ON UNDERSTANDING EVALUATION OF TOOL SUPPORT FOR IS DEVELOPMENT

Björn Lundell
School of Humanities and Informatics
University of Skövde
P.O. Box 408
SE-541 28 Skövde
Sweden
bjorn@ida.his.se

Brian Lings
School of Engineering, Computer Science and Mathematics
University of Exeter
Exeter EX4 4PT
England
B.J.Lings@dcs.ex.ac.uk

ABSTRACT

Evaluation of IS tools has received considerable attention in the literature, yet no consensus is apparent on characterising the important dimensions of an evaluation process. From an analysis of the literature, we identify a weakness in extant methods to support evaluation activities, and posit a framework for IS tool evaluation. We conclude that there is a need for enriched method support to deal with the complex socio-technical issues involved in the summative assessment of IS tools.

Keywords: Tool evaluation, Evaluation dimensions, Socio-technical evaluation, IS development tool

INTRODUCTION

Several overlapping interpretations of the concept of evaluation have been presented in the literature (Serafeimidis and Smithson, 1996, p. 205). With acknowledgement to Remenyi *et al.* (1997), Forss *et al.* (1994) and Ramage (1997), in this paper we consider evaluation as

> involved learning through a series of activities related to understanding, feedback and assessment.

Symons and Walsham (1988) consider evaluation not in terms of tools and techniques, but as a *process to be understood*, emphasising that such understanding "should not be confused with an understanding of the situation under study" (p. 123). From this interpretive perspective one can argue that such understanding can only come with an awareness of: *why* evaluation is undertaken; the *nature* of the process in which it is undertaken; and the impact of and impact on the *context* in which evaluation is undertaken.

The aim of this paper is to posit a framework which encompasses the key elements of evaluation, developed from a review of the evaluation literature, and to use this framework to draw attention to perceived weaknesses in current method support for evaluation – and in particular setting up an evaluation activity.

Whilst initially based on the Serafeimidis and Smithson (1996) framework for considering the elements of evaluation, the analysis behind our framework draws broadly from reported research in a number of fields in order to explore the variety of issues that might affect an evaluation activity. Three key dimensions are identified in the framework: the *stakeholder dimension*, the *contextual dimension* and the *activity dimension*. We first present the new framework, and then present the analysis behind it focussing on each dimension in turn. We do this not only to justify the elements within the framework, but to act as a rich description of the elements identified. Finally, we consider implications for practical evaluations and for research into method support for such evaluation activities.

DIMENSIONS OF EVALUATION

Serafeimidis and Smithson (1996) introduce what they consider to be the primary elements of Information Systems evaluation (see Figure 1), presenting a "multiple perspective approach" (p. 206). They identify context, process and content as central.
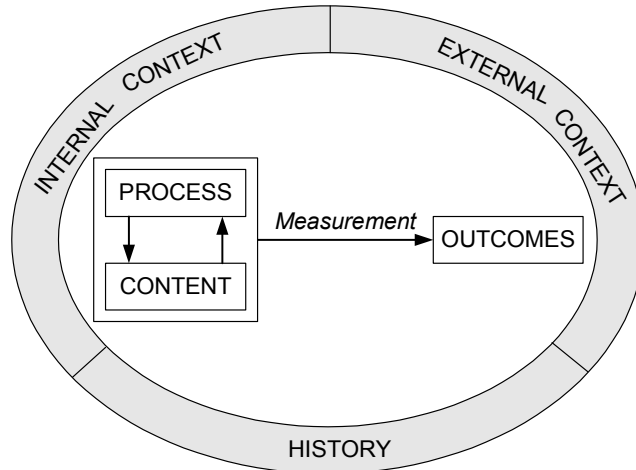
FIGURE 1: ELEMENTS OF INFORMATION SYSTEMS EVALUATION (SERAFEIMIDIS AND SMITHSON, 1996)

Whilst agreeing with their general formulation, our analysis and experience suggests the utility of making further elements of evaluation explicit (see Figure 2). In particular, the centrality of the stakeholder dimension is highlighted within the context of an evaluation.
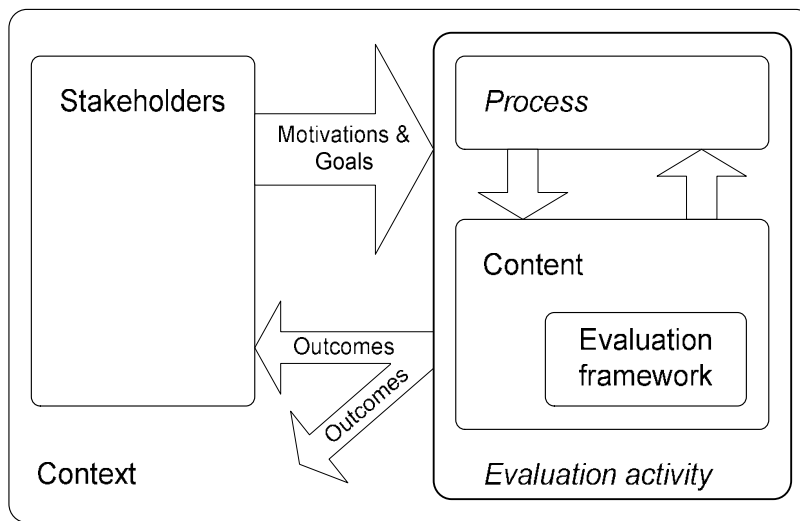
FIGURE 2: ELEMENTS OF IS DEVELOPMENT TOOL EVALUATION

There is a resulting dynamic between an evaluation activity and the context in which it takes place. In particular, stakeholder motivations and goals are explicitly represented as strongly affecting an evaluation activity. Further, outcomes are not represented as a distinct element, but affect as well as effect is included. This cycle accounts for the dynamic of an evaluation activity in context. The final

addition is the element of an evaluation framework, which is seen by many as fundamental to any evaluation activity. For example, according to Kitchenham (1996), writing in a series of papers on evaluating software engineering methods and tools, this is the position taken in DESMET, a project concerned with evaluation "within a particular organisation" (p. 11). In the fifth paper in the series, Kitchenham and Jones (1997) claim that "A common framework is necessary to make any sort of comparative evaluation of methods and tools." (p. 14) (for further examples see Bailer *et al.*, 1993; IEEE, 1992; ISO, 1999; Powell *et al.*, 1997; Smyth, 1998; Lundell and Lings, 2003).

