Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 00 (2013) 000–000
Procedia Computer Science
www.elsevier.com/locate/procedia

Conference on Systems Engineering Research (CSER'13)
Eds.: C.J.J. Paredis, C. Bishop, D. Bodner, Georgia Institute of Technology, Atlanta, GA, March 19-22, 2013.

Systems Engineering Research Methods

Gerrit Muller*

Buskerud University College, Kongsberg, Norway

Abstract

One of the challenges of systems engineering research is that the expertise and the application happen "in the field". The field can be an industrial company or a government agency. Researchers need methods to research in the field; methods to try-out ideas, collect data, analyze, and evaluate.
In this paper, we present a simple research model to help researchers to shape their research. We provide and discuss a number of elementary research methods. The model and the elementary research methods are based on our systems engineering research experiences in the past five years. We conclude the paper with some validation challenges and pitfalls.

Selection and/or peer-review under responsibility of Georgia Institute of Technology.
Systems Engineering; research methods

1. Systems Engineering Research

Systems Engineering is a young discipline that is mostly practiced "in the field", e.g. in industrial companies. Many systems engineers do their work based on experience [1]. An experienced group of INCOSE editors filled the Systems Engineering Body of Knowledge (SEBoK) [2] with a rich collection of best practices. Hot questions in the systems engineering community is what the return on investment is of applying systems engineering [3], or what the value is of applying systems engineering [4]. More general the academic question is "how do we know the effectiveness of these best practices?"

In this paper, we will discuss research methods to study effectiveness of systems engineering methods and techniques in practice. We focus on field research, under the assumption that researchers can only observe effectiveness in practice in the field. Dominant research approaches to study in the field are action research [5] and industry-as-laboratory [6, 7]. The researcher combines active participation in the systems engineering activities with the researcher role. The researcher is wearing two hats in these research approaches: as systems engineer and as researcher. The attitude of the systems engineer is result oriented and cooperative. The attitude of the researcher is questioning and challenging. Where the systems engineer will promote the use of a proposed method or technique, the researcher needs to question its validity.

* Corresponding author. Tel.: +4732869594; fax: +4732869551.
E-mail address: Gerrit.muller@gmail.com.

Researchers in systems engineering are typically educated in the technical domain. However, effectiveness of systems engineering depends largely on human aspects, such as competence and behavior of individual stakeholders, social interaction between stakeholders, political circumstances, organization and governance, and many more. Research in systems engineering has to build on available scientific methods, both technical, as well as from the social sciences. Bhattacherjee wrote good introduction in social science research [8].

Figure 1 shows the logical order of steps to define a research project in systems engineering. The starting point is a need for improvement in the field, triggered by an industrial problem. Researchers reformulate the problem into an industrial goal; "where do we want to go" instead of "where are we". The problem triggers multiple research questions, e.g. how much can method x mitigate the problem? The next step is to sharpen the research by looking for quantifiable propositions, e.g. full requirements traceability will reduce the change request rate after the project definition with 80%. Researchers can use the research questions and the quantified propositions to formulate a hypothesis as basis for evaluation. Often a number of explicit criteria help in such evaluation. Preferably, the research leaves room to study multiple options. The researcher will need some baseline to evaluate. The baseline can be past performance, benchmarking with other projects or organizations, or comparison of different solutions.

