Empirical Research in Software Engineering: A Literature Review

Petr Pícha, Přemysl Brada
Department of Computer Science and Engineering
University of West Bohemia
Pilsen, Czech Republic
{ppicha, brada}@kiv.zcu.cz

***Abstract*-Software engineering (SE) and the empirical research in this area have both become a large fields with significant bodies of knowledge as well as methodical problems. Mentions, descriptions and examples of these problems are spread throughout the literature, but only a handful of suggestions and attempts of their rectification are being presented. This paper summarizes the goals and history of the SE field and focuses on the empirical research area within it. It highlights the most frequent problems affecting empirical SE research efforts and their most promising suggested solutions. Both the problems and suggested solutions were collected from a carefully selected sample of research publications. The presented overview should serve as a starting point for researchers or other professionals trying to get the first broad and shallow insight into the context of SE empirical research, as well as a theoretical basis for subsequent research.**

Keywords-empirical research, literature review, software engineering, history, problems, solutions

I. INTRODUCTION

The field of Software Engineering (SE) has been developing since the 1960s, and has, therefore, amassed a substantial amount of knowledge. Even empirical SE research, which specializes in studies based on experiments and observation, is very rich and obtaining the full grasp of its topics and findings poses a significant challenge.

The goal of this work was to perform a broad and shallow study of literature concerning the current state of Empirical Research in Software Engineering (ERSE). The motivation for this study was to obtain an initial understanding of the area as a prerequisite for further, deeper research on its particularities, details and various research methods and approaches.

The challenge of the study was obviously the extent of the literature body obtained. Therefore, this paper focuses on summarizing the findings about ERSE and the SE discipline in general. It highlights their context, goals, major problems in conducting empirical SE research and some suggested solutions.

The study will be useful mainly for researchers starting in the SE research field trying to gain a general overview of its content and context quickly and with minimal effort.

In section II, we briefly describe the motivation for and design of the study. Section III speaks about similar papers included in the studied literature sample and the main differences between this them and this paper. Section IV presents facts and findings about ERSE and SE in terms of their goals and history. Section V describes the major problems of the field found in literature while Section VI mentions the proposed solutions to these problems. Finally, Section VII discusses the accuracy and relevance of this study. Section VIII summarizes the principal findings and explains their proposed usefulness for other researchers and other readers.

II. STUDY CONTEXT AND DESIGN

This section explains the main reasons for creating this literature review and its goals as well as a process of its execution.

A. *Reasons for the Study*

The main motivation for this study was to create a pivotal resource and knowledge basis for ERSE within the local research group (Reliable Software Architecture - ReliSA) at the University of West Bohemia. To establish a baseline before any particular studies could be conducted we needed to obtain an overview of the ERSE field itself in a broad, but not necessarily too detailed manner.

In particular, we wanted to find out key information about the following aspects of ERSE: its context; methodologies and taxonomies used; leading experts, organizations and outlets; major SE problems studied in the research community and pitfalls of the research approaches both in general and in their particular steps (including ways to avoid or mitigate them). The literature review fitted all these goals as the best way to achieve them.

B. *Process of Study Execution*

The design of the study (meaning the process of literature gathering and exploration in order to extract the key information sought) breaks down to several steps.

At first, the main sources of material were selected. These were the academically oriented part of the search engine Google - Google Scholar, and digital libraries (DL) of scientific publications, such as ACM DL, IEEE DL and Springer Link. The key expressions, such as “empirical research”, “software engineering experiment”, or “research in software engineering” were searched in these sources to get a preliminary set of publications and organizations.

 ISBN: 978-1-61208-367-4

Then, the websites of the organizations that occurred most frequently and seemed most promising were explored to widen the set of publications. The relevance of the publications was determined mainly by reading their abstracts.

The whole set of selected publications was sorted using a simple citation metric which takes the sum of the publication's citation counts on Google Scholar and Web Of Science and divides this number by the age of the publication in years plus one (so that the papers published in the current year do not end up with count divided by zero).

The next step was the actual reading of the materials. Mainly, only the papers were read; the books were omitted for reasons of the exceedingly long time required to read them.

After reading each paper, the referenced publications were checked for relevance and, if found useful, added to the set of useful publications with their citation metric value calculated.

