# Theoretical basis of health IT evaluation

Brender McNair, Jytte

*Published in:*
Evidence-Based Health Informatics

*DOI (link to publication from Publisher):*
10.3233/978-1-61499-635-4-39



*Publication date:*
2016

*Document Version*
Publisher's PDF, also known as Version of record

*Evidence-Based Health Informatics*
*E. Ammenwerth and M. Rigby (Eds.)*


*doi:10.3233/978-1-61499-635-4-39*

# Theoretical Basis of Health IT Evaluation

Jytte BRENDER McNAIR[a,1]
[a]*Dept. of Health Science & Technology, Aalborg University*

**Abstract.** The focus of this contribution is on the theoretical principles and concepts behind evaluation of IT-based systems, discussing their presuppositions, implications and interrelationships; for instance in relation to a series of issues to consider: terminology for the concepts used as that is a reason for many disputes, bias as that is a common reason for less accuracy and trustworthiness in conclusions, culture as the tacit driver of everything we do and design, constructive evaluation as this has strict time and timing issues, preparing for meta-analyses as that is in the near future, and top-level issues in choice of methodology. Awareness in these respects will lead to avoidance of major pitfalls and perils at evaluation and thereby improve the validity and trustworthiness of an evaluation outcome, supporting the initiative towards evidence-based health informatics.

**Keywords.** Evaluation, systems theory, IT system, bias.

## 1. Introduction

> *"Evaluation is the act of measuring or exploring properties of a health information system (in planning, in development, in implementation, or in operation), the result of which informs a decision to be made concerning that system in a specific context. Evaluation of health information systems has to deal with the actors, the artifacts, and their interaction to best support the decisions to be made."* [1].

Many tend to believe that evaluation is something everyone can do. It is indeed a common day activity in one's life and it appears so easy. Nevertheless, there is a difference between providing somewhat random or subjective evidence and providing measures based on systematic judgements, and one has to know which of the two extremes to apply when. When you buy a new car, would you do it only based on your own test-drive? Probably not, you would likely read the professional associations' assessment from test-driving, crash tests and more, and then use your own assessment to add a judgement on whether the car really is suited for your practical purpose, because there are always compromises to be made.

'Providing evidence' requires a stringent approach adopting the principles and methods used in science, and evaluation of health IT applications is a scientific discipline. The purpose of this book is to contribute to healthcare through the concept of 'evidence-based health informatics'. Therefore, in order to secure maximum impact,

[1] Corresponding author: Emeritus associate professor Jytte Brender, Dept. of Health Science & Technology, Aalborg University, Fredrik Bajer's Vej 7D, DK-9220 Aalborg East, Denmark, jytte@brender.dk or jbr@hst.aau.dk.

the focus of this contribution shall be on those basic principles and concepts of evaluation that have a major impact on the validity and trustworthiness of an evaluation outcome. Thus, the aim of this contribution is to provide a scientifically-minded reader with the theoretical background for starting to design an evaluation study, and to show the non-scientific reader the importance and content of a structured objective approach.

The theoretical foundation of any scientific discipline is the philosophy of science, where the dictionary definitions of 'philosophy' includes two relevant for our purpose, both from [2]: "the academic discipline concerned with making explicit the nature and significance of ordinary and scientific beliefs and investigating the intelligibility of concepts by means of rational argument concerning their presuppositions, implications and interrelationships"; and "the critical study of the basic principles and concepts of a discipline.". Similarly, we see a theory as "a set of hypotheses related by logical or mathematical arguments to explain and predict a wide variety of connected phenomena in general terms." [2]. It is such arguments that this contribution will outline, arranged under the following headings:

1) Grounding possibilities: Theoretical assumptions and methodological considerations founding an evaluation study; value norms and raison d'être.
2) Communicative interactions: Matching scope, practical assumptions and delimitations.
3) Identifying and balancing the risk of bias in health IT evaluation.
4) Decision-making preferences: Culture is the driver of our decision-making whether we know it or not, and whether we want it or not.
5) Time and timing of evaluation: The concept of a constructive evaluation as opposed to traditional (summative) evaluations.
6) The next stage – that is, the indeterministic nature of systems development demands sustainability through flexibility and fluidity, and the demand for evidence enforce a next stage of methodological approaches.
7) Selecting/choosing the appropriate and sufficient methodology.

## 2. Grounding Possibilities: Theoretical Assumptions and Methodological Considerations

Which theoretical assumptions and methodological considerations can and should be the basis for an evaluation study? And when are they relevant to consider?