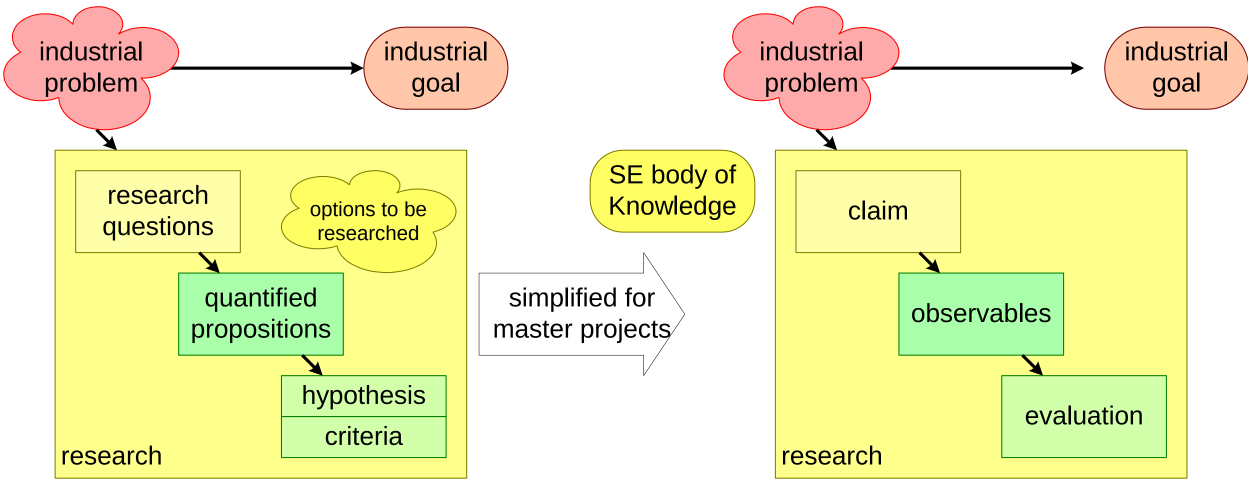

Figure 1. From industrial problem to validated research; left the generic approach, right a simplified approach used for master projects

We simplified the research model shown in Figure 1 left for research done in master projects of six months duration. The claim and observables form the core of the master project research. The claim relates to the contribution of the method or technique from the body of knowledge. The observables focus on supporting or invalidating this part of the body of knowledge. Thinking in terms of claim and observables is useful for other research too.

Figure 2 shows a context diagram of the research and many entities that are relevant for the research. If we study how effective a concept selection technique is, then we have to observe and describe the people using them, the process in which the concept selection is applied, the stakeholders and their concerns and objectives, the artifacts used, and the concepts themselves. Researchers have to strive for understanding the causes of success or failure. Is it the technique, the embedding process, the people, or some other factor the cause?

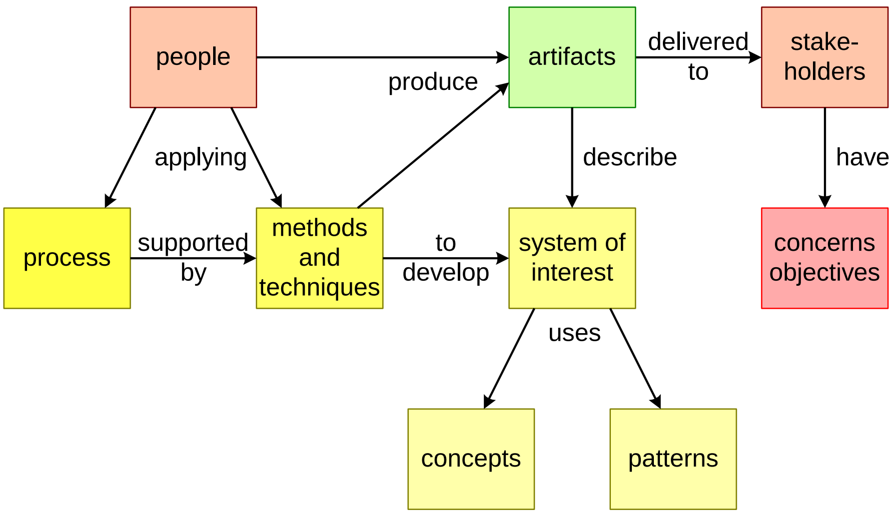

Figure 2. Context diagram with entities relevant for the research.