ANALYSIS

While a majority of sources for the analysis could be classed as "general" sources from the field of information systems (interpreted broadly), sources from other fields have also been considered. These include: *IS management*, *IS assessment*, *evaluation research*, *educational evaluation* and *Software Engineering* (among others). Such a broad basis for the study seems appropriate, since the issue of evaluation is important in all these fields and each brings its own insights. The development of the new framework was our response to understanding a seeming lack of consensus in the literature.

Focus on the stakeholder dimension

<u>Different stakeholders for an evaluation</u>

One view of the term stakeholder (for example adopted by Rossi and Freeman (1993)) is that it denotes a *human actor* who *is affected by* or *can affect* a system (or an activity), and it is this view that will be adopted here.
Farbey *et al.* (1999) claim, based on findings from field studies, that "[t]here are few attempts to analyse the stakeholder role" (p. 202). However, there is a growing interest in the stakeholder dimension. For example, Jones and Hughes (2001) stress the importance of stakeholders in arguing that the "views, beliefs and assumptions of stakeholders must be exposed and considered with the IS evaluation process" (p. 200), and Mcauley *et al.* (2002) "extends Farbey *et al.*'s (1999) concern that stakeholder analysis should be included as an important evaluation theme" (p. 242) based on a study concerning an outsourcing decision.
From evaluation studies in the field of educational programmes, Brandon (1998) concludes that "stakeholder participation can help enhance the likelihood of producing valid evaluation findings." (p. 333) There are many reports in the literature concerning *why* it is important to seriously consider the impact of stakeholders on evaluations. Etzerodt and Madsen (1988) highlight the need to view problems and issues from different perspectives, in order to achieve an in-depth understanding, a view supported by Ramage (1997). Stakeholders' individual experiences, perceptions, values and interests can all have an impact on an evaluation process. This highlights the importance of carefully considering which stakeholders to directly include in the activity.
When selecting stakeholders it is important to acknowledge the importance of a broad and representative input to the evaluation process. For example, Fetterman (1997) observes that the "consumer's voice" is often excluded from the activity. Consequently, if important 'input' to the process is left out, there is a potential risk that an evaluation activity might consider aspects that are not considered relevant by the community for which it is undertaken. It is also important to consider the closeness of a role to the "usage context" for any tool being evaluated (Lundell and Lings, 2003). This is not only an issue of the richness of the inputs to an evaluation activity, but affects the politics of evaluation.
It has been claimed that "[e]valuation is, to a large extent, a political activity depending on the particular interests of the stakeholders, their personal values, and judgements." (Avgerou, 1995, p. 428) For example, with respect to tool support for IS development, Hughes and Wood-Harper (2000) note in reporting from a case study that a "common theme in respect of tool and methodology evaluation is politics, many suggested 'power dominates' evaluation in practice" (p. 401). Walsham (1993), influenced by the work of Guba and Lincoln (1989), similarly elaborates on the need to

consider multiple interests among different stakeholders in an IS evaluation. Transparency of process helps to maintain motivation and encourage a sense of ownership. An important part of transparency is direct feedback, and Rehbinder *et al.* (2001) claim that an evolving evaluation framework can act as an effective feedback mechanism.

Brandon (1998) found, from studies in the field of educational programmes, that stakeholder participation improves acceptance among stakeholders that are not directly involved in the evaluation process because they know that their colleagues have been involved.

There is an important distinguished role of the *evaluator* identified in the literature, which has been characterised as a "key role" in IS evaluation (Serafeimidis and Smithson, 2003, p. 270). The evaluator may come from the defined organisational context or from outside. According to Avgerou (1995):

"The evaluator is an organizer and facilitator of the evaluation process, rather than an objective assessor, capable to calculate or to prove the merits and the pitfalls of an information system." (p. 436)

The ability of an evaluator to acknowledge different viewpoints and perspectives becomes more important when multiple stakeholders are involved (Symons, 1990, p. 195). According to Symons (1990), an "evaluation is never value-free" (p. 195), and consequently certain aspects can be seen "very differently" (p. 195) among different stakeholders. Therefore, it is important that an evaluator has the ability to understand such alternative viewpoints, since a failure to understand can "lead to unanticipated resistance and hence operational problems" (Symons, 1990, p. 195). However the role of evaluator is defined, Scriven (1996) acknowledges that it will be very demanding.

Different motivations and goals for addressing an evaluation

The dynamics of a usage context are strongly influenced by the motivations of stakeholders for addressing a particular evaluation. Therefore, the importance of clarifying its *purpose* is considered, by Walsham (1993) and others, to be a "key element of the evaluation" (p. 173) irrespective of whether it is undertaken at an early proposal stage or at a post-implementation stage. Budgen and Thomson (2003, p. 58) present a variety of different motivations, or "agendas", for conducting tool evaluation, their specific concern being CASE-tools.

In the field of evaluation research, Chelimsky (1997, p. 10) distinguishes three different purposes for an evaluation, namely: *accountability* (e.g. to measure results or efficiency), *development* (e.g. for helping managers to improve "institutional performance"), and *knowledge* (e.g. to achieve an in-depth understanding with respect to some specific task or area). While acknowledging that these three purposes are not mutually exclusive and exhaustive, it is argued that this distinction can be helpful when analysing suitable approaches for an evaluation activity with respect to (potentially) different expectations of outcome.

In terms of *accountability*, as organisations spend large proportions of their budgets on information systems and technology they consider it important to assess and measure organisational impact. This is true even where organisational benefits have become more difficult to estimate in financial terms so "measurement involves subjective estimates" (Avgerou, 1995, p. 428) provided by human assessments.

With reference to *development*, Rossi and Freeman (1993), from the field of evaluation research, claim that:

"Evaluations are undertaken to influence the actions and activities of individuals and groups who have, or are presumed to have, an opportunity to tailor their actions on the basis of the results of the evaluation effort." (p. 46)

From this perspective, it can be argued that evaluations are important in the change process, during which an organisation considers IS development tools and information systems adoption (Symons, 1991, p. 211).