There are a set of interlinked concepts, like methodology, perspective and culture, for which a deep understanding will support the initiation of a successful planning and/or accomplishment of an evaluation study. In the following, these terms will be discussed briefly, also showing how important it is to always make one's terminology and values explicit.

"The term 'methodology' signifies "the science of methods" (BIPM et al. 1993) from the Greek 'logos', which means "the science of". In functional terms it relates to the knowledge of how to prepare and use methods. Expressed in structural terms a methodology consists of "a coherent set of methods covering all the sub-tasks necessary for a given undertaking". In other words, a methodology is supposed to a) provide the answer to what to do next, when to do what and how to do it, and b) to describe the ideas behind such choices and the suppositions (for instance the philosophical background) behind them." ([3], p. 14).

Designing an evaluation study is in some ways like peeling an onion, because one answer brings out a next level with a number of new questions. For instance, before one can even start thinking of which methodology to choose one has to have a goal in order for the methodology to make the starting point and the end point meet – that is implicit from the structural definition above. At one such 'deeper' level before deciding on the methodology, one has to make clear one's theoretical assumptions, for instance, a high-level reductionistic[2] versus a holistic perspective.

Such decisions are drivers towards the actual planning of an evaluation study. "The concept of 'perspective' stands for hidden aspects and assumptions deeply buried in the design and application of methods, see for instance (Mathiassen & Munk-Madsen 1986; Arbnor & Bjerke 1997; Brender 1997). In a generalized version, the perspective is the implicit assumptions of (cause-effect relations within) the object of study. So, the perspective is synonymous with "that aggregation of (conscious or unconscious, epistemological) assumptions of how things relate in combination with imprinted attitudes guiding our decision making e.g. in a problem solving situation"." ([3] p. 18).

Few method designers are aware that our cultural background (professional, religious and national) maintains a series of tacit assumptions affecting our way of doing and perceiving things; see for instance [4] and [5]; as well as a brief overview in [6]. Caused by the tacit nature (i.e. completely unknown to inexperienced users of a given method or methodology), some perspectives may contain pitfalls, where the perspective of a method conflicts with the actual purpose which the method is intended to be used for. This is why this concept is so important in a profession-oriented context, here evaluation of health IT applications. A couple of simple examples of the implications of culture will illustrate this:

- In some Asian and African cultures it is highly impolite to answer a question with a 'No'. Then think of many traditional questionnaires (which are typical evaluation instruments) or radio-buttons in the screen interactions between a computer and an end-user (i.e. relevant in a usability test). It doesn't matter that the application as such isn't situated in either of the cultures mentioned, because with today's intensive migration of labour forces these cultures will be present everywhere, and adaptation to a new local culture is not something that comes overnight. Hence, such culture may unintentionally impose a bias in the evaluation outcome, or even worse may unintentionally compete with the design principles behind the screen functionality and falsify the input from a user-computer dialogue.
- In some cultures, a manager is considered the ultimate decision-maker (actually a decision-taker), he/she is always right, and the accuracy of his/her information or the appropriateness of his/her decision-making is never questioned by his/her employees. Then obviously, interviews and questionnaires involving end-users on the floor have a built-in risk of bias.

Clearly an evaluator has to manoeuvre within such local organisational context and conditions.

[2] In a reductionistic perspective one can observe a system's components individually and then combine all such observations to provide a true and complete explanation of the whole system under investigation or of that part of the system which those components represent. In a holistic perspective, component parts of a system are dynamically interrelated and should be viewed within that wholeness.

Evaluation[3] from a user perspective is always about an IT system operating within a context, and that context is the user organisation (i.e. another 'system'). There are two definitions of 'system' that we find beneficial, and together they capture the essence of the concept of a 'system' for the present purpose. A system is:

> *"all the components, attributes and relationships needed to accomplish an objective"* [7], p. 483 – that is, it has purpose, structure, behaviour and interactions, internally and across its boundary; this implicitness led in turn to the following definition, which however, is still not perfect:
>
> *"An organisation in which all structural components and dynamics are interrelational, participating internally, and affecting conditions externally"* [8], p. 480. An 'organisation' here is to be understood in a business or administrative sense.

Both definitions emphasise the intricate relationship and hence dynamics between a system's components and their properties. One can definitely use a reductionistic approach, and most methods applicable for evaluation are somewhat reductionistic; however, one has to be aware of the implications of the assumptions that one implicitly takes for given methods. It is always relevant to take this into consideration.