2. The local research situation in Kongsberg

Buskerud University College has been building a systems engineering department since 2006 in close cooperation with local industry. The author has described this process in earlier papers [9, 10]. Master students have done a major part of the research until now, during their half-year (full-time) master project. Gradually, we start to recruit PhD and other researchers. The results of this research are promising; we have published ten papers at conferences and journals (see http://www.gaudisite.nl/MasterProjectPapers.html). We evaluate the master project process every year, among others by sending the students a questionnaire. The academic staff uses the outcome of such survey to determine improvements for the next year. Last year, we added a workshop “Academic writing” as result of the feedback. This year, we add a presentation about research methods to the set of preparation workshops. We base this presentation on our experiences of the last few years, which form the basis of this paper too.

The limited duration of the master projects triggered us to simplify the research approach as shown at the right hand side of Figure 1. The researcher looks in the systems engineering body of knowledge for methods or techniques that could mitigate the industrial problem. The master student captures the expected improvement in a claim, similar to the quantified proposition in the generic approach. The next step is that the student identifies observables that can support or invalidate the claim. Students can then evaluate the claim by analyzing the observables.

3. Spectrum of research methods

Researchers observe or measure and will create artifacts as part of their research. In the hard sciences we are used to formalisms, often math-based, that support our research. However, the challenge in systems engineering research is that problems are often a mix of hard engineering and “soft” human factors. The softer factors affect the research approach significantly [11]. We have to learn from human sciences how we can research in this human context with psychological, social, political, and cultural aspects. One of the challenges is to collect data in such way that researchers can use it for analysis.

Figure 3 shows a spectrum of possibilities on how the researcher can collect data, can extract data from other people, and what artifacts the researcher can produce. This spectrum runs from free format at the left-hand side to standardized format at the right-hand side. A free format uses a free representation without a formal definition. Such free format works well in discovery and exploratory phases, when researchers know so little that they need flexibility to explore. Main disadvantage of free formats is that analysis is difficult, due to a lack of definitions. Comparison and aggregation of data is difficult for the same reason. The formal-informal dimension does not directly match the physical versus social science dichotomy. Physical sciences are used to (informal) logbooks, while social sciences use formalized data collection.

