Jung, 2005a) while there are reports on usage in Australia (Cater-Steel et al., 2006; Rout et al., 1998; Rout et al., 2001).

The development of specific PRMs and PAMs for individual domains is also indicative of the level of adoption of

the Standard; in particular, the release of Automotive SPICE (Automotive SIG, 2005; Fabbrini et al., 2002b) has driven an extensive volume of assessments.

## 5. Conclusions

Our intention in this paper has been to attempt to document the achievements of the SPICE Project and the accompanying standardisation effort in the domain of process assessment. We have demonstrated the significant advances in the understanding of the nature of process capability and its evaluation that have been made possible through these efforts.

The integration of theoretical, definitive and empirical studies in the development of ISO/IEC 15504 is unique in the domain of software and systems engineering. It has resulted in the development of a framework that has application well beyond the initial field of application of software engineering; there have been demonstrated applications in fields outside the information technology area (Coletta et al., 2005).

In addition to the growth in theoretical understanding, there has been a parallel growth in knowledge of the impact of process improvement approaches and the process of process assessment, through the empirical investigations of the trials. The experiences of SPICE have raised expectations throughout the software engineering community in terms of the need for effective standards that have been validated through empirical investigation. Efforts such as SPICE will increasingly be expected in large-scale standardisation efforts.

Finally, the overall strategy has resulted in a new Standard that is widely known and extensively used; this is the key criterion of the success of the effort. It is an effort that continues, and we expect that the patterns of past achievement will continue to be expressed in the future.

## Acknowledgements

The contribution of all of the professionals who have participated in the development and validation of ISO/IEC 15504 is acknowledged; particular individuals need to be credited for their contribution, particularly the initiators of the original Study Group, Alec Dorling (Convenor of WG10 since its inception); Peter Simms (the original Project Editor for the development of the Technical Report) and Harry Barker (original Project Editor for ISO/IEC TR 15504-3). Other members of the original Management Board were Jean Normand Drouin and Mark Paulk. Prominent in the planning and conduct of the SPICE Trials were Fiona McLennan, Bob Smith, Angela Tuffley and Robin Hunter.

## References

- Automotive SIG, 2005. Automotive SPICE process reference model, V 4.2. SPICE users group. <<http://www.automotivespice.com>>.