At an early point in time, make the policy and values as evaluator explicit; policy may be derived from one's values. They show your view on what evaluation really is – that is, its role in a larger context and interrelations with components within the systems under evaluation. In systems development, the different development methodologies and methodologies for project management have implicit and/or explicit built-in values. This is not any different for evaluation. Further, policy and values constitute a commitment towards external parties such as the user organisation and potential sponsors of the evaluation, but they also dictate the trajectory for achieving the goal, and hence the choice of methods to be involved.

Examples of value statements for health IT evaluations are:

- "User involvement is essential."
- "The users' capacity, skills and responsibilities will be respected."
- "Any interaction with users will take place on their premises, professionally, linguistically and otherwise."

Examples of policy statements are:

- "(Constructive) evaluation is a dynamic (non-deterministic) process that obeys the information need of the decision makers and not unnecessarily hampers or delays ongoing processes within the user organisation or the development organisation."
- "User involvement will be designed as specified tasks in manageable bites, in compliance with their respective managerial and professional competence, and will be continuously adjusted according to their respective relevance for the evaluation."

[3] Cf. the definition of evaluation, our perspective is the user's as opposed to the developers' debugging.

- "Even if there are theoretical considerations behind the practical tools and prescriptions, the user organisation shall not be unnecessarily distracted by being presented with this. That is, you will speak with the user organisation in their language and on their contextual premises."

Given the definition of 'evaluation', its outcome is going to be used within some context. Users speak their profession-oriented language, so one should not enforce one's own terminology upon them. Thus, the implication of for instance the last statement includes conditions for interacting with the users, including reports from the evaluations and perhaps even choice of evaluation method(s).

## 3. Communicative Interaction: Matching Scope, Practical Assumptions and Delimitations

Note how heavily this entire paper is filled with definitions. Definitions are indeed difficult to make, but when they are finally right they are extremely helpful instruments. Make the applied terminology clear before starting an evaluation study and then again when you wish to publish your study. Examples are the central terms 'evaluation', 'verification', 'validation' and 'assessment', where this author has witnessed so many heated disputes among colleagues – because of different use of the terms. Moreover, different domain professionals use the same terms as evaluators do, but in other contexts and with (slightly) different meanings; for instance, the term 'phase'.

It is not necessary to make the definitions oneself, but find in the literature those that resonate with the study' purpose and need. Being explicit about the terminology prevents a lot of miscommunication and misinterpretations, and may help harmonise the domain in which you are operating, and last but not least it contributes to securing 'evidence-based health informatics'.

Then, make practical assumptions and delimitations explicit: Even when one aims at performing a scientific evaluation study in order to secure an appropriate level of evidence, there are toes and heels that have to be cut before Cinderella's glass shoe will fit: available funding and local conditions, publication strategy/restrictions, confidentiality and personal data security, … The level of ambition and the evaluation set-up have to match the local realities, and that requires a communicative interaction with and within the organisational context to identify the local conditions, a necessity for aligning with the practical reality.

While taking all of these issues into account, make the scope explicit: What is the question that the evaluation is going to answer? This is a top-level decision, the target of the entire study, and that which the methodology has to fulfil.

In [9] you will find a lot more details and prescriptions on how to design an evaluation study in practice, like which aspects to consider in each part of the study design, implementation and reporting phases of the evaluation study, such as operationalisation (making practical) of methods from their abstract version.[4]

An example, who are the stakeholders, who are the beneficiaries and who are the victims? Victimhood needs not be in term of power, salary or esteem, but may be in

[4] For further discussion of these issues, see part II "Methodological considerations of health IT evaluation" in E. Ammenwerth, M. Rigby (eds.), Evidence-Based Health Informatics, Stud Health Technol Inform 222, IOS Press, Amsterdam, 2016.

terms of overruled professional responsibility. For example, a new IT system will change the business processes both in structure and content, and thereby potentially interfere with the user's responsibility or even liability (example: decision-support systems, or CPOEs).

## 4. Balancing the Risk of Bias

The necessity for stringency of scientific work emphasises the need to control all potential sources of variation, any bias within the study. A bias is an inclination towards a systematic deviation of measurements from the 'true' value for the population under investigation – that is, biased data may still be factual and objective but will not be accurate.

Bias is hard to avoid in evaluation of health IT applications, like in any scientific study, just think of the rule in physics that the mere observation of an object inevitably will change the object; so it is important to recognize and to the extent feasible control it. The Hawthorne effect is a similar bias identified in evaluation studies, and in a generalised version it simply states than an organisation (the study object and/or individual components and processes within it) under observation will change; see for instance [3]. It is important to recognize and control the bias at risk, or balancing parameters in the approach so that the bias will have minimal impact on the validity and trustworthiness of the study outcome. The issue is that one needs to be aware of and manoeuvre with the risk of biases in the planning, while constantly remembering that the theoretical and practical impact of biases shall be balanced against the study purpose and the study's role in a (users') decision-making context.