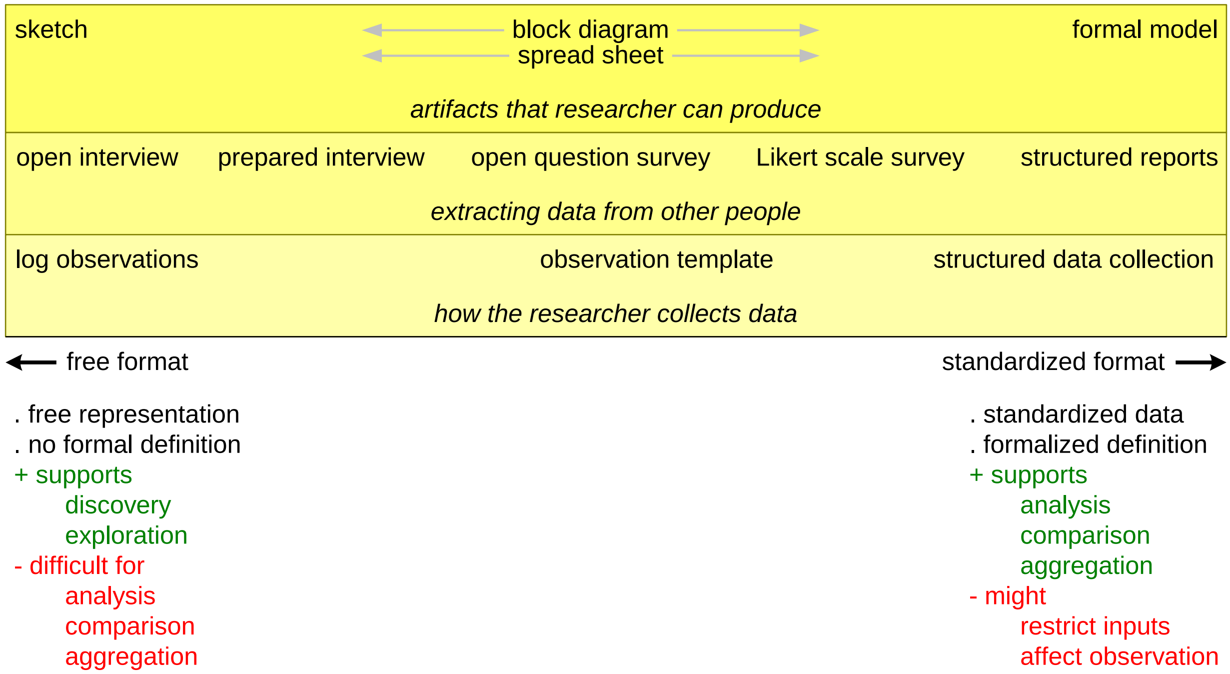

Figure 3. Spectrum of research methods from informal, free format, to formal with a standardized format.

Standardized formats standardize the format of the data and use formal definitions. Such standardization makes it possible to compare and aggregate data and more in general to analyze the data. However, the formalization affects our observations and restricts our inputs. Researchers might lose or ignore any event that does not fit our formalized model, or it might distort it to fit it in our formalism. Figure 3 shows on purpose a spectrum, since the soft context is not black and white; we might formalize part of the data, or we might use a more flexible formalism.

3.1. Collecting data

3.1.1. Logbook

The most fundamental data collection for scientists is the research logbook. In past centuries, scientists always carried a physical notebook and pencil. They were taking note after note. Such note taking requires discipline. In modern times, the logbook can be maintained electronically, for instance as Word or PowerPoint files. Entries always need a date/time stamp and then the *what*, *how*, *why*, *when*, *where*, and *who* key words guide the rest of the entry. Researchers should log any references to data, documents, or publications in electronic format, e.g. an URL. We recommend researchers to make a copy of these inputs too; today's storage media facilitate a practically unlimited logbook. Finally yet importantly, all "raw" data, such as submitted questionnaires, and intermediate data, such as spreadsheets, must be stored. The prime characteristic of a logbook is that it is "append" only by nature; never delete any data from the logbook! As consequence, all entries, including raw and intermediate data need a clear labeling with version and date. For larger research projects, configuration management is a serious issue. The logbook as described here is at the free format end of the spectrum, although more standardized information may be entered as well.

In the action research as we promote for systems engineering, the logbook will contain many confidential or company-sensitive data. Researchers and supervisors have to define upfront how to cope with confidentiality. At Buskerud University College, we have agreed that logbooks are confidential, and that supervisors have access under the condition that they respect the confidentiality.

3.1.2. Observation template

Session attributes – date (year/month/day)		
Kind of session:		Communicate information/status
		Sell a idea/concept
		Brainstorming/generate ideas
		Decision making
		Solve/discuss problem(s)/issue(s)
		Planning
		KPI/Performance/Action log
		Team building/training
		Presentation
Physical location of session:		Defined meeting room
		Colleague own office
		In the factory – "on the shop floor"
Planned session or not:		Planned
		Unplanned
A3 purpose:		
A3 name/link:		
A3 usage/iteration number:		
A3 usage time with stakeholders:		
Number of participants:		
Did everyone understand the A3:		
Did it answer some of the stakeholders questions:		
Create any new questions/concerns:		
Models changed/added:		
Stakeholder participation:		
Prefer A3 instead of A4:		
Observations/recordings:		

from Master Project by Espen Polanscak

Figure 4. Example of an observation template (courtesy Espen Polanscak).

A first step of standardization is to standardize the elements to be observed. Figure 4 shows an example of an observation template for meetings by Polanscak. He studied the effectiveness of using A3s to assist in communication in a factory of airplane engine components [12]. He designed this template to support note taking during every session.