With reference to *knowledge*, Almstrum *et al.* (1996), in an educational context, adopt a learning perspective on the purpose for an evaluation, asserting that "[e]valuation is an educational process, not an end in itself" (p. 201). Related to this, Patton (1996), in the field of evaluation research, uses the term *developmental evaluation* when characterising long-term, ongoing evaluation processes that are aimed at facilitating an *in-depth* understanding of need. In such "long-term improvement-oriented evaluation" (p. 135), the "development is an end in itself" (p. 135), and therefore the *process itself* is the goal for the evaluation activity. The primary concern for such formative evaluation is, according to Farbey *et al.* (1999), to "illuminate and learn", contrasting with summative evaluation, which is for "accounting and control." (p. 200)

Whatever the underlying motives, it is important to be *explicit* about them and reasons for an evaluation being undertaken (Walsham, 1993, p. 182).

In the context of software evaluation, Brown and Wallnau (1996) argue that it is important to have "well-defined goals before starting the evaluation" (p. 40).

As the concept of evaluation itself is so broad, it is not surprising that individual stakeholders can have very different goals, some of which may well be implicit. Brown and Wallnau's (1996) statement refers to agreed, explicit, and shared goals only, and these will collectively be seen as organisational goals.

When considering sources that have an emphasis on goals related to *learning* and *understanding* in the light of IS development tool evaluation, we observe that they typically concern evaluation in a usage situation. In other words, the phenomenon to be evaluated is *dynamic* since human actors are involved.

In the field of information systems, tools are often evaluated with the goal of facilitating learning among stakeholders within organisations. For example, Serafeimidis and Psoinos (1995) consider evaluation to be an important feedback mechanism for organisations that facilitates learning, and characterise this activity as a complex process that is far from rational. Instead, they view it as an inherently social activity, influenced by the human bias of the evaluator. Ramage (1997) identifies two benefits from taking this view. Firstly, it allows involved stakeholders to derive different benefits from the evaluation outcome. Secondly, it acknowledges that *learning* is stimulated simply due to the presence of an evaluator.

A fundamentally different perspective obtains when evaluations emphasise the *assessment* of functionality against stated requirements for a specific IS development tool. In other words, the phenomenon to be evaluated is to be considered as a *static*, technical artefact. This perspective is adopted in the ISO standard for software product evaluation (ISO 14598). According to an overview (ISO, 1999b), the goal is "to decide whether relevant quality characteristics meet the requirements of the system." (p. iv) A principal objective of an evaluation process, according to part 5 of this ISO standard (ISO, 1998, p. 4), is to promote four process characteristics: *repeatability* ("repeated evaluation of the same product to the same evaluation specification by the same evaluator should produce results that can be accepted as being identical"), *reproducibility* ("evaluation of the same product to the same evaluation specification by a different evaluator should produce results that can be accepted as being identical"), *impartiality* ("the evaluation should not be biased toward any particular result"), and *objectivity* ("the evaluation results should be factual, i.e. not coloured by the feelings or the opinions of the evaluator"). We interpret this standard as an example of an approach to evaluation that corresponds to what Jackson and Sulaksono (1998), in the context of information systems evaluation, refer to as evaluations being undertaken in "the efficiency zone" (p. 42). They claim that the overall aim for such evaluation activities is "the achievement of more precise measurement" (p. 42) of the evaluated systems' functions.

A similar example of this view, commonly found in the field of software engineering, in which the *assessment* and *judgement* aspects of an evaluation are more central, is exemplified by Carney and Wallnau (1998) with respect to commercial off-the-shelf (COTS) software product evaluation. They argue that, by definition:

"to evaluate is to assign value to something, or to judge something. We can improve this definition by considering why we evaluate: essentially, we evaluate in order to decide, that is, to choose between different possibilities." (p. 852)

In terms of the formative/summative debate (Patton, 1996), we would categorise this view of evaluation as *summative*, since its role typically is to constitute a basis for a "Go/No-Go" (Rossi and Freeman, 1993, p. 47) decision.

For the evaluation of IS development tools *in use*, an otherwise static phenomenon (a tool) is to be evaluated in a dynamic (usage) context. The dynamic nature of this context (IS development organisations) implies that assessments may be formative as well as summative, and that the ISO objectives must be treated with caution. Adoption of a product is typically expensive and will precipitate a change of context, so any evaluation must take a long-term perspective.

Focus on the contextual dimension

The *context* of an evaluation activity refers to those aspects that might influence and affect the activity, and thereby the outcome of the activity. Given that one is considering IS development tools in use, the context of an evaluation activity is at least partly defined by its stakeholders. Being human actors, this context will therefore change as the activity progresses. However, Ramage (1997) observes that software evaluation "often tends to go on in isolation from the 'real world' of work, in labs and the like." (p. 775) From our own review of ISO 14102 (Lundell and Lings, 2002), it is apparent that no consideration is given to the stakeholder dimension during an evaluation. We characterise such evaluations as 'context-free', in that no recognition is given to the feedback into the context. However, this is perhaps not surprising as the purpose of evaluation implied in the standard is summative, namely a go/no-go decision.

Even when considering evaluation of IS development tools for software selection in context, there is a lack of consensus on what constitutes an appropriate context for an evaluation activity. In fact, Silverman (1998) claims that "'Context' itself is by no means a straightforward concept" (p. 6) as there may be a mismatch between an evaluator's view and that of the (other) stakeholders. Here we concentrate on the context for an evaluation, but in so doing elaborate further on the role of an evaluator.

Symons (1991) argues against what she sees as a perception common to much of the evaluation literature, namely that evaluation is a largely deterministic process, since such systems are *part of* and hence "inseparable from the organizational context" (p. 206). This view is based on the perception of an information system not *only* as a technical system but also as a dynamic social system. More recently, Jones and Hughes (2001) have noted that evaluation methods still tend to be "based on mechanistic and functional principles that are flawed in practice" (p. 189). This is clearly evident in the ISO standard, in which the selection process involves the application of a "selection algorithm" (ISO, 1995, p. 24) in order to decide whether to "acquire a tool or set of tools." Based on their experience, Jones and Hughes stress the importance of "an explicit concern for the social context" (p. 200).

One can distinguish between an *internal* and an *external* context for an evaluation (Serafeimidis and Smithson, 1996) - also referred to as *inner* and *outer* context (e.g. Symons, 1991). An *internal context* consists of factors and 'circumstances' within an organisation that have an impact on the evaluation activity. Examples include: stakeholder perspectives; stakeholder experiences and organisational history; informal communications and information flows within an organisation; organisational structure; political and organisational culture within an organisation; and corporate policies. It can also be important to consider which parts of an internal context represent the usage context (Lundell and Lings, 2003). Typically, stakeholders in the usage context perceive that they may be strongly affected by the outcomes of an evaluation, and their motivations and goals are strongly affected by this belief. By contrast, stakeholders in the internal context but outside the usage context have higher-level (for example policy level) goals, which act more as constraints on evaluation activities.