Whenever a significantly large subset of newly added papers originated from one organization previously unexplored, the website of the organization was checked and relevant papers were added as before.

The whole process stopped at the point where we had a list of the top six reachable papers which remained unchanged after studying all relevant referenced publications, i.e., when the process stopped uncovering new significant papers.

C. *The Analyzed Literature Sample*

The end result of the process described in the previous subsection was a set of almost 90 collected publications, with a subset of 53 publications found both available and useful. The publication years of this subset range from 1986 to 2014.

In terms of outlet, in which these publications were published, most of them (approximately 72%) came from journals, magazines, newsletters and other periodic publications. Most significant contributions amongst them made Information and Software Technology (Elsevier), Transactions on Software Engineering (IEEE), Empirical Software Engineering (Springer), and Software (IEEE). Almost 21% of all the publication studies came from conferences, symposiums and meetings with International Conference on Software Engineering (ICSE), International Symposium on Empirical Software Engineering and Measurement (ESEM) and The Future of Software Engineering Symposium (FOSE) leading the list. The rest of the publications (roughly 8%) originated elsewhere, e.g., were a technical report or an available chapter of a book.

On the level of the organizations from where the publications originated in the majority of cases (approximately 70%) at least one author was affiliated with Simula Research Laboratory (Lysaker, Norway), Keele University (UK), University of Maryland (USA), NICTA (National Information Communications Technology Australia – University of New South Wales, Sydney), Lund University (Sweden), or SINTEF (The Foundation for Scientific and Industrial Research, Trondheim, Norway). Other organizations included Microsoft's ESE (Empirical Software Engineering Group), Karlsruhe Institute of Technology (Germany), Chalmers University of Technology (Gothenburg, Sweden), Carnegie Mellon University (Pittsburgh, USA) and other institutions from USA, Ireland, Canada, Norway, UK and France.

III. Related Work and Similar Studies

This study certainly does not constitute the first attempt in the research community to assess the state and context of the SE field and ERSE area. This section mentions the similarly oriented studies closest to ours. It also highlights the differences and individual properties of our paper in relation to them.

Glass et al. [1] discuss the history of SE research and then try to categorize 369 selected papers from 1995 to 1999 on several levels. Perry et al. [2] try to assess the contemporary state and future (at the time) challenges of empirical studies in SE followed by describing the recommended structure of an individual study and concrete steps for assuring its quality. Sjöberg et al. [3] give the description of primary and secondary research methods followed by a vision for increased research quality for the future and suggestions leading to its fulfilling. Victor Basili published his study overview of the field in [4] including the nature of the discipline, research paradigms, vision and its attainability, goals of studies, types of studies and a description of a maturing process of the research.

None of the mentioned (or other studied) publications deals strictly with summarizing the general context of SE and ERSE in terms of their goals, history, major problems and solutions. The main contribution of this particular paper lies in summarizing the problems and solutions on the level of the field and individual studies (not the concrete phases and steps while conducting research). In addition, the literature sample used in our study spreads through almost three decades. That makes its findings not only contemporary, but also a broad overview, since none of the above mentioned works describes all the problems and solutions included in this paper.

IV. Context of Software Engineering Research

This section presents findings about SE, SE research and ERSE in terms of goals and history discovered throughout the studied literature.

A. *Goals of Research*

As in other scientific disciplines, the main goal of SE research is to contribute to the effectivity and provide knowledge for decision making in its respective field both in further research and in practice. However, the outcomes of the research can be used by several different types of audiences, such as common readers, other researchers, reviewers, meta-analytics, different kinds of committees and practitioners. While the resulting papers should ideally bring some level of benefit to them all, the authors may have to decide beforehand who is their target audience and how the study should benefit them and adjust the study and its goals properly.

Of course, there is the matter of subsequent research. The occurrence of one paper which brings great benefit to the field as a whole on some 10 pages is hardly imaginable. So the research in particular organizations (or several cooperating organizations) should be carefully designed and structured to support the preliminary vision and high-level goals.