- Carmines, E., Zeller, R., 1979. *Reliability and Validity Assessment*. Sage University Paper Series on Quantitative Applications in Social Sciences. Sage, Newbury Park.
- Cater-Steel, A., Toleman, M., Rout, T.P., 2006. An evaluation of the RAPID assessment-based process improvement method for small firms. *Information and Software Technology* 48, 323–334.
- Chrissis, M.B., Konrad, M., Schrumm, S., 2003. *CMMI: Guidelines for Process Integration and Product Improvement*. Addison-Wesley, Boston.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1), 37–46.
- Coletta, A., Piazzola, A., Ruffinazzi, D., 2005. An industrial experience in assessing the capability of non-software processes using ISO/IEC 15504. In: *Proceedings of SPICE 2005*, Klagenfurt, Austria, April 2005.
- Cronbach, L., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16 (3), 297–334.
- Dorling, A., 1993. SPICE: software process improvement and capability determination. *Information and Software Technology* 35 (6/7), 404–406.
- El Emam, K., 1998. The internal consistency of the ISO/IEC 15504 software process capability scale. In: *Proceedings of 5th International Symposium on Software Metrics*, pp. 72–81.
- El Emam, K., 1999. Benchmarking kappa: interrater agreement in software process assessments. *Empirical Software Engineering: An International Journal* 4 (2), 113–133.
- El Emam, K., Birk, A., 2000a. Validating the ISO/IEC measure of software development process capability. *Journal of Systems and Software* 51 (2), 119–149.
- El Emam, K., Birk, A., 2000b. Validating the ISO/IEC measure of software requirement analysis process capability. *IEEE Transactions on Software Engineering* 26 (6), 541–566.
- El Emam, K., Briand, L., Smith, R., 1996. Assessor agreement in rating SPICE processes. *Software Process: Improvement and Practice* 2 (4), 291–306.
- El Emam, K., Garro, I., 2000. Estimating the extent of standards use: the case of ISO/IEC 15504. *Journal of Systems and Software* 53 (2), 137–143.
- El Emam, K., Goldenson, D., 1995. SPICE: an empiricist's perspective. *Proceedings of Second IEEE International Software Engineering Standards Symposium*. IEEE Computer Society Press, Los Alamitos, California, pp. 84–97.
- El Emam, K., Goldenson, D., 2000. An empirical review of software process assessments. *Advances in Computers* 53, 319–423.
- El Emam, K., Jung, H.-W., 2001. An evaluation of the ISO/IEC 15504 assessment model. *Journal of Systems and Software* 59 (1), 23–41.
- El Emam, K., Marshall, P., 1998. Interrater agreement in assessment ratings. In: El Emam, K., Drouin, J.-N., Melo, W. (Eds.), *SPICE: The Theory and Practice of Software Process Improvement and Capability Determination*. Sage, Los Alamitos, California, pp. 357–362.
- Fabbrini, F., Fusani, M., Lami, G., 2002a. Assessing software processes: the rating dilemma between measurement and human judgment. In: *Proceedings of SIMPROS 2002*, Recife, Brazil, September, pp. 10–13.
- Fabbrini, F., Fusani, M., Lami, G., Sivera, E., 2002b. Software process assessment as a mean to improve the acquisition process of an automotive manufacturer. *Software Process Improvement CMM&SPICE in Practice*. Verlag UNI-DRUCK, Munchen, Germany, pp. 142–154.
- Fenton, N., Pfeleger, S.L., 1996. *Software Metrics: A Rigorous and Practical Approach*. International Thompson Computer Press, London.
- Hunter, R., 1998. SPICE trials assessment profile. *IEEE Software Process Newsletter* 12, 12–18.
- Hunter, R., Jung, H.-W., 2003. Annotated bibliography for phase 2 of the SPICE trials. SPICE Project. <<http://www.cis.strath.ac.uk/research/papers/SPICE/part3.pdf>>.
- ISO/IEC 15504, 2003. *Information Technology, Process Assessment, Parts 1–5*. ISO/IEC TR 15504, 1998. *Information Technology, Software Process Assessment, Parts 1–9*.
- ISO/IEC JTC1/SC7, 1992. *The need and requirements for a software process assessment standard*, Study Report, Issue 2.0, JTC1/SC7 N944R, 11 June 1992.
- Jung, H.-W., 2003a. Evaluating interrater agreement in SPICE-based assessments. *Computer Standards and Interfaces* 25 (5), 477–499.
- Jung, H.-W., 2003b. Evaluating the internal consistency of ISO/IEC TR 15504 process capability measures. *Software Process Improvement and Practice* 8 (1), 5–26.
- Jung, H.-W., 2005a. Current state of software process improvement and assessment in Korea. In: *Proceedings of SPES 2005 Symposium*, Japan, pp. 60–73.
- Jung, H.-W., 2005b. Evaluating the ordering of the SPICE capability levels: an empirical study. *Information and Software Technology* 47 (3), 142–149.
- Jung, H.-W., Hunter, R., 2001. The relationship between ISO/IEC 15504 process capability levels, ISO 9001 certification and organization size: an empirical study. *Journal of Systems and Software* 59 (1), 43–55.
- Jung, H.-W., Hunter, R., Goldenson, D., El Emam, K., 2001. Findings from phase 2 of the SPICE trials. *Software Process Improvement and Practice* 6, 205–242.
- MacLennan, F., Ostrolenk, G., Tobin, M., 1998. Introduction to the SPICE trials. In: El Emam, K., Drouin, J.-N., Melo, W. (Eds.), *SPICE: The Theory and Practice of Software Process Improvement and Capability Determination*. IEEE CS Press, Los Alamitos, California, pp. 269–286.
- McCaffery, F., Donnelly, P., Dorling, A., Wilkie, F.G., 2004. A software process development, assessment and improvement framework for the medical device industry. In: *Proceedings of SPICE 2004*, Lisbon, Portugal, pp. 28–29.
- Paulk, M.C., Curtis, B., Chrissis, M.B., Weber, C.V., 1993. *Capability maturity model for software*, Version 1.1. Report CMU/SEI-93-TR-24, Software Engineering Institute, Pittsburgh.
- Radice, R.A., Harding, J.T., Munnis, P.E., Phillips, R.W., 1985. A programming process study. *IBM Systems Journal* 24 (2), 91–101.
- Rice, J., 1995. *Mathematical Statistics and Data Analysis*. Duxbury Press.
- Rodriguez-Dapena, P., 2003. ECSS-Q-80-02, A standard for software process assessment and improvement for the European space domain. In: *Proceedings of 3rd International SPICE Conference on Process Assessment and Improvement*, Nordwijk, Netherlands, pp. 7–14.
- Rout, T.P., 2003. ISO/IEC 15504 – evolution to an international standard. *Software Process Improvement and Practice* 8, 27–40.
- Rout, T.P., Tuffley, A., 2005. The ISO 15504 measurement framework for process capability and CMMI. In: *Proceedings of SPICE 2005*, Klagenfurt, Austria.
- Rout, T.P., Tuffley, A., Hodgen, B., 1998. Process capability in the Australian software industry – results from the SPICE trials. In: *Proceedings of Australian Software Engineering Conference – ASWEC'98*, Adelaide, Australia, pp. 12–19.
- Rout, T.P., Tuffley, A., Cahill, B., Hodgen, B., 2001. The rapid assessment of software process capability. In: Hunter, R.B., Thayer, R. (Eds.), *Software Process Improvement*. IEEE Computer Society Press, Los Alamitos, California.
- Sanders, M., (Ed.) 1998. *The SPIRE Handbook: Better, Faster, Cheaper Software Development in Small Organizations*. The SPIRE Project Team: ESSI Project 23873.
- SEI, 2001. SCAMPI V1.1, Standard CMMI appraisal method for process improvement: method definition document, Version 1.1 (CMU/SEI-2001-HB-001, ADA3399204). Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA.
- SPICE Project, 1995. *Phase 1 Trials Report*, Version 1.00.
- SPICE Project, 2003. *Final Report of Phase 2 of the SPICE Trials: Volume 1, Executive summary and detailed reports of some of the studies*, Version 1.01.
- Tuffley, A., Rout, T.P., 2004. CMMI conformance to ISO/IEC 15504: Mapping to ISO/IEC 12207. SEPG Australia Conference, Adelaide Australia.