3.1.3. Structured data collection

Sometimes the intended improvement correlates with performance data that available in the company. Examples are databases with project management data, problem or field reports, engineering data, regression testing, or development repositories. The researcher has to "design" the data collection to facilitate further research. Main challenge is to collect background information to interpret such data properly, and to understand its accuracy and possible constraints. For example, effort and time are often relevant components of a claim and hence the evaluation. Most employees have to register their activities in a time reporting system. However, the classification as used for time reporting is not always usable for the research.

3.2. Extracting data from other people

3.2.1. Interviews

Interviewing is a powerful way to get information from other people. The benefit of an interview is that it is interactive. Researchers can respond on answers from interviewees, e.g. ask for clarification, or enter a new direction. Interviews can be completely free format, where researchers are entirely perceptive, open for every piece of information. Alternatively, researchers can prepare themselves for the interview by writing down questions or key words. More preparation tends to correspond with moving in the direction of standardized format on the spectrum axis.

3.2.2. Surveys

Surveys are typically performed as one-way communication. The survey is sent to the recipient, who answers the survey without the possibility of interaction; clarification of intent is not possible. Surveys can use open questions (more to the left in Figure 3) or multiple choice answers (more to the right in Figure 3). Researchers often use the Likert scale [13], which is a so-called ordinal scale [14]. Figure 5 shows an example of a question with a 5 point scale with possible answers. Below the question, a possible presentation of accumulated answers is shown.

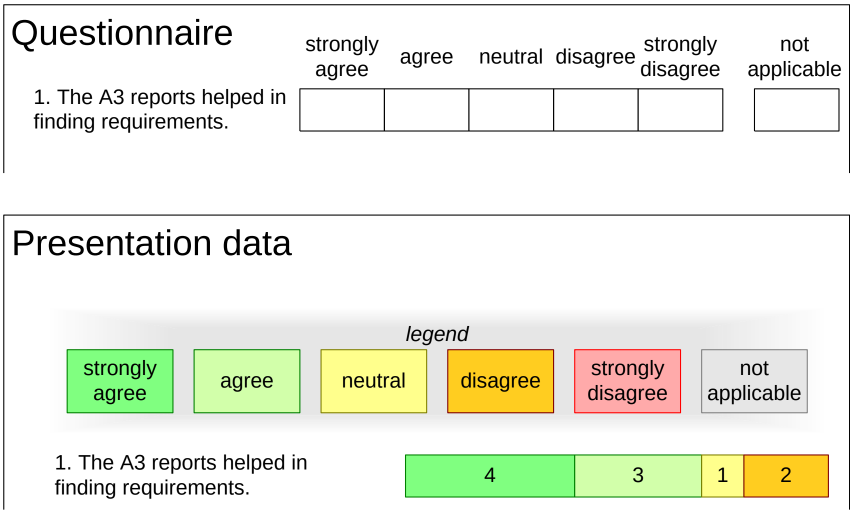

Figure 5. Survey using a Likert scale; at the top the questionnaire, at the bottom a possible presentation of results.

Benefit of a Likert scale is that researchers can accumulate and compare answers of respondents. When researchers use the same questions and scale, then they can even do this across research projects. However, in further analysis of the data we need to be careful with our interpretation of the results. There are many discussions in literature about the validity of interpreting this type of data, see for instance [15].

Researchers typically will map the scale on a numeric scale, as shown in Figure 6; e.g. fully disagree = 1, and fully agree = 5. The researchers can apply numeric analysis after such mapping. However, we do not know if respondents have perceived the scale in such linear fashion. Assuming a linear scale we can determine *mean* and *variance*, *median*, and *mode*. Unfortunately, the resulting values are still not easily interpretable. When can we conclude that respondents are satisfied: mean > 3 (e.g. better than neutral)? Or do we need at least a mean of 4 (better than agree?).