An *external context* includes that part which is typically beyond the control of the organisation in which an evaluation is undertaken. Serafeimidis and Smithson (1996, p. 207) claim that "the organization and its members need to respond to and accommodate" external factors to an evaluation, such as "national and local government policy, level of government support, markets and

market demands, competition, supplier availability and expertise, and other environmental pressures."

For the purpose of IS development tool evaluations, ISO-standards (and other 'de jure' standards, being defined external to an organisation) represent further examples of *external contextual factors* for an evaluation, since simply their existence (whether being considered or not) might, to a greater or lesser extent, affect an IS development organisation. Additionally, as part of a (high level) corporate policy within an IS development organisation, there can also exist various corporate standards that individual projects must adhere to, for example, an organisational 'standard' for a modelling method and/or notation.

Finally, it is clear that the evaluator role is closely related to contextualisation. Fetterman (1997) observes from the field of social program evaluations, based on the work of Greene (1997), that an evaluator is always on somebody's side, and that the choice of side being made by an evaluator is strongly related to whose question is being addressed in the evaluation. Consequently, the evaluator's choice of relevant factors and criteria to be considered in an evaluation activity is always related to the context in which the question to be addressed in the evaluation is expressed (Fetterman, 1997).

The Activity dimension

There is much general discussion in the literature of activities for tool evaluation, (e.g. Budgen and Thomson, 2003; Kitchenham *et al.*, 1997). Few sources give detailed descriptions about how to conduct a tool evaluation activity, but a significant exception is the ISO 14102 standard for evaluation, which identifies a rich set of activities. Figure 3 gives an overview of a full evaluation process, which is divided into four main phases. Most papers cover only a sub-set of these phases, and do not attempt to cover the whole process. However, the ISO standard can be criticised for largely ignoring the other dimension of evaluation (Lundell and Lings, 2002), and in particular is weak with respect to acknowledging the stakeholder dimension. Output is also limited to a simple, quantitative decision.
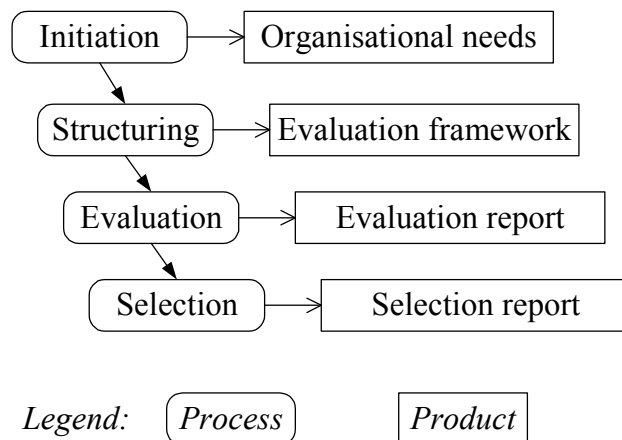
FIGURE 3: A FULL EVALUATION PROCESS FOR IS DEVELOPMENT TOOLS, BASED ON THE ISO 14102 STANDARD

Evaluation Activity Process: Initiation and Structuring

The survey instrument has been extensively used in a variety of contexts for assessing different stakeholder views (e.g. developer, user, researcher) of tools, and limitations associated with their use for tool evaluation have been discussed (e.g. Coupe, 1994; Coupe and Onodu, 1997). Although exceptions from large international organisations have been reported (e.g. Maccari and Riva, 2000,

2002), such surveys are often undertaken in the broad context of a general user (or developer) community. In other words, the interviewees in such assessments represent a context external to that in which a specific tool is used (or intended for use). For example, a variety of surveys have been undertaken to assess stakeholder perceptions (and/or experiences) of tools (e.g. Lending and Chervany, 1998; Maansaari and Iivari, 1999; Post *et al.*, 1998) in order to establish a general summative outcome from an evaluation activity. The *content* for such evaluation is typically based on a pre-defined set of issues/questions to be addressed. These may have been designed with the specific outcome in mind of making a recommendation for action, which can range from important lessons learnt, through guidelines for adoption practice to a definitive recommendation for adoption. Alternatively, they may target the outcome of making a general assessment of the qualities of competing tools (or specific features of competing tools).

In contrast, when an evaluation aims to establish an in-depth understanding of a specific organisation or usage context, qualitative approaches are often preferred. According to Irani and Love (2001, p. 185) there "is no doubt that interpretative approaches to IT/IS evaluation are needed so that a manager can gain an insight into how IT/IS investments will influence employee behaviour and performance." Sometimes, qualitative methods, such as Grounded Theory, have been used in evaluation for establishing developers' or users' experiences within specific organisations after adoption of specific tools. For example, Cronholm (1995) used Grounded theory to evaluate specific organisations after the adoption and use of specific CASE-tool products, and given that such tools are among the most complex products marketed today (Post *et al.*, 1999, p. 215) one could argue that statements made about CASE-tools generalise to all IS products.

There is an (implicit or explicit) underlying assumption in evaluation methods of an *evaluation framework*. Such a framework comprises a representation of potentially relevant tool features and how they interrelate, sometimes referred to as a checklist. Some published frameworks are aimed at general tool classes (e.g. Ovum, 1993); others are tailored for use in pre-usage evaluation of (perhaps limited aspects of) a specific type of tool. For example, there are those designed for the analysis of: hypertext functionality in CASE-tools (e.g. Kaipala, 1997); data modelling tools (e.g. Moriarty, 1998); collaborative work (e.g. Vessey and Sravanapudi, 1995); conformance to a specific modelling notation (e.g. UML - Juric and Kuljis, 1999); conformance to method rules in structured analysis (e.g. Jankowski, 1997); usability of CASE-tools (e.g. Cronholm, 1998).

A number of problems have been acknowledged with published frameworks. Firstly, published frameworks can differ significantly. For example, both Henderson and Cooprider (1990) and Ovum (1993) have developed rather extensive frameworks for CASE-tools, but they differ in scope, structure and detail. Secondly, frameworks must be continually updated to reflect technology changes. This problem is recognised explicitly in some frameworks (e.g. ISO, 1995, p. 25). Thirdly, even the developers of frameworks would accept that they cannot be used without modification in a specific usage situation, not least because frameworks may be interpreted differently in different contexts.