Much has been written about how the individual studies should be done and research papers structured [2][5][6][7][8][9][10]. Most of the findings in this area will be summed up in a future work; here were we would like to present just one of the more general ideas suggested in [3], which claims that the point of every study is (or at least should be) the exploration or description of the relations between four “archetype classes”. These classes are: Actor, Technology, Activity and Software System. In this context, Actor can be not only a person or a group of people but also another SW or HW system and so on. Furthermore, Technology class representative need not to be a piece of HW or SW exclusively. It can also refer to a technique, method, practice, diagram, model, guideline, etc. This seems to be a pretty straightforward, simple and achievable principle.

B. *Software engineering history*

Now, let us talk a bit about SE and ERSE history. The most useful, yet general and complex description of history was found in [1], which is the main source of the information presented here.

The existence of SE can be traced back to the early 50s. The SE research (including ERSE) appeared almost simultaneously, although it was mainly ad hoc research and unfortunately, though much has been researched and discovered, there were very few outlets to present the findings in. This lead to the fact, that many of the findings from this era were published roughly 10 years afterwards.

And so the real traceable history of SE research dates back to the late 60s when the first SE conferences were being held. However, back in that time the SE was still just one part of the much larger Computer Science field, and did not begin to separate its presence in academia off from it until the early 80s.

Of course, that was not the end of the genesis of the field that continues up until this day. For example, qualitative research (e.g., [11]), which is aimed more on how things work (especially important in SE, since the impact of the human factor on the field is very significant) than on causal links and numbers, did not really appear in noticeable measure until 10 to 20 years ago and still does not have very large volume compared to standard and more “technical” quantitative research. (There is a simple and useful differentiation between qualitative and quantitative research methods in [7].)

V. PROBLEMS IN SOFTWARE ENGINEERING RESEARCH

The studied literature identifies and describes many problems of SE research impacting both the field in general (presented in this section), or just one or more steps in particular studies (described in a future work). Many of these problems appear in more than one publication and this section summarizes and presents the ones that seem the most significant and frequently encountered.

A. *The Whole Field*

At the top of the list is probably the **absence of** any **unified** and universally recognized **guidelines, methodologies, taxonomy and** even **terminology** which could be used and followed while conducting studies and presenting the results in research papers. Some authors [2][5][6][7][12][13][14] try to come up with first suggestions or their own models for research or its evaluation, but no field-wide agreement has been reached so far. The result of this is an inconsistency of studies and papers throughout the research community, which leads to sort of isolation of the studies, that cannot be compared, widened, summarized by meta-analysis or theory building, followed up by other studies, or even properly disproved.

A related problem is often mentioned **little support in recognized authority** which could come up with these so needed guidelines and methodologies, or at least support their development and mediate a discussion on them, or even make them a standard in the end. Examples from other fields given in literature include the *Cochrane Collaboration* and their *Cochrane Database of Systematic Reviews* (Australia; mentioned in [15]) in medicine, or the *Human Genome Project* (mentioned in [3]) in genetics. Some sort of comfort or hope lies in *SWEBOK* which describes issues related to SE but still does not include guidelines for research.

Another problem is that **not enough research** in general. According to some (e.g., [3], [16]), too few studies are being conducted, presented and published in the SE and ERSE field. This is of course a matter of opinion, but the truth is the more studies undertaken the better. Even if their quality (further discussed later in this paper) is in general poor, the total number of good papers is higher in bigger overall amount than in smaller. There are some subareas where the current situation is even worse.

One is **qualitative research** (already mentioned in Section IV), which **is** somewhat **rare** in itself in the SE field. Although the situation started to get better in recent years, as finally more and more qualitative research (described for example in [11] or [17]) is done, the pace is not fast enough.

The other is **theory building** [3][18][19][20][21], which **suffers** from subsequent issues from incorrect conclusions (mentioned in [5]), ignoring negative results (mentioned in [22]), research questions not being insightful enough (mentioned in [2]), misused statistical methods (described, e.g., in [23], [24]) and poor quality of studies in general (many of these problems are mentioned in following subsection B) which, of course, makes it difficult to prove hypotheses and build theories, which are necessary for establishing some ground on which to build further activities in the field.

The next problem often discussed and mentioned is the **gap between practice** of the field **and academia** (or the research community in general). This results in research not covering the real needs of the industry and in return not applying the findings of the research in practice. This has an immediate link to the next issue (next paragraph), but another statement found in [1] linked to this talks about so-called "**assimilation gap**", which refers to the time frame in between the first acquisition of a new technology (tool, technique, method, practice, model, etc.) and its 25% penetration into software development organizations in practice, and according to [1], is nowadays 9 to 15 years wide.