Since this contribution is limited with respect to space, only an abstraction of the issue of biases will be provided here, and the reader is referred to [3] (pages 253-313), where biases are discussed in terms of a meta-framework for assessing evaluation studies. Threats to validity of health IT evaluation studies are discussed elsewhere. [5]

Examples of biases are:

- Selection skewness – that is, when matching structures or components during the recruitment process for controlled studies intended comparables may easily be incomparable (like apples and bananas) and will provide different outcomes, such as comparing a given system when applied in different medical specialties, or recruiting physicians at different professional competence levels (e.g. registrar versus chief physician) for evaluating a decision-support system.
- Skewed frame of reference, for instance, placebo effects, Hawthorne effects, carry-over effects, checklist effects, confounding factors, and more.
- Value-based (emotional inclination), for instance, 1) a developer's versus a user's assessment and therefore a developer should have no influence on an evaluation study before, during or after the event; and 2) technophile or technophobia-based judgements.

[5] See C.R. Weir, Ensuring the quality of evidence: Using the best design to answer health IT questions, in: E. Ammenwerth, M. Rigby (eds.), Evidence-Based Health Informatics, Stud Health Technol Inform 222, IOS Press, Amsterdam, 2016.

- Cognitive-based, for instance a range of judgemental observer effects that have been identified in cognitive science, such as judgement of probabilities, insight bias, as well as post-rationalisation, and many more.
- Culture-based (judging phenomena based on one's own culture): cultural decision-making preferences (see later in this paper) will make people behave differently (stubbornly so and significantly so) for instance in the manager-context mentioned briefly in an earlier section.
- One particular bias worth mentioning here is circular inference: "Circular inference arises when one develops a method, a framework, or a technique dedicated to a specific (population of) case(s) and applies it on the very same case(s) for verification purposes." [3], p. 265. So, when designing a method for one particular case/purpose one cannot reuse that particular case to assess the validity of the same evaluation method.

The essence is that a bias when present may render one or more variables unable to reflect objectively the necessary population characteristics. Moreover, since bias is not black and white but comes from a scale of grey nuances, it need not have a significant impact in practise. The 'art' of science (actually 'craftsmanship', i.e. the ability to juggle with methods and handling perspectives) is to know which bias matters and which not for a given setting, and the size of the impact of biases at risk, while reflecting both such awareness and the impact on the conclusions of the study.

## 5. Decision-Making Preferences: The Impact of Culture

The driver of our decision-making is our respective 'culture' whether we know it or not, and whether we want it or not. Therefore, a simple awareness of its presuppositions, implications and interrelations with (evaluation) methodologies and methods is relevant to briefly address.

"Our understanding of the concept of 'culture' may be expressed shortly this way: *"By cultural behavior, we mean the stability across generations of behavioral patterns acquired through social communication within a group, and valued by the group"* (Maturana 1987, cited and discussed in (Demeester 1995). Culture is the style of working in the field, or the mental, tacit (learned) behavioral pattern behind the style of working (Hampden-Turner & Trompenaars 1993 and 1997; Trompenaars & Hampden-Turner 1997). Thus, culture is guiding the preferences; culture is what comes before starting a discussion of strategy, …, in a chain of causal events towards problem solving. When specifically talking of the interpretation of culture in an organizational context it means "the acquired preferences in problem solving", where problem solving should be understood in the broadest sense and not only as problem solving in a profession oriented perspective." ([3], p. 289). This is this author's perspective on the concept of culture.

Smaldino brings a thought-provoking example [11] (p. 251): "Perception is constrained in part by our biology, but culture also constrains even our basic perceptions of a situation (Nisbett & Miyamoto 2005; Smaldino & Richerson 2012). For example, (Masuda & Nisbett 2001) showed American and Japanese university students animated underwater scenes with a focal fish. In a recall task, Americans were much better identifying fish they had seen independent of background information, but

Japanese students were much better at remembering details of the background scenes." Considering that our perception (visual and other) provides the input for our decision-making, such difference is noteworthy. Therefore, beware that different cultures 'see' different things when they observe the same object. Further, Smaldino explicitly concludes that cultural differences in patterns of perception and memory fit larger cultural differences in epistemology and styles of thinking that exist between East and West [11].