Most sources emphasise that the frameworks presented are to be taken as comprising potentially relevant aspects, to be considered in an evaluation if found appropriate for the evaluation task at hand. The method by which aspects are to be classified as appropriate is usually not elaborated.

More fundamentally, it has been shown to be problematic to arrive at a broadly shared, in-depth understanding of even a single criterion that is defined *a priori*. This is clearly shown in an empirical and exploratory study (van Reeken and Trienekens, 1992) concerning method and CASE-tool usage, undertaken in sixteen of the largest organisations in the Netherlands. Each organisation was found to use the same terms in fundamentally different ways. This gives a strong motivation for developing an evaluation framework within a particular organisational setting, and even for a specific evaluation activity.

Evaluation Activity Process: Evaluation and Selection

Once an evaluation framework has been established an evaluation activity can proceed with a detailed analysis of tool features. Various strategies have been proposed in the literature for this

stage, and several authors stress the importance of enforcing a rigid process for the actual evaluation and selection phases. For example, Powell *et al.* (1997) claim that "evaluators should provide evidence that all relevant issues from [the content of the evaluation framework] have been addressed." (p. 173) It should be emphasised that a selection decision can result in a decision not to adopt a tool.

Most descriptions of tool evaluation approaches are rather vague and provide limited detail on how to undertake an evaluation and selection process, and in this sense Powell *et al.* (1997) is not untypical. However, as noted earlier, ISO 14102 (ISO, 1995) provides more detailed guidance, most notably on how to undertake the later phases of an evaluation activity. During the evaluation process, the tool "is evaluated in comparison with" the content of the evaluation framework that has been established. It is apparent that view of an evaluation is seen as a process "of measurement, rating and assessment" (p. 19). The measurement can be undertaken in a variety of different ways, including examination of vendor-supported documentation, examination of code and the tool, and interviews with actual users of the tool. It is apparent that rating and assessment is mainly quantitative, and that measurements are undertaken at several levels of granularity using defined rating scales. Specifically, an evaluation framework consists of the following levels: *a set of characteristics* (e.g. "Functionality - characteristics related to life-cycle processes" (p. 25)); *characteristics* (e.g. "Development process" (p. 26)); *subcharacteristic* (e.g. "Modeling" (p. 27)); and *atomic subcharacteristic* (e.g. "Diagram Development" (e.g. p. 27)). Actual measurements are then aggregated "up to the characteristics level" (p. 21), and used as the basis for an evaluation report. In such, all evaluation results are provided and for each subcharacteristic, "the metric value measured" (p. 21) is given in terms of a rating level for that metric.

For the actual selection process, ISO (1995) use a "selection algorithm" (p. 23) and several different types are suggested in an appendix to the standard (ISO, 1995), including those suggested for an organisation which addresses "evaluation data which are fuzzy or sparse and is having difficulties in amalgamating the viewpoints of multiple evaluators and/or selectors" (p. 53). Among those mentioned are (ISO, 1995, p. 53): Borda's algorithm (which is a "sum of the ranks" algorithm), Condorcet's algorithm (a "pairwise comparison" algorithm), Dodgson's algorithm (a "preference measurement" algorithm), Fishburn's algorithm (a "preference ordering" algorithm), and Analytic Hierarchy Process (a "structured" algorithm).

After completing the evaluation, a final validation of the selection decision (Chapter 8.4 in ISO 14102 (ISO, 1995)) must be made against the organisation's goals.

> "A check should be made to ensure that if the recommendation is accepted, the high level goals (or a sufficient number of them) will be met." (ISO 1995, p. 24)

Clearly, this validation is far from formal. The standard gives no advice on how to perform this task, however an associated technical report (ISO, 1999) on how to use ISO 14102 suggests using pilot studies in the later stages of evaluation:

> "A pilot project process should be conducted to provide a realistic trial for the CASE tools in their intended environment. While the tool was exercised during evaluation and selection, that process does not require a realistic use of the tool. Evaluation and selection identifies the tool among the candidates with the most potential for the organization. The pilot project aims to be sure it can really perform for the organization in a real application." (ISO, 1999, p. 9)

Powell *et al.* (1997) also suggest the use of pilot studies, but in their case the suggestion is for their 'early' use in the overall activity. They stress that:

> "Significant tool selection decisions should not be made without recourse to some form of pilot study - either shadowing part of a project or applying it to a low risk part of a live project." (p. 171)

It is our belief that both views imply the importance of qualitative methods if the desirable goals of evaluation are to be met, since an inherent property of this type of evaluation is the potential subjectivity that comes with it, as expressed within the standard (ISO, 1995) itself at several places.

<u>Different outcomes of an evaluation activity</u>

With the complexity involved in an evaluation activity, it is not always the situation that an evaluation outcome is necessarily a satisfactory problem solution. Instead, Almstrum *et al*. (1996) argue that a possible outcome can be simply "a better question." (p. 202) Their view of an evaluation process, as a long-term iterative activity that takes place in the context in which the phenomenon being evaluated is to be used, ultimately aims to illuminate the implications of a complex technology for that usage situation (Almstrum *et al*., 1996, p. 202). Thereby, the process can be viewed as a long-term learning process.

The learning aspect has been emphasised by other authors with respect to evaluation of information systems. For example, Etzerodt and Madsen (1988) argue that:

> "Information system evaluation should be seen as a process where the actors involved learn from their own experiences and not just as a process of measuring a fixed set of system values." (p. 334)

A very different view is exemplified by Williams (1992) who, in the context of software product evaluation, states that:

> "[i]n order to compare products it is beneficial to reduce the information collected to a single score." (p. 121)

Such an outcome can, essentially, be used as 'input' to decision making. As mentioned earlier, the goal for such evaluations has been characterised as "Go/No-Go" decisions (Rossi and Freeman, 1993, p. 47) and is, obviously, fundamentally different from an outcome which emerges as a consequence of an evaluation being undertaken with a *learning* emphasis. For such evaluations, the outcome can, instead, be an in-depth understanding that has been gained during the process of undertaking the evaluation activity. In other words, the outcome for such an activity is *process* oriented (where *understanding* and *insight* are central), whereas a 'single score' outcome from an evaluation very much has a *product* focus (where *assessment* and *judgement* are central).