A large amount of **research is purposeless** as mentioned, e.g., in [2]), meaning its being done simply for something to be done without the aims and goals of the research corresponding to or reflecting on the needs of the practice. The cause of this, apart from the above mentioned gap between industry and academia, is the fact that often a major evaluation criterion of researchers (including the PhD students) is the quantity of published papers and citations. And so, if the researchers have no experience or relationship with the actual practice in the respective field, they end up doing research just for its own sake.

The last big issue of the field as a whole is the **lack of resources**, both financial and personal. The ERSE still struggles to get sufficient support from governments, industry or other entities, such as different kinds of initiatives and foundations. The comparison with other fields, especially medicine, is often given in the literature studied [3].

This concludes the problems of the field in general. Other listed problems refer more or less to the issues that the individual studies and papers suffer from.

B. *Individual Studies*

The main problem concerning individual studies is **low quality** (see, e.g., [25])**, relevance** (meaning significance of results) **and usability** (also called *impact* – factor of research being interesting for the field; [8]). Many studies suffer from **misusage of statistical methods**, which are often poorly understood by IT practitioners [23]. They also often just present results or state conclusions with **improper or no validation and evaluation** whatsoever, as mentioned in [8][12][15][16][26][27].

Another contributing factor is **incorrect conclusions drawing** (see, e.g., [5])**, fishing for results** (presenting insignificant results so that there is something to publish) and mostly **ignoring negative results** (as mentioned, e.g., in [22]). Although not positive (in terms of proving the hypothesis or showing something is wrong with the studied phenomenon), such results still need to be published or made public at the very least to prevent other researchers from repeating the same thing once it has been proven meaningless. Furthermore, these results can still be followed by validation studies, replication studies (see, e.g., [28]), qualitative studies shedding light on what went wrong, etc.

The follow-up studies are often difficult to conduct for one more reason and that is **publishing** study results **without the** input **data** used. This, of course, has its source in the fact that the data are mostly industrial by origin and as such considered sensitive and in need of keeping them confidential, which is unfortunately something SE researchers can hardly influence or overcome (as mentioned, e.g., in [29]).

Last but not least, in reference to the above mentioned gap between industry and academia (subsection A), the researchers are often given no choice but to use **students as subjects in** their **studies**. The reasons are obvious: students are easily available and their usage is cheap [9][22][30]. But, of course, practitioners tend to dismiss the results of such studies as irrelevant or not representative of reality instead of considering the results and if found interesting trying to supply replication study with their own people.

VI. Suggested Solutions to the Problems

This section deals with solutions to the problems in Section V, suggested in the studied literature.

As many publications compare the state of SE and ERSE to other disciplines with longer history and more stable infrastructure, the first suggested course of action is **patterning** our field **after,** such **other disciplines** (e.g., [26][31]) like medicine, psychology and sociology (both for their overlap with IT in SE and long tradition of qualitative research), information systems, or computer science. This could potentially bring some "order" to the "chaos" and establish a platform for a developing more field-specific empirical research framework.

Another suggestion was to **include practitioners and statisticians** into the research activities (as mentioned in [5][9]). This would significantly improve the situation concerning the gap between academia and industry, purposeless research, misuse of statistical methods, validation and evaluation of the results. The problem here is of course the above mentioned resources issue.

The resources could be brought by initiatives and organizations (see, e.g., [9]) established for these purposes and also for the purposes of stabilizing the terminology, taxonomy, methodology and guidelines, as well as theory building and authority foundation. Of course, to **establish** such **authorities and initiatives**, community-wide cooperation towards these goals is essential, but once done this should have major impact on research quality and effectivity, thus fulfilling the purpose of research in the first place (section IV). A way to collaborate on the community level through social networking is described in [32].

But before that can be done, the lower-level cooperation needs to appear and grow. Research organizations such as universities, institutes and even companies should strive to **build relationships** on a common goal of research and improvement [2][3]. Furthermore, each research entity should have some degree of **long-term** focus on a particular **topic** (or topics, depending on size and manpower), and conduct not isolated studies, but whole **families of studies** surrounding the topic [2][28]. The similar topic focus among several entities should supply the common ground for the above mentioned cooperation.