An alternate way to analyze such data is to use the Net Promotor Score (NPS), (see [16]). In NPS, we assume that respondents who provide an answer “strongly agree” will be promotors of our idea, while respondents who score “neutral”, “disagree”, or “fully disagree”, probably will complain to others about the idea. The overall score is number of promotors minus number of complaining respondents; if this number is larger than zero, then we have a positive outcome. There is again a lot of literature about NPS and its applicability; see, for instance, [17]. The net effect of using NPS is that “good” results, e.g. results that indicate appreciation of respondents, require relatively high scores, e.g. a significant number of “fully agrees”. Interestingly, a mean of 3 corresponds with a negative NPS; in other words, a slightly positive set of answers is not a good result. Many researchers intuitively start using 3 as threshold value, which is by far not good enough from NPS perspective.

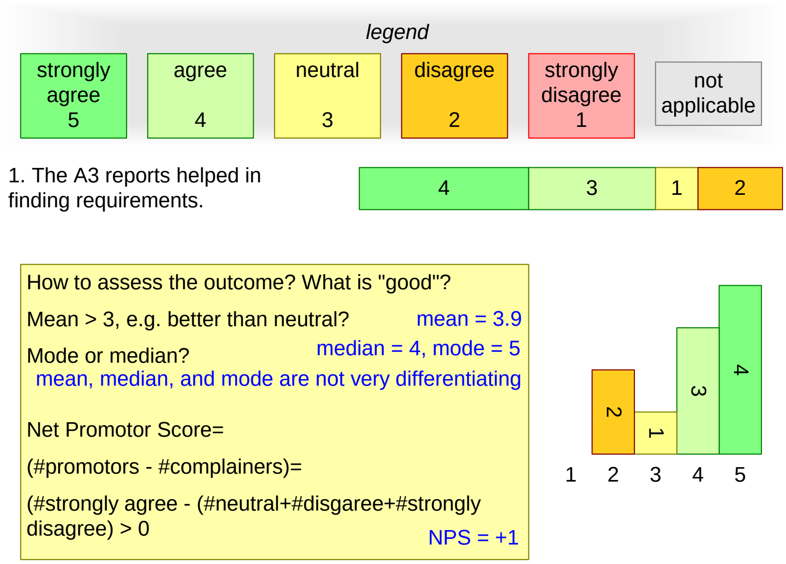

Figure 6. Evaluation of a Likert scale survey can be more complicated than researchers realize.

3.2.3. Structured reports

Researchers can institutionalize data collection by asking employees to report data regularly in structured reports. For example, the researcher can ask programmers to quantify effort and productivity figures every day. Such structured reports provide a wealth of data that allows, for example, trend analysis and performance analysis. Disadvantages are that data collection becomes more intrusive and costs more effort. Finally, researchers should be aware that their data collection might affect behavior or worse that people fill in data “as desired” rather than “as is”.

3.3. Producing artifacts

Researchers can validate the claim by producing artifacts, such as prototype, proof-of-concept implementations, models, or simulations. These artifacts can entail any set of entities in Figure 2. For example, if the claim is that a certain type of filter improves navigation, then modeling of navigation and the filters will provide evidence about the claimed performance improvement. If the claim is that a certain process will result in better organizational performance, then a model of the organization will provide insight in the relation between process and performance of the organization.

Researchers will produce, as part of the action research, many artifacts in their role as developer. They will make sketches, block diagrams, spreadsheets, and models. These artifacts might be a consequence of the method or technique that the researcher studies. Researchers need to consider these artifacts and their quality, when evaluating the method or technique.

4. Validation challenges and pitfalls

The ultimate result of research is the (in)validation of the claim or hypothesis. Challenges in systems engineering are to formulate a proper claim, to measure or observe relevant data for evaluation, and to analyze and interpret properly. Unfortunately, there are many complicating factors in systems engineering research that complicate validation. For instance, the proper measures for success of systems engineering methods are often far in the future, e.g. life cycle cost, adaptability for changes during the life cycle, etc. Typically, what *can* be measured or observed is not directly what we *need* for validation. Consequently, we need analysis and reasoning to get from data to conclusion. Another complication is that measurements typically take place in a context that if full of “soft” factors, e.g. humans and their behavior. The context is typically uncontrolled; researchers do their work in the given dynamic circumstances. Research strategies with double blind control groups are not easy to achieve for most systems engineering methods and techniques. Given all these complications, researchers have to scrutinize every step in their process as part of the validation.