In reviewing work on a so-called "crisis in utilization", Blackler and Brown (1988) observe a "general tendency to treat evaluation in isolation" (p. 361) from corporate policy and decision making, and not to use evaluation results "to aid long term organizational objectives." (p. 362)

For example, whatever the purpose of an evaluation activity, significant effort will have been expended on developing an evaluation framework. This framework will reflect both long-term objectives and pragmatic issues concerning specific tools under evaluation. It has been our experience that these aspects are distinct but related, and that in practice two evaluation frameworks should be maintained. The first of these represents only long-term organisational need, the second is instantiated for use in the specific evaluation activity in progress. Even in a 'go/no-go' evaluation activity, the long-term evaluation framework can be seen as a useful outcome and could form a basis for organisational learning.

Dissemination of evaluation findings is an important aspect for any evaluation effort. Communicating evaluation frameworks to stakeholders is an important facet of dissemination. However, other in-depth insights, and experiences gained by stakeholders involved in an evaluation activity undertaken within an organisational context, do not speak for themselves (Fetterman, 1997). Therefore, it is important that each specific evaluation activity has an appropriate means for communicating findings and results to stakeholders that are (directly or indirectly) affected by the outcome of the activity. Fetterman (1997) argues that the evaluator has an important role in this communication process.

CONCLUSIONS

In this paper we have critically considered the literature related to the evaluation of IS development tools, and in particular that related to evaluation for possible adoption. Where relevant, we have critically considered the ISO standard on evaluation, both in terms of its contribution to the debate and in terms of its implied shortcomings.

From our analysis, three primary dimensions are evident in which any evaluation is planned and carried out: the *stakeholder dimension*, the *contextual dimension*, and the *activity dimension*. We

have described each dimension, and explained why each should be actively considered in any evaluation effort. We summarise these findings below.

<u>Stakeholders</u>

For an evaluation to succeed, that is produce valid and accepted findings, stakeholders must feel able to actively participate, and their beliefs and assumptions must be considered. Validity is enhanced by the multiple perspectives offered by the inclusion of representative stakeholders, but effective and ongoing feedback is essential to preserve a sense of ownership. However, evaluation is a political activity, potentially strongly affecting stakeholders, so selection of stakeholders and the choice of evaluator are critical factors for success.

When selecting stakeholders, it is important to have a clear strategy. For example, there may be a need to cover all aspects of a process life-cycle, or to include all roles within the context potentially affected by the evaluation activity. It is important to establish clear goals, but also to acknowledge that each stakeholder will have individual motivations and personal goals. Goals can be very diverse, from educational goals to financial assessment and control. It is also possible to target several goals simultaneously, for example targeting both organisational learning and tool selection within the same activity. Finally, there is the special role of evaluator, which may be taken by stakeholders themselves or may involve an external agency. When selecting an evaluator, there is a clear choice between appointing a stakeholder and using an external agency.

These selection decisions are clearly political decisions, and should ideally be resolved transparently, and to the satisfaction of all stakeholders, prior to the commencement of any evaluation activity.

Failure to consider the stakeholder dimension can result in a feeling of lack of ownership of the process by stakeholders, and undesirable outcomes. These can range from poor tool selection, with consequential financial penalties, to political fallout.

<u>Context</u>

Serafeimidis and Smithson (1996) separate context into *internal*, *external*, and *history*. There is also an argument for discriminating further with respect to internal context, differentiating between *usage* context and *organisational* context according to closeness to the phenomenon being evaluated (Lundell and Lings, 2003). For example, a tool user will be considered part of the usage context whereas a manager involved in policy documents for tool adoption will be considered as part of the more general organisational context. Further, in the framework presented, history is considered only as one facet of each context, rather than a separate element, as it is important to emphasise the continuing change of contexts.

Failure to consider the contextual dimension can result in outcomes of limited relevance to the organisational setting, and in the extreme an evaluation becomes a context free comparison of tool features.

<u>Activity</u>

A number of different approaches, and associated sets of activities, have been proposed for evaluation. The most descriptive of these is the ISO 14102 standard (ISO, 1995; Lundell and Lings, 2002) for CASE-tool evaluation and selection. It is clear that an appropriate set of activities for a given evaluation situation is very dependent on underlying assumptions about the context of the activities and the stakeholders involved in those activities.

Failure to consider the activity dimension can result in poorly conducted evaluations and unreliable outcomes, leading to a lack of stakeholder confidence in the whole activity.

Further Implications

In considering the literature in the light of this framework, several specific weaknesses in current approaches are apparent. In brief, these relate to:

- (implicit) assumptions about the ability to separate consideration of organisational needs from those of current technology when composing an evaluation framework;
- lack of method support for considering both of the above aspects within an evaluation, and for adapting pre-existing frameworks;
- the sequential nature of the assumed overall process in ISO 14102, with an implied mutual independence of the outputs from each component process.

We have earlier claimed that evaluation methods require improved support for evaluation framework development (Lundell and Lings, 2003). In the light of our analysis we would go further, and suggest that all phases of an evaluation process require enhanced support. In particular, we see the following as open questions worthy of further research:

- How can evaluation frameworks be enhanced to effectively capture information related to the impact on an organisation should an IS development tool fail to offer support for an implied requirement?
- What decision mechanisms are appropriate for supporting selection through a full evaluation given an evaluation framework so enhanced?

We believe that these questions must be answered in the full light all of the elements of evaluation identified in this paper, and in particular the stakeholder and contextual dimensions.

REFERENCES

Almstrum, V.L., Dale, N., Berglund, A., Granger, M., Little, J.C., Miller, D.M., Petre, M, Schragger, P. & Springsteel F. (1996) "Evaluation: turning technology from toy to tool: Report of the Working Group on Evaluation", **ACM SIGCSE Bulletin**, Vol. 28 (Special Issue), pp 201-217.

Avgerou, C. (1995) "Evaluating Information Systems by Consultation and Negotiation", **International Journal of Information Management**, Vol. 15 (6), pp 427-436.

Bailer, B., Bauknecht, K. & Schatzmann, C. (1993) "Ways to CASE: An Analysis of Selected CASE Experiences", In H.-Y. Lee, T.F. Reid & S. Jarzabek (Eds.) **Proceedings of the Sixth International Workshop on Computer-Aided Software Engineering: CASE '93**, IEEE Computer Society Press, Los Alamitos, California, pp 36-43.