 ISBN: 978-1-61208-367-4

In terms of individual studies one more issue has to be mentioned and that is the **insightfulness of** research **questions** (see, e.g., [2]). The importance of this is that if you do not have research questions or hypotheses insightful enough, you undermine your studies from the very beginning. And although you may still finish the studies and present your results, such studies are hard to follow up on, replicate or include in meta-analysis (see, e.g., [33][34]), rendering them isolated, and therefore, useless.

The last main suggested solution looks to the future in a different way. Instead of trying to change the field, current researchers, practitioners and their mindset, the solution lies in the upbringing of good researchers from the very beginning – of course through **education**. The literature [2][35] suggested special SE research courses at colleges and universities to try to prepare students with research aspirations for the obstacles and circumstances of such a career and to teach them the basic knowledge on how to do research properly. This would, however, better be premeditated by establishing proper, unified and universally acknowledged methodologies and guidelines (mentioned before), or else every education entity will end up teaching their students something different and the resulting chaos will be little to none better than the current one.

VII. Discussion

Here, we discuss the usage and relevance of this work. As it is a simple literature review we do not claim that the information presented is fully comprehensive and accurate. Much wider and deeper analysis would be needed for a statistically valid study.

Through the course of this paper, we often talk about SE research, while the topic should be mainly ERSE. This is caused by the broad and shallow nature of this study. In its high-level context, the separation of empirical research off from SE research in general is difficult and probably not feasible to some degree. Nevertheless, the findings apply to both areas in most cases and where not, it is impossible to describe the issues of ERSE without the wider context of SE research.

Lastly, this presented paper does not cover the full breadth of the studied material and findings. As future work, we plan to address the methodology and taxonomy aspects of ERSE, various types of studies, suggested steps in conducting and describing the studies in papers and of course basic techniques, principles, pitfalls, and frequent mistakes made in each of these steps.

VIII. Conclusions

This literature review summarized and described information and findings discovered throughout the studied publications concerning SE and ERSE in this field.

The studied literature sample consisted of 53 papers and articles published from 1986 to 2014, coming from major journals and conferences in the field and spread throughout the world in their origin. The key problems in the area of SE research, as found through this study, are overall insufficient quality, relevance and impact. The steps towards mitigation and avoidance of these problems are being taken throughout the respective community, although it is a long-term and slow process.

The general advice for improving the research quality includes rigorous study design and description (correct usage of statistical methods, drawing of conclusions, insightful research question, validation and evaluation), including statisticians and as much of practitioners participation (or at least industrial data) to the research as possible, education of researchers through specialized courses, establishing authorities and financial support systems, and using other scientific disciplines as a pattern.

Our hope is that this review helps especially other scientists starting in the ERSE field as a base of general knowledge and overall overview. In addition, the findings can be used as an information basis for both further and more detailed research and conducting individual empirical studies in the field of software engineering.

Acknowledgment

The work was supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems.

References

[1] R. L. Glass, I. Vessey, and V. Ramesh, “Research in Software Engineering: An Analysis of the Literature”, Information and Software Technology, vol. 44, issue 8, Jun. 2002, pp. 491-506, doi: 10.1016/S0950-5849(02)00049-6.

[2] D. E. Perry, A. A. Porter, and L. G. Votta, “Empirical Studies of Software Engineering: A Roadmap”, Proceedings of the Conference on The Future of Software Engineering (ICSE ‘00), ACM New York, Jun. 2000, pp. 345-355, ISBN: 1-58113-253-0, doi: 10.1145/336512.336586.

[3] D. I. K. Sjöberg, T. Dyba, and M. Jörgensen, “The Future of Empirical Methods in Software Engineering Research”, Future of Software Engineering (FOSE ‘07), IEEE, May 2007, pp. 358-378, ISBN: 0-7695-2829-5, doi: 10.1109/FOSE.2007.30.

[4] V. R. Basili, “The Role of Experimentation in Software Engineering: Past, Current and Future”, 18th International Conference on Software Engineering (ICSE ‘96), IEEE Computer Society, 1996, pp. 442-449, ISBN: 0-8186-7246-3.