4.1. Quality and interpretation of data

Fundamental questions are accuracy and credibility of the data. However, in this type of research we have to worry also about calibration. Can researchers compare or aggregate data from different? Can measurements suffer from any bias, e.g., "very poor" is an unacceptable score? A related worry is classification. When various stakeholders register data (e.g. how they spend their working hours), then the decomposition and semantics behind the decomposition has to be compatible. E.g., some engineers may allocate their time as project management, while others may classify the same work as systems engineering.

4.2. Validity and applicability of literature

The scientific body of knowledge depends on references to previous work. Authors have to support statements in a publication either by references to literature or by observations. Question is whether the reference actually supports the statement. How can we know that the publication we refer to is dependable? In the Netherlands, this is currently a hot topic, after discovery of scientific fraud in different disciplines [18]. Another risk is that the reference restricts its validity to a particular domain of application. In systems engineering, we often "borrow" insights from other domain. Is our reference to literature from other domains appropriate?

4.3. Cogency of arguments and reasoning

Researchers analyze and use experiments and observations to form an evaluation. The argument often uses an inference step. The inference can be valid or invalid, or in less black and white terms, strong or weak. Tang [19] shows many examples of faulty reasoning. Tang mentions cognitive biases and illogical reasoning as root causes of such reasoning failures.

For systems engineering research in particular, the question is whether researchers can distinguish cause and effect. If we introduce a new method, and the development went well, can we then conclude that our method is instrumental for the success? Alternatively, are any of the following factors the cause of success: the researcher's presence, the team, the system-of-interest, the organization, the leader, etc.?

4.4. Assessment of the final result

We expect from an evaluation a judgment on the goodness of systems engineering methods and techniques. The question is, on what scale do we assess the goodness? Typically, we lack an absolute scale of goodness. One approach to achieve meaningful judgments is to compare alternatives, e.g. benchmarking. Such comparison is far from trivial in itself. Are the alternatives comparable? Do they address the same problem? Do we have a clear set of criteria for comparison? Are the criteria meaningful to judge goodness?

4.5. Generalization of the outcome

This type of research takes place in a specific context, e.g. a domain (defense, subsea, maritime, etc.), an organization, a process context, specific staffing, etc. Can researchers generalize conclusions based on research in this specific context?

4.6. Pitfall observed in practice

We observed a number of common pitfalls in our local research efforts:

- When respondents to a survey are enthusiastic then the conclusion is that the respondents perceive the method as positive. Such perception in itself does not yet mean that a method or technique is “good”; respondents appreciated it.
- When a model predicts success, then this is an indication that reality might be a success. It depends on how close to realistic such model is and how credible, whether there is a closer correlation between model and success in reality.
- The skills and capabilities of the researcher were root cause for a positive or negative outcome, rather than the applied methods and technique. In general, unraveling cause and effect is quite difficult.
- When the outcome of the research is disappointing, e.g. a method is less productive than expected, there is a tendency to “blame” a specific circumstance; sometimes researchers draw the conclusion that in other circumstances the method is more beneficial without actually researching this assumption.

5. Conclusions

Research in systems engineering starts with a clear articulation of the current situation and its perceived problems. Researchers can transform the problem into objectives. They can look for new ideas, or methods or techniques from the existing body of knowledge to improve the current situation. The academic challenge is to evaluate the effectiveness objectively. Researchers need a yardstick for evaluation. We propose that researchers formulate such yardstick by articulating a claim or hypothesis with criteria. The next challenge is to observe, measure, and collect data to evaluate fact-based. We have discussed a spectrum of elementary observation and data collection methods and techniques.