'Culture' is one example of a nation and/or profession-oriented perspective that is hidden in the methods and methodologies that we use. From the two examples (Asian/African culture, and managers as decision-takers) in the Section on "Grounding Possibilities: Theoretical Assumptions and Methodological Considerations" plus one specific bias, it is obvious that culture has an impact on the choice of methodology and of methods; for instance, if 'No' is not an option for a fraction of the target end users, then questionnaires with yes/no answers should be excluded as candidate tools. Further, this should be seen in the context of our – subconscious/tacit – culturally conditioned way of perceiving a situation and interpreting observations or designing solutions. Few method designers make the cultural assumptions explicit, and therefore the evaluators need to be aware.

## 6. Time and Timing of Evaluation

There is a huge difference between constructive (also called formative) and summative evaluation. Constructive evaluation comprises evaluation activities that are completely intertwined with the systems development activities throughout the project (or for a circumscribed period/phase), while summative evaluation is concerned with evaluation at an end point in a developmental path or phase. Constructive evaluation may for instance take place during rapid prototyping, but also at usability studies, and even at the requirements specification phase. So, naturally there is a time and timing issue, because the evaluations cannot and must not significantly delay the systems development. Characteristic is that the outcome of constructive evaluation studies provides substantial input for revision and/or continued systems development, i.e. the trajectory for the future work, rather than merely a verification of contractual fulfilment such as is often the case with summative evaluation. Summative evaluation is also often used to gain insight or measurement of properties without a pressing decision-making information need.

Figure 1 shows the Dynamic Assessment Methodology. It is not a waterfall systems development model, irrespective of its depiction as four sequential phases. The difference is reflected in the contents laid in the arrows. It comprises an example of rapid prototyping developments. It is a model of constructive evaluation in a systems development context that is defined at four phases, from the point of conception of an idea, over iterations of a solution while iterating with usability evaluation, and till impact assessment in a mature clinical setting.

Now zooming out: the implication for scientific evaluation activities (i.e. those that you want to publish in the scientific literature) is tremendous: a) There is no fixed single frame of reference for the evaluation that is valid throughout, which means that the traditional approach to a user requirements document has to differ accordingly, for instance through rapid prototyping. b) There is a risk of a circular inference bias, see above (or more in [3]). c) Evaluation methods have to be chosen accordingly;

applicable methods for different phases and types of user assessment are marked in [3] by means of icons.