[5] B. A. Kitchenham, S. L. Pfleeger, D. C. Hoaglin, K. E. Eman, and J. Rosenberg, “Preliminary Guidelines for Empirical Research in Software Engineering”, IEEE Transactions on Software Engineering, vol. 28, issue 8, Aug. 2002, pp. 721-734, doi: 10.1109/TSE.2002.1027796.

[6] M. Shaw, “What Makes Good Research in Software Engineering?”, International Journal on Software Tool For Technology Transfer, vol. 4, issue 1, pp. 1-7, Oct. 2002, doi: 10.1007/s10009-002-0083-4.

[7] C. Wohlin, M. Höst, and K. Henningsson, “Empirical Research Methods in Software Engineering”, Empirical Methods and Studies in Software Engineering, Springer Berlin Heidelberg, pp. 7-23, 2003, ISBN: 978-3-540-40672-3, doi: 10.1007/987-3-540-45143-3_2.

[8] T. Dyba, B. A. Kitchenham, and M. Jörgensen, “Evidence-based Software Engineering for Practitioners”, IEEE Software, vol. 22, issue 1, Jan./Feb. 2005, pp. 58-65, doi: 10.1109/MS.2005.6.

[9] S. S. Briliant and J. C. Knight, “Empirical Research in Software Engineering: A Workshop”, ACM SIGSOFT Software Engineering Notes (Newsletter), vol. 24, issue 3, May 1999, pp. 44-52, doi: 10.1145/311963.311998.

[10] B. A. Kitchenham, ”Procedures for Performing Systematic Reviews”, Joint Technical Report, July, 2004, ISSN: 1353-7776.

