

Addressing the Challenges of Replications of Surveys in Software Engineering Research

Aileen Cater-Steel
University of Southern
Queensland,
Toowoomba Australia
caterst@usq.edu.au

Mark Toleman
University of Southern
Queensland,
Toowoomba Australia
markt@usq.edu.au

Terry Rout
Griffith University,
Brisbane Australia
T.Rout@griffith.edu.au

Abstract

Surveys are a popular research tool often used in empirical software engineering studies. While researchers are urged to replicate existing surveys, such replication brings with it challenges. This paper presents a concrete example of a replication of a survey used to determine the extent of adoption of software development best practice. The study replicated a European survey which was adapted and administered in a different context of Australian software development organisations. As well as discussing problems encountered, this paper presents a set of recommendations formulated to overcome identified challenges. Implementation of the recommendations will strengthen the value and contribution of surveys to the body of knowledge of empirical software engineering research.

1. Introduction

It has been claimed recently that the standard of empirical software engineering (SE) research is poor and that research guidelines are needed to improve the research and reporting process [1]. There is growing awareness of the need to aggregate empirical results to overcome difficulties in interpreting empirical studies, and to avoid 'study graveyards' [2].

Independent replication lies at the heart of scientific methodology [3]. Since no single study can simultaneously control all extraneous variables, replications by independent investigators are the only guarantee that a phenomenon is robust [4]. Although SE researchers are urged to replicate experiments and surveys [5], there is little advice in the field of software engineering about potential difficulties in undertaking

such studies. According to Basili, the act of replication is also a research problem, and researchers can learn from replication too, in the form of new insights [6].

This paper commences with a discussion about the use of surveys in SE research citing examples of surveys which have made a significant contribution to the field. The need for surveys as a means to determine the state of practice in SE, is then established with a summary of best practice adoption research. The best practice survey conducted by the European Software Institute (ESI) is detailed and contrasted with the replication of the ESI survey in Queensland. The analysis of the Queensland survey provides illustrations of specific difficulties in replicating surveys and provides the basis for a set of recommendations to assist researchers overcome such challenges.

2. Surveys in SE research

Many SE researchers have made effective use of surveys, with the results presented at conferences and published in academic journals. A survey is defined as a 'research technique in which information is gathered from a sample of people by use of a questionnaire' [7]. Surveys may be administered in person, or distributed by post, e-mail, or via the Internet. Within the SE field, surveys are used for three main purposes: for market research (e.g. to determine the adoption of a CASE tool); to provide government agencies with information about the software industry for policy making; and, for research, often as part of higher education research qualifications. There is an overlap between the last two purposes, as a government agency may sponsor a research centre or student to conduct a survey on its behalf.

2.1. Prominent SE surveys

One of the most well-known surveys was that used to determine the capability maturity level of organizations. The Software Engineering Institute's (SEI) capability maturity questionnaire of 1987 [8], along with its 1994 revision [9], has had a major influence on software development organisations, not only throughout the US Defense contractor population, but also globally [10]. Although the SEI questionnaire was intended to be used to assess the capability of software development groups by assessors in a face-to-face interview setting, it has been mailed by researchers to organizations to report and compare capability maturity of developers in the US and Japan [11], Singapore [12], and Australia [13]. Another example of an often cited replicated survey also relates to the Capability Maturity Model (CMM). In 1995, a mail survey was conducted to explore the benefits of CMM [14] and parts of that questionnaire have been replicated to gain feedback about the benefits of ISO/IEC 15504 (SPICE) [15].

2.2. The role of surveys in determining current practice

In 1987, Fred Brooks observed that 'the gap between the best software engineering practice and the average practice is very wide—perhaps wider than in any other engineering discipline' [16]. The view that the *state of practice* lags far behind the *state of the art* is still widely held today [17, 18].

Recently, SE research has been criticised on two counts. Firstly, that researchers, while eagerly prescribing practices, are not aware of what practitioners are actually doing [19, 20]; and secondly, that many of the prescribed practices, methods, techniques and standards are not thoroughly evaluated [21, 22] and may not even be suitable for all software development organisations [23-26]. Surveys can help to redress these problems.

Paulk, Goldenson and White [27] recommend the use of surveys to gain a 'good feel for the breadth of deployment of specific techniques across industry'. Broad investigations on the actual state of software engineering are rarely done because 'collecting the information is a huge effort, while few people appreciate the results' [18]. Consequently, there are few empirical studies reported of actual software processes used [28, 29]. Prior to instigating process improvement programs,

it is imperative to have an understanding of current software development practices throughout the industry.

The need for an industry-wide survey was highlighted in Australia in 1997, in a government report about the future of the Information and Communication Technology (ICT) industry. The report recommended to 'support the collection and dissemination of improved industry statistics and undertake regular benchmarking' [30]. Surveys of the software development industry are essential to ensure policies relating to ICT are based on a factual representation of the industry.

A survey is considered to be a feasible means of providing data for any study investigating the state of practice [31]. The replicated survey analysed in this paper provided a broad industry-wide snapshot of the adoption of best practice techniques in use throughout software development groups in Queensland.

3. Software process best practice research

Although many authors refer to software developers using dominant, prevalent, or common practices, there has been little research to date to document actual current use. A survey conducted in 1992 found marked differences in the practices used by 102 European firms compared to 326 Japanese companies [32], but a different study of software management practices in US, Japan and Western Europe concluded that, globally, firms appear to be remarkably similar in the practices used [33]. Other researchers have focused on a particular location or group of practices, for example, a survey of adoption of software processes of 280 companies in Singapore [12]; large US projects [34]; comparison of software practices in the US and Japan [35]; quality practices in the UK [36]; requirements engineering in Canada [37]; comparison of Australian and US software development practices [38]; software metrics in Australia [39]; and project management practices in the US [40].

The most widely reported survey of best practice in Europe was that conducted by the ESI [29, 41]. This paper discusses the challenges encountered when the ESI survey was replicated in Queensland.

3.1. ESI best practice survey

In 1995, the European Commission (EC) launched the European Systems and Software Initiative (ESSI) program with the aim of motivating organisations to test and deploy software best practices. The ESSI

program was administered by the ESI as part of the EC's Information Technologies program [42]. Organisations were encouraged to apply for EC funding to enable them to adopt a specific software process improvement (SPI) project in a real-life commercial environment over a period not exceeding 18 months. The ESSI program included a longitudinal study of European software practices to assess and monitor the level to which European software developers were adopting best practices.

The ESI developed the Software Best Practices Questionnaire (SBPQ) to collect data for the ESSI program. Previous research in software process improvement, and popular models such as the CMM, Bootstrap and ISO 15504, influenced the development of the questionnaire. On three occasions (between 1995 and 1997), the questionnaire was distributed by the ESI as part of the call for proposals for ESSI funding. Respondents were explicitly informed that the questionnaire was independent of the funding proposal review process [42].

A best practice is defined as 'a management practice that is widely recognised as excellent and is recommended by most practitioners and experts in the field' [43]. The SBPQ represents the 'subjective consensual views of multiple experts' [44