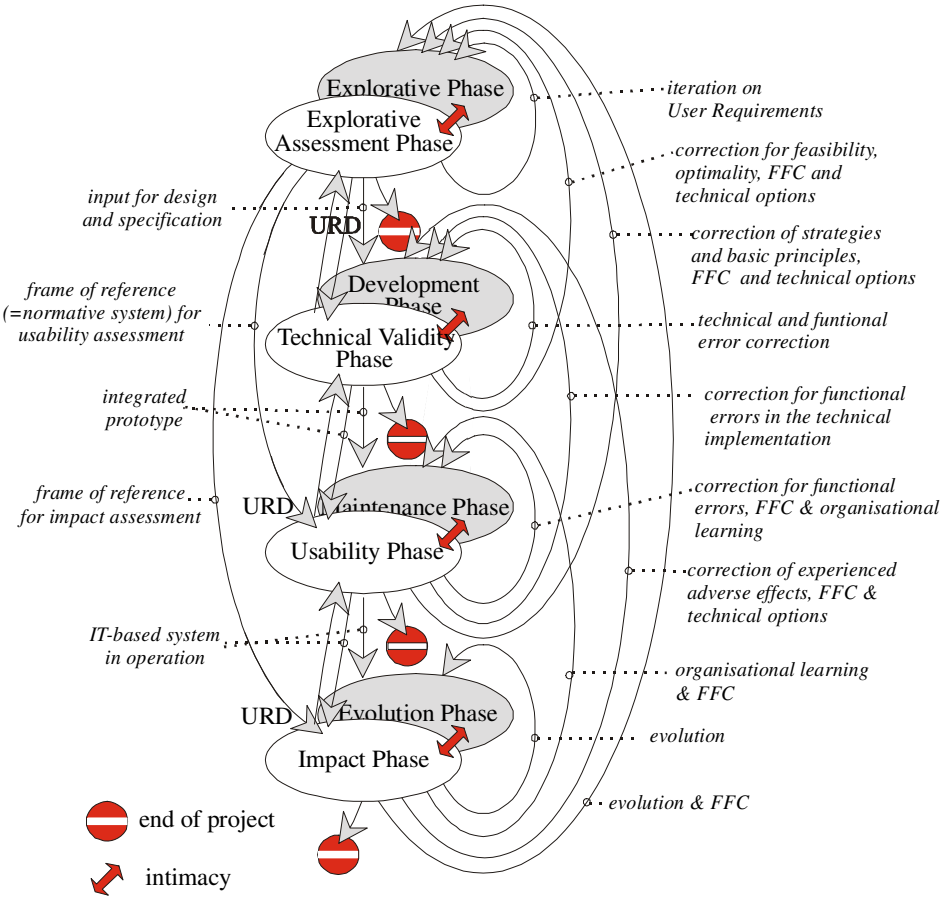


**Figure 1.** The Dynamic Assessment Methodology, complete with descriptions of feed-forward loops (providing frames of reference and preventive measures) and feed-back-loops (initiating corrective activities) and indicating the contents of this information flow. The shaded ellipses illustrate the technical or development activities, whereas the white ellipses illustrate the corresponding constructive assessment activities in a four-phased structure. The thick arrows indicate unspecified interaction between the technical development and assessment activities (co-ordination and collaboration). (URD = User Requirements Document; FFC = Four Founding Capacities, i.e. characteristics regarding the capability and capacity of accommodating changes), reproduced (modified) from ([3]).

The difficult aspects of constructive evaluation are the indeterminism of systems development and the consequential demand for creativity and innovation in order to continuously comply with the project's information need without delay. When publishing a study on constructive evaluation, one has to watch out for a circular

inference bias (see this above): a method or theory may indeed be highly successfully applied, but one cannot judge the method's various kinds of validity until assessed independently in another setting. Such aspects of validity include for instance: construct validity (does it really measure that which we believe or intend), internal validity (degree of compliance between the perceived meaning and the reality, i.e. with minimal bias), external validity (generalizability to other contexts of investigation), empirical validity (accuracy towards the true value of a measurement), rational validity (coverage or representativeness of characteristics), reliability (consistent outcome), and more; see e.g. [10] – or even Wikipedia – for more detail. When one is aware of this problem, the solution is to phrase the conclusion accordingly when relevant, for instance making a potential risk explicit and/or phrasing the certainty regarding the conclusion with caution (i.e. with weaker words).

## 7. The Demand for Evidence Enforce a Next Stage of Methodological Approaches

Systems development is indeterministic in nature (i.e. one cannot plan everything in detail, because things change or demand a new decision) and consequentially so is evaluation or at least that of constructive evaluation. Even at this point in the theoretical considerations, it is still important to keep an eye on whether – given the thoughts and considerations so far – the information need (that the evaluation is going to feed, cf. the definition of 'evaluation' in the introduction to this contribution) is likely to be appropriately fulfilled – that is, objectives fulfilment.

In an indeterministic context change is inevitable, so it is important to design the evaluation methodology while taking its sustainability into account. It means that the scientific evaluator needs the competence and experience to be able to incorporate sustainability into the evaluation methodology through fluidity and flexibility, creativity and innovation based on scientific premises.

At this point, a brief helicopter view will inform us whether all of the ends may converge into a coherent sustainable wholeness that in the end will fulfil the information need. They will; after all, evaluation of health IT systems have taken place for decades, meaning that a knowledge base of evaluation studies has accumulated that will pave the way for further studies and will support the validity of future evaluation studies.

Systematic reviews (in the literal sense of this concept) are already practised in the domain of evaluation in health informatics, as for instance seen from [12]. The next stage in evidence-based health informatics is the emergence of meta-analyses (in the Cochrane sense) of concrete cases of evaluation of health IT applications. Meta-analyses and systematic reviews provide answers to different questions, or answers with different levels of certainty attached.[6] The former has a quantitative nature while the latter has a qualitative (or quasi-quantitative) nature; see also [12]. Key to meta-analyses is comparability among studies as well as degrees of errors and bias, which again points at the importance in the reporting of specific details of the evaluation study rather than raising new methods; see the Section 'Recommended further

[6] See also: C. Urquhart et al., Systematic reviews and meta-analysis of health IT, in: E. Ammenwerth, M. Rigby (eds.), Evidence-Based Health Informatics, Stud Health Technol Inform 222, IOS Press, Amsterdam, 2016.

readings'. Errors and bias are really not easy to compensate for even if approaches are available, not least because of the nature of evaluation studies in health informatics as compared to those of clinical trials in medicine in general, or social sciences; nevertheless there may be food for thought and partial help in for instance [13].