[11] C. B. Seaman, "Qualitative Methods in Empirical Studies of Software Engineering", IEEE Transactions on Software Engineering, vol. 25, issue 4, Jul./Aug. 1999, pp. 557-572, doi: 10.1109/32.799955.
[12] V. R. Basili and R. W. Selby, D. H. Hutchens, "Experimentation in software engineering", IEEE Transactions on Software Engineering, vol. 12, issue 7, Jul. 1986, pp. 733-743, doi: 10.1109/TSE.1986.6312975.
[13] B. A. Kitchenham et al., "Can We Evaluate the Quality of Software Engineering Experiment?", The Forth International Symposium on Empirical Software Engineering and Measurement (ESEM '10), ACM New York, Sep. 2010, article no. 2, ISBN: 978-1-4503-0039-1, doi: 10.1145/1852786.1852789.
[14] M. Höst and P. Runeson, "Checklists for Software Engineering Case Study Research", The First International Symposium on Empirical Software Engineering and Measurement (ESEM '07), IEEE, Sep. 2007, pp. 479-481, ISBN: 978-0-7695-2886-1, doi: 10.1109/ESEM.2007.46.
[15] B. A. Kitchenham, T. Dyba, and M. Jörgensen, "Evidence-based Software Engineering", The 26th International Conference on Software Engineering (ICSE '04), IEEE, May 2004, pp. 273-281, ISBN: 0-7695-2163-0, doi: 10.1109/ICSE.2004.1317449.
[16] W. F. Tichy, "Should Computer Scientists Experiment More?", IEEE Computer, vol. 31, 1997, pp. 32-40.
[17] M. Dixon-Woods, "How Can Systematic Reviews Incorporate Qualitative Research? A Critical Perspective", Qualitative Research, vol. 6, issue 1, Feb. 2006, pp. 27-44, doi: 10.1177/1468794106058867.
[18] M. Jörgensen and D. Sjöberg, "Generalization and Theory-Building in Software Engineering Research", The 26th International Conference on Software Engineering (ICSE '04) - Workshop "8th International Conference on Empirical Assessment in Software Engineering (EASE 2004)", May 2004, pp. 29-36, ISBN: 0-86341-435-4, doi: 10.1049/ic:20040396.
[19] A. Zendler, "A Preliminary Software Engineering Theory as Investigated by Published Experiments", Empirical Software Engineering, vol. 6, issue 2, Jun. 2001, pp. 161-180, doi: 10.1023/A:1011489321999.
[20] D. I. K. Sjöberg, T. Dyba, B. C. D. Anda, and J. E. Hannay, "Building Theories in Software Engineering", Guide to Advanced Empirical Software Engineering, Springer London, pp. 312-336, 2008, ISBN: 978-1-84800-043-8, doi: 10.1007/978-1-84800-044-5_12.
[21] J. E. Hannay, D. I. K. Sjöberg, and T. Dyba, "A Systematic Review of Theory Use in Software Engineering Experiments", IEEE Transactions on Software Engineering, vol. 33, issue 2, Feb. 2007, pp. 87-107, doi: 10.1109/TSE.2007.12.
[22] W. F. Tichy, "Hints for Reviewing Empirical Work in Software Engineering", Empirical Software Engineering, vol. 5, issue 4, Dec. 2000, pp. 309-312, doi: 10.1023/A:1009844119158.
[23] T. Dyba, V. B. Kampenes, and D. I. K. Sjöberg, "A Systematic Review of Statistical Power in Software Engineering Experiments", Information and Software Technology, vol. 48, issue 8, Aug. 2006, pp. 745-755, doi: 10.1016/j.infsof.2005.08.009.
[24] B. A. Kitchenham, D. R. Jefferey, and C. Connaughton, "Misleading Metrics and Unsound Analyses", IEEE Software, vol. 24, issue 2, Mar./Apr. 2007, pp. 73-78, doi: 10.1109/MS.2007.49.7.
[25] M. Jörgensen, T. Dyba, K. Liestöl, and D. I. K. Sjöberg, "Incorrect Results in Software Engineering Experiments: How to Improve Research Practices", unpublished.
[26] M. V. Zelkowitz and D. Wallace, "Experimental Validation in Software Engineering", Information and Software Technology, vol. 39, issue 11, 1997, pp. 735-743, doi: 10.1016/S0950-5849(97)00025-6.
[27] B. A. Kitchenham et al., "Evaluating guidelines for reporting empirical software engineering studies", Empirical Software Engineering, vol. 13, issue 1, Feb. 2008, pp. 97-121, doi: 10.1007/s10664-007-9053-5.
[28] V. R. Basili, F. Shull, and F. Lanubile, "Building Knowledge through Families of Experiments", IEEE Transactions on Software Engineering, vol. 25, issue 4, Jul./Aug. 1999, pp. 456-473, doi: 10.1109/32.799939.
[29] V. R. Basili, M. V. Zelkowitz, D. I. K. Sjöberg, P. Johnson, and A. J. Cowling, "Protocols in the Use of Empirical Software Engineering Artifacts", Empirical Software Engineering, vol. 12, issue 1, Feb. 2007, pp. 107-119, doi: 10.1007/s10664-006-9030-4.
[30] D. I. K. Sjöberg, B. Anda, and T. Dyba, "Conducting Realistic Experiments in Software Engineering", 2002 International Symposium on Empirical Software Engineering, IEEE, Oct. 2002, pp. 17-26, ISBN: 0-7695-1796-X, doi: 10.1109/ISESE.2002.1166921.
[31] D. Budgen et al., "Cross-domain Investigation of Empirical Practices", IET Software, vol. 3, issue 5, Oct. 2009, pp. 410-421, doi: 10.1049/iet-sen.2008.0106.
[32] A. Begel, J. Bosch, and M. A. Storey, "Bridging Software Communities through Social Networking", IEEE Software, vol. 30, issue 1, Jan. 2013, pp. 26-28, doi: 10.1109/MS.2013.3.
[33] J. Miller, "Applying Meta-analytical Procedures to Software Engineering Experiments", Journal of Systems and Software, vol. 54, issue 1, Sep. 2000, pp. 29-39, doi: 10.1016/S0164-1212(00)00024-8.
[34] L. M. Pickard, B. A. Kitchenham, and P. W. Jones, "Combining Empirical Results in Software Engineering", Information and Software Technology, vol. 40, issue 14, Dec. 1998, pp. 811-821, doi: 10.1016/S0950-5849(98)00101-3.
[35] M. Jörgensen, T. Dyba, and B. A. Kitchenham, "Teaching Evidence-Based Software Engineering to University Students", Proceedings of the 11th IEEE International Software Metrics Symposium (METRICS '05), IEEE Computer Society, Sep. 2005, pp. 24-31, doi: 10.1109/METRICS.2005.46.