Current research in systems engineering is handicapped by the uniqueness of every field situation. It is close to impossible to make hard claims of success related to specific methods, since we cannot unravel the causes of success. We need to evolve gradually our research methods to “harden” our research results, and we will need many of such results to achieve conclusions that are more generic.

Acknowledgements

Discussions with Mike Pennotti about evaluation of surveys helped to shape the survey evaluation part. The interaction with master project students in the past helped to order the methods.

References

1. I. Dixit and R. Valerdi Challenges in the Development of Systems Engineering as a Profession, to be submitted to Journal of Systems Engineering, preprint 2012.

2. A. Squires, A. Pyster, D. Olwell, S. Few, and D. Gelosh. 2009. Advance Systems Engineering.” INSIGHT 12 (4): 69–70.

3. B. Boehm, Valerdi, R., and Honour, E., The ROI of Systems Engineering: Some Quantitative Results for Software Intensive Systems, Journal of Systems engineering, Vol. 11, p221-234, 2008

4. G. Muller, Eirik Hole (editors) The State-of-Practice of Systems Architecting: Where Are We Heading? White Paper Resulting from Architecture Forum Meeting , 2005 (Helsinki, Finland), http://www.architectingforum.org/whitepapers/SAF_WhitePaper_2005_1.pdf

5. R. O'Brien, 2001. Um exame da abordagem metodológica da pesquisa. In Roberto Richardson (Ed.), *Teoria e Prática da Pesquisa Ação [Theory and Practice of Action Research].* João Pessoa, Brazil: Universidade Federal da Paraíba; accessed in the English translation.An Overview of the Methodological Approach of Action Research, University of Toronto http://www.web.ca/robrien/papers/arfinal.html#_edn2, accessed July 25, 2012.

6. C. Potts, Software-engingeering research revisited. IEEE Software, Vol. 10, No.5:19–28, September/October 1993.

7. G. Muller and Heemels W. P. M., Five Years of Multi-Disciplinary Academic and Industrial Research: Lessons Learned; CSER 2007 in Hoboken NJ

8. A. Bhattacherjee, Social Science Research: Principles, Methods, and Practices, USF Open Access Textbooks Collection, 2012, http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks, retrieved 2 Decemebr 2012

9. G. Muller, Validation of Systems Engineering Methods and Techniques in Industry; CSER 2012 in St Louis

10. G. Muller, Creating an Academic Systems Engineering Group; Bootstrapping Systems Engineering Research; CSER 2009 in Loughborough

11. R. Valerdi, S. Brown, and G. Muller, Towards a framework of research methodology choices in Systems Engineering; CSER 2010 in Hoboken NJ

12. E.L. Polanscak and G. Muller, Supporting Product Development: A3-assisted Communication and Documentation, master project paper at Buskerud University College, 2011.

13. R. Likert, (1932). A Technique for the Measurement of Attitudes. Archives of Psychology 140: 1-55

14. S. S. Stevens, (1946). On the Theory of Scales of Measurement. Science 103 (2684): 677-680

15. S. Jamieson, Likert scales: how to (ab)use them. Medical Education, (2004). Retrieved 30 July 2012, http://xa.yimg.com/kq/groups/18751725/128169439/name/1LikertScales.pdf

16. F. Reichheld, The One Number You Need to Grow, Harvard Business Review, 2003, http://hbr.org/2003/12/the-one-number-you-need-to-grow/ar/1

17. T. Keiningham, L. Aksoy, L. Cooil, B. Andreassen, T, W. (2008). Net Promoter, Recommendations, and Business Performance: A Clarification on Morgan and Rego. Marketing Science. Vanderbilt University. http://www2.owen.vanderbilt.edu/bruce.cooil/Documents/Publications/2008--Marketing%20Science.pdf

18. F. van Kolfschooten, Ontspoorde wetenschap over fraude, plagiaat en academische mores, De Kring, 2012 (translation: Derailed science about fraud, plagiarism and academic mores)

19 A. Tang, Software Designers, Are You Biased? SHARK'11, ACM 978-1-4503-0596-9/11/05