Blackler, F. & Brown, C. (1988) "Theory and Practice in Evaluation: The case of the new Information Technologies", In N. Bjørn-Andersen, G.B. Davis (Eds.) **Information Systems Assessment: Issues and Challenges - Proceedings of the IFIP WG 8.2 Working Conference on Information Systems Assessment**, North-Holland, Amsterdam, pp 351-367.

Brandon, P.R. (1998) "Stakeholder participation for the purpose of helping ensure evaluation va-lidity: Bridging the gap between collaborative and non-collaborative evaluations", **American Journal of Evaluation**, Vol. 19 (3), pp 325-337.

Brown, A.W. & Wallnau, K.C. (1996) "Framework for evaluating software technology", **IEEE Software**, Vol. 13 (5), 39-49.

Budgen, D. & Thomson, M. (2003) "CASE tool evaluation: experiences from an empirical study", **Journal of Systems and Software**, Vol. 67 (2), pp 55-75.

Carney, D.J. & Wallnau, K.C. (1998) "A basis for evaluation of commercial software", **Information and Software Technology**, Vol. 40 (14), pp 851-860.

Chelimsky, E. (1997) "The Coming Transformations in Evaluation", In E. Chelimsky & W.R. Shadish (Eds.) **Evaluation for the 21st Century: A handbook**, SAGE Publications, Thousand Oaks, California, pp 1-26.

Coupe, R.T. (1994) "A critique of the methods for measuring the impact of CASE software", **European Journal of Information Systems**, Vol. 3 (1), pp 28-36.

Coupe, R.T. & Onodu, N.M. (1997), "Evaluating the impact of CASE: An empirical comparison of retrospective and cross-sectional survey approaches", **European Journal of Information Systems**, Vol. 6 (1), pp 15-24.

Cronholm, S. (1995) "Why CASE Tools in Information Systems Development? - an Empirical Study Concerning Motives for Investing in CASE Tools", In B. Dahlbom, F. Kämmerer, F. Ljungberg, J. Stage & C. Sørensen (Eds.) **Proceedings of the 18th Information Systems Research Seminar in Scandinavia: IRIS 18 - "Design in Context"**, Gothenburg Studies in Informatics, Report 7, pp 132-144.

Cronholm, S. (1998) **Metodverktyg och användbarhet - en studie av datorstödd metodbaserad systemutveckling**, Ph.D. thesis, Department of Computer Science, Linköping University Linköping, Sweden (in Swedish).

Etzerodt, P. & Madsen, K.H. (1988) "Information Systems Assessment as a Learning Process", In N. Bjørn-Andersen & G.B. Davis (Eds.) **Information Systems Assessment: Issues and Challenges - Proceedings of the IFIP WG 8.2 Working Conference on Information Systems Assessment**, North-Holland, Amsterdam, pp 333-345.

Farbey, B., Land, F. & Targett, D. (1999) "Moving IS evaluation forward: learning themes and research issues", **Journal of Strategic Information Systems**, Vol. 8 (2), pp 189-207.

Fetterman, D. (1997) "Empowerment evaluation: A response to Patton and Scriven", **Evaluation Practice**, Vol. 18 (3), pp 253-266.

Forss, K., Cracknell, B. & Samset, K. (1994) "Can evaluation help an organization to learn?", **Evaluation Review**, Vol. 18 (5), pp 574-591.

Greene, J.C. (1997) "Evaluation as Advocacy", **Evaluation Practice**, Vol. 18 (1), pp. 25-35.

Guba, E.G. & Lincoln, Y.S. (1989) **Fourth Generation Evaluation**, SAGE Publications, Newbury Park.

Henderson, J.C. & Cooprider, J.G. (1990) "Dimensions of I/S Planning and Design Aids: A Functional model of CASE Technology", **Information Systems Research**, Vol. 1 (3), pp 227-254.

Hughes, J. & Wood-Harper, T. (2000) "An empirical model of the information systems development process: a case study of an automotive manufacturer", **Accounting Forum**, Vol. 24 (4), pp 391-406.

IEEE (1992) **IEEE Recommended Practice for the Evaluation and Selection of CASE Tools**, IEEE Std 1209-1992, December 13, IEEE Standards Board, New York.

Irani, Z. & Love, P.E.D. (2001) "Information systems evaluation: past, present and future", **European Journal of Information Systems**, Vol. 10(4), pp 183-188.

ISO (1995) **Information Technology - Guideline for the evaluation and selection of CASE tools**, ISO/IEC JTC1/SC7/WG4, ISO/IEC 14102:1995(E).

ISO (1998) **Information Technology - Software Product Evaluation - Part 5: Process for evaluators**, ISO/IEC 14598-1:1998(E), 1998-07-01.

ISO (1999) **TR 14471 - Software Engineering - Guidelines for the adoption of CASE tools**, ISO/IEC JTC1/SC7 WG4, ISO/IEC, 1999-04-20.

ISO (1999b) **Information Technology - Software Product Evaluation - Part 1: General Overview**, ISO/IEC 14598-1:1999, 1999-04-15.

Jackson, B. & Sulaksono, S. (1998) "Going Soft on Information Systems Evaluation", **The Australian Journal of Information Systems**, Vol. 5 (2), pp 41-50.

Jankowski, D. (1997) "Computer-aided software systems", **Empirical Software Engineering: an international Journal**, Vol. 2, 11-38.

Jones, S. & Hughes, J. (2001) "Understanding IS evaluation as a complex social process: a case study of a UK local authority", **European Journal of Information Systems**, Vol. 10 (4), pp 189-203.

Juric, R. & Kuljis, J. (1999) "Building an Evaluation Instrument for OO CASE tool Assessment of Unified Modelling Language Support", In **Proceedings of the Thirty-second Annual Hawaii International Conference on System Science** [CD-ROM], IEEE Computer Society Press.

Kaipala, J. (1997) "Augmenting CASE Tools with Hypertext: Desired Functionality and Implementation Issues", In A. Olivé & J.A. Pastor (Eds.) **Advanced information Systems Engineering: 9th International Conference, CAiSE'97**, Springer, Berlin, pp 217-230.

Kitchenham, B.A. (1996) "Evaluating Software Engineering Methods and Tool - Part 1: The Evaluation Context and Evaluation Methods", **Software Engineering Notes**, Vol. 21 (1), pp. 11-15.