## 8. Selecting the Appropriate and Sufficient Methodology

The key issue in selecting an appropriate and sufficient methodology while taking into account all of the above issues is match-making the scope with the methodology and applicable methods while building a plan based on strategic, tactical and operational issues, as seen in [9]. However, inherent within methodology is the supposition regarding the overall approach: is the evaluation study likely to be a desk-top exercise, a laboratory experiment or an application in a real case scenario? When it comes to evaluation of health IT applications aiming at providing evidence-based facts, then real-life evaluation studies are the most relevant. This is the issue to be discussed in the following.

Evaluation researchers need to demonstrate the validity of new evaluation approaches, methods, or theories, and users need information for their decision-making; thus together there is a potential for a fruitful partnership, which may be achieved through action-case research, and actually, some kinds of evaluation research have no alternative to action-case research. By 'action-case research' is meant the intentional trial application of a researcher's creation (a theory or method) in a real-life case with the purpose of verifying the validity of that creation in real practice. With this definition, 'action-case research' is not the same as the traditional definition of a 'case study' in the sense of illuminating the rationality and implication of a set of decisions, as defined by Yin [14]. We see 'action-case research' as being case-based action research.

McKay and Marshall define action research this way, "Action research is, quite literally, a coming together of action and research, or rephrased, of practice and theory." [15] (p. 219), that is, that such research is accomplished through action, or in other words utilising the research in a practical application. They discuss a set of approaches to action research, all involving informed action and reflection, but with varying degrees of control. The approaches range from, at one extreme, exerting full control – that is, the research interests have precedence in the decision-making regarding the evaluation issues; to an intermediate form where the real-world situation shapes the research interests and questions; to the other extreme, 'consultancy masquerading as action research', in which the real world have precedence over the research interests. "Apart from PhD projects, the majority of IT-systems development and implementation projects are for real-world usage, and the real-world is not an environment where a researcher can try out his methods or methodologies without consequence. Rather, the researcher has to accept the conditions of the real world. … the user organisation is … responsible for any decision that will impact their future practice. As an example, the health informatics applications may have huge implications for individual patients and/or for a hospital's economy, and hence may also have liabilities. That is why decision-making in such cases is a serious issue that the researcher can intervene with only within certain limits." [16], p. 51.

Action-case research needs a theoretical foundation and a methodology to guide the real-world problem situation, for instance a model for decision-making. The

research themes and the decision-making as well as the problem-solving are serious factors that must go hand in hand within the methodological design. In such a case, McKay and Marshall's recommendation of "Dual imperatives of Action Research" are taken appropriately into account. Further, in [16] the particular biases at risk in action-case research are discussed: Tacit knowledge and post-rationalisation, intention to treat, insight bias, circular inference, hypothesis fixation, as well as local minima. Apart from the last, they are all outlined above and discussed in more detail in [3]. The bias 'local minima' is a risk in large development or implementation projects where there is a succession of decision-making points; a non-optimal situation may arise when the basis for decision-making in a given situation points at a solution that constitutes a local minimum – that is, the decision appears optimal within the specific context, but may not be in a larger perspective.

With the above understanding it should be possible to start designing an evaluation methodology according to the guidelines in [9].

## 9. Discussion

The definition of philosophy states that "…making explicit the nature and significance of ordinary and scientific beliefs and investigating the intelligibility of concepts by means of rational argument concerning their presuppositions, implications and interrelationships", and this is what this contribution addressed. The theoretical considerations beneath stringent (evidence-based) evaluation of health IT applications belong under strategic factors in the framework comprising strategy, tactics and operations that are known from military operations and ISO9000. This contribution has dealt with only the strategic aspects at an early point of planning an evaluation study. The same issues have to be revisited during and after implementation - that is, at a follow-up; for example, (at least some of the) biases may be verified by measurement during the implementation of the evaluation study. The tactical aspects relate to making the study real in terms of choice of methodology, methods and action plans, while the operational aspects are concerned with the practical implementation of the evaluation study.

This contribution has mainly dealt with the presuppositions and implications, while the interrelations of the theoretical issues discussed have not yet been addressed. A framework was applied as a template for the entire contribution, namely that provided in [8]. This framework comprises seven sequential functions, each having a specific role in the wholeness of a system and each has an emergent property as output that serves as input for the subsequent function. This framework reflects a system in itself, and the dynamics, the interrelationships between the issues at hand, are handled through successive traversals of the framework – that is, *an iterative and incremental progression* of the issues dealt with; each of the issues discussed in this contribution are dependent on the solution of its predecessor issues.

## Recommended further readings

1. J. Brender, *Handbook of Evaluation Methods for Health Informatics*, Academic Press, New York, 2006.
2. J. Talmon, E. Ammenwerth, J. Brender, N. de Keizer, P. Nykänen, M. Rigby, STARE-HI - Statement on reporting of evaluation studies in Health Informatics, *Int J Med Inform* **78**(1) (2009), 1–9.
3. J. Brender, J. Talmon, N. de Keizer, P. Nykänen, M. Rigby, E. Ammenwerth, STARE-HI – Statement on Reporting of Evaluation Studies in Health Informatics, Explanation and Elaboration, *Appl Clin Inform* **4** (2013), 331–358.

## Food for thought

1. Why is the systems development model in Figure 1 not a waterfall model, even if it includes four primary and sequential phases? Explain the difference in nature and the implication for the practical evaluation activities.
2. What is the best approach to avoid or circumvent a circular inference bias and provide indicators of internal validity of one's study?
3. Explain the 'local minima' problem at constructive evaluation, described in the Section 'Balancing the Risk of Bias'. How does this concept relate to the other concepts discussed in the above? (e.g. bias, rapid prototyping, time and timing, …), and what are the implications?

## References

[1] E. Ammenwerth, J. Brender, P. Nykänen, H.-U. Prokosch, M. Rigby, J. Talmon, Visions and strategies to improve evaluation of health information systems - Reflections and lessons based on the HIS-EVAL workshop in Innsbruck, *Int J Med Inform* **73**(6) (2004), 479-491.
[2] *Collins English Dictionary,* Third Updated Edition, HarperCollins Publishers, Glasgow, 1995.
[3] J. Brender, *Handbook of Evaluation Methods for Health Informatics*, Academic Press, New York, 2006.
[4] F. Trompenaars, C. Hampden-Turner, *Riding the waves of culture – understanding cultural diversity in business,* 2nd edition, Nicholas Brealey Publishing Ltd, London, 1997.
[5] S.C. Schneider, J.-L. Barsoux, *Managing across cultures*, FT Prentice Hall, Harlow, 2003.
[6] Á. Cabrera, E.F. Cabrera, S. Barajas, The key role of organizational culture in a multi-system view of technology-driven change, *Int J Inf Manag* **21** (2001), 245-261.
[7] Y.Y. Haimes, C. Schneiter, Covey's seven habits and the systems approach: A Comparative Approach, IEEE Trans Syst, *Man Cybern* **26**(4) (1996), 483-487.
[8] L. Dennis, J. Brender McNair, L.H. Kauffman, eds., *The Mereon Matrix: Unity, Perspective and Paradox*, Elsevier, London, 2013.
[9] P. Nykänen, J. Brender, J. Talmon, N.F. de Keizer, M. Rigby, E. Ammenwerth, Guideline for good evaluation practices in health informatics (GEP-HI), *Int J Med Inform* **80**(12) (2011), 815-827.
[10] J.C. Alderson, C. Clapham, D. Wall, *Language test construction and evaluation*, University Press, Cambridge, 1995.
[11] P.E. Smaldino, The cultural evolution of emergent group-level traits, *Behav Brain Sci* **37**(3) (2014), 243-254.
[12] M. Rigby, E. Ammenwerth, M.-C. Beuscart-Zephir, J. Brender, H. Hyppönen, et al., Evidence based health informatics: 10 years of efforts to promote the principle. Joint contribution of IMIA WG EVAL and EFMI WG EVAL, *Yearb Med Inform* **8**(1) (2013), 34-46.

[13] J.E. Hunter, F.L. Schmidt, *Methods of Meta-Analysis, Correcting Error and Bias in Research Findings,* second edition, Sage Publications, Inc, Thousand Oaks, 2004.
[14] R.K. Yin, *Case Study Research, Design and Methods,* 3rd edition, Sage Publications, Inc, Thousand Oaks, 2003.
[15] J. McKay, P. Marshall, *Action Research. A Guide to Process and Procedure*, in: D. Remenyi, editor, Proceedings of the European Conference on Research Methodology for Business and Management Studies 2002. MCIL, Reading, 2002. pp. 219-227.
[16] J. Brender, *Research Methods for IT/IS Systems' Evaluation, A Critical Appraisal*, in: D. Remenyi, editor, Proceedings of the 2nd European Conference on Information Management and Evaluation, Royal Holloway University of London, 11-12th of September 2008, Academic Publishing Limited, Reading, 2008. pp. 51-58.