**Terence P. Rout** is an associate professor in the School of Computing and Information Technology at Griffith University, Queensland, and is associated with the Software Quality Institute. He is the overall project editor for ISO/IEC 15504 – Software Process Assessment and has contributed numerous publications in this field. He has broad experience as a leader of ISO/IEC 15504-based assessments, and has advised companies on the implementation of effective assessment-based improvement programs. He has led the Australian effort in the SPICE Project since its inception, developing an extensive network of industry participants in the evaluation of approaches to process assessment. He is recognised as one of the leading international authorities on SPICE (Software Process Improvement and Capability dEtermination), in which area he has both published extensively and been an invited speaker on frequent occasions.

**Khaled El Emam** is Chief Scientist at TrialStat Corporation, a Senior Scientist at the Children’s Hospital of Eastern Ontario Research Institute, where he is leading the eHealth research program. In addition, he is an Associate Professor at the University of Ottawa, Faculty of Medicine, and a Canada Research Chair in Electronic Health Information. Previously He was a senior research officer at the National Research Council of Canada, where he was the technical lead of the Software Quality Laboratory, and prior to that, he was head of the Quantitative Methods Group at the Fraunhofer Institute for Experimental Software Engineering in Kaiserslautern, Germany. In 2003 and 2004, He was ranked as the top systems and software engineering scholar worldwide by the *Journal of Systems and Software* based on his research on measurement and quality evaluation and improvement. Currently, he is a visiting professor at the Center for Global eHealth Innovation at the University of Toronto (University Health Network) and at the School of Business at Korea University in Seoul. He holds a Ph.D. from the Department of Electrical and Electronics Engineering, King’s College, at the University of London (UK)

**Mario Fusani** has been with the National Research Council of Italy since 1973. First he worked in the field of distributed operating systems. Since the 80’s, he has been involved in the investigation of quality issues of software products and processes. He visited, both for software project evaluation and for process assessment, many public and private organizations in Europe, USA and Asia. He is member of the International Process Research Consortium, launched by the SEI/CMU. He is responsible of a joint research program with the SEI for the automatic analysis of requirements expressed in natural language.

**Dennis R. Goldenson** is a senior member of the technical staff at the Software Engineering Institute (SEI). A principal author of the Measurement and Analysis process area for Capability Maturity Model Integrated (CMMI) models, He is the technical lead for the SEI’s empirical investigations into the impact of CMMI-based process improvement. He previously served as co-lead of test and evaluation for CMMI. A charter member of the SPICE trials team in support of ISO/IEC 15504, he served for a time as international trials coordinator. Goldenson came to the SEI in 1990 after teaching at Carnegie Mellon University since 1982.

**Ho-Won Jung** is a professor in the Department of Business Administration at Korea University. He is also the international SPICE research coordinator for the emerging ISO/IEC 15504 (Process Assessment) International Standard and an authorized instructor for introductory courses in the SEI Capability Maturity Model Integration. He is a charter research member of the International Process Research Consortium and is on the *Software Quality Journal* editorial board. He received his BS in Industrial Engineering from Korea University, his MS in the same field from the Korean Advanced Institute of Science and Technology (KAIST), and his Ph.D. in Management Information Systems from the University of Arizona.