Kitchenham, B.A. & Jones, L. (1997) "Evaluating Software Engineering Methods and Tool - Part 5: The Influence of Human Factors", **Software Engineering Notes**, Vol. 22 (1), pp 13-15.

Kitchenham, B., Linkman, S. & Law, D. (1997) "DESMET: a methodology for evaluating software engineering methods and tools", **Computing & Control Engineering Journal**, Vol. 8 (3), pp 120-126.

Lending, D. & Chervany, N.L. (1998) "CASE tools: understanding the reasons for non-use", **ACM SIGCPR Computer Personnel**, Vol. 19 (2), pp 13-26.

Lundell, B. & Lings, B. (2002) "Comments on ISO 14102: the standard for CASE-tool evaluation", **Computer Standards & Interfaces**, Vol. 24 (5), pp 381-388.

Lundell, B. & Lings, B. (2003) "The *2G* Method for Doubly Grounding Evaluation Frameworks", **Information Systems Journal**, Vol. 13 (4), pp 375-398.

Maansaari, J. & Iivari, J. (1999) "The evolution of CASE usage in Finland between 1993 and 1996", **Information & Management**, Vol. 36 (1), pp 37-53.

Maccari, A. & Riva, C. (2000) "Empirical Evaluation of CASE Tools Usage at Nokia", **Empirical Software Engineering**, Vol. 5 (3), pp 287-299.

Maccari, A. & Riva, C. (2002) "On CASE Tool Usage at Nokia", In **17th IEEE International Conference on Automated Software Engineering (ASE'02)**, IEEE Computer Society, 23 - 27 September 2002, Edinburgh, pp 59-68.

McAuley, L., Doherty, N. & Keval, N. (2002) "The stakeholder dimension in information systems evaluation", **Journal of Information Technology**, Vol. 17 (4), pp 241-255.

Moriarty, T. (1998) "Searching for the Right Data Modeling Tool", **DM Review**, Vol. 8 (6), pp 74-76.

OVUM (1993) **Ovum Evaluated: Case Products**, Ovum, London, England, January.

Patton, M.Q. (1996) "A World Larger than Formative and Summative", **Evaluation Practice**, Vol. 17 (2), pp 131-144.

Post, G., Kagan, A. & Keim, R.T. (1998) "A comparative evaluation of CASE tools", **Journal of Systems and Software**, Vol. 44 (2), pp 87-95.

Post, G.V., Kagan, A. & Keim, R.T. (1999) "A Structural Equation Evaluation of CASE Tools Attributes", **Journal of Management Information Systems**, Vol. 15 (4), pp 215-234.

Powell, A., Vickers, A, Williams, E. & Cooke, B. (1997) "A practical strategy for the evaluation of software tools", In S. Brinkkemper, K. Lyytinen & R.J. Welke (Eds.) **Method Engineering: Principles of method construction and tool support - Proceedings of the IFIP TC8 WG8.1/8.2 Working Conference on Method Engineering**, Chapman & Hall, London, pp 165-185.

Ramage, M. (1997) "Developing a methodology for the evaluation of cooperative systems", In K. Braa, E. Monteiro (Eds.) **Proceedings of the 20th Information Systems Research In Scandinavia, Conference Proceeding Nr. 1**, Department of Informatics, University of Oslo, Norway, pp 769-789.

van Reeken, A.J. & Trienekens, J.J.M. (1992) The Practical Importance of Methods and Case Tools: Results of Empirical Research, MERIT 92-015, Maastricht Economic Research Institute on Innovation and Technology, University of Limburg, Maastricht, Netherlands.

Rehbinder, A., Lings, B., Lundell, B., Burman, R. & Nilsson, A. (2001) "Observations from a field study on Developing a Framework for Pre-usage Evaluation of CASE-tools", In N.L. Russo, B. Fitzgerald & J.I. DeGross (Eds.) **New Directions in Information Systems Development: IFIP WG 8.2 2001 Conference**, Kluwer, Boston, pp 211-220.

Remenyi, D., Sherwood-Smith, M. & White, T. (1997) **Achieving Maximum Value From Information Systems: A Process Approach**, John Wiley & Sons, Chichester.

Rossi, P.H. & Freeman, H.E. (1993) **Evaluation: A Systematic Approach**, SAGE Publications, Newbury Park, London.

Scriven, M. (1996) "Types of Evaluation and types of Evaluator", **Evaluation Practice**, Vol. 17 (2), pp 151-161.

Serafeimidis, V. & Psoinos, A. (1995) "Emerging Issues and a Holistic Approach for Information Technology Investments Evaluation into the 1990's", In **5th Hellenic Conference on Informatics**, Athens, 7-9 December 1995.

Serafeimidis, V. & Smithson, S. (1996) "The Management of Change for Information System Evaluation Practice: Experiences from a Case Study", **International Journal of Information Management**, Vol. 16 (3), pp 205-217.

Serafeimidis, V. & Smithson, S. (2003) "Information systems evaluation as an organizational institution – experience from a case study", **Information Systems Journal**, Vol. 13 (3), pp 251-274.

Silverman, D. (1998) "Qualitative research: meanings or practices?", **Information Systems Journal**, Vol. 8 (1), 3-20.

Smyth, R.W. (1998) "Use of CASE tools by a software vendor: a study of success indicators", In W.R.J. Baets (Ed.) **Proceedings of the 6th European Conference on Information Systems**, Euro-Arab Management School, Aix-enProvence, pp 571-583.

Symons, V. (1990) "Evaluation of information systems: IS development in the Processing Company", **Journal of Information Technology**, Vol. 5 (4), pp 194-204.

Symons, V.J. (1991) "A review of information systems evaluation: content, context and process", **European Journal of Information Systems**, Vol. 1 (3), pp 205-212.

Symons, V. & Walsham, G. (1988) "The Evaluation of Information Systems: A Critique", **Journal of Applied Systems Analysis**, Vol. 15, pp 119-132.

Vessey, I. & Sravanapudi, A.P. (1995) "CASE Tools as Collaborative Support Technologies", **Communications of the ACM**, Vol. 38 (1), pp 83-95.

Walsham, G. (1993) "Evaluation", In G. Walsham (Ed.) **Interpreting Information Systems in Organizations**, John Wiley & Sons, Chichester, England, pp 165-185.

Williams, F. (1992) "Appraisal and evaluation of software products", **Journal of Information Science: Principles & practice Institute of information scientists**, Vol. 18 (2), pp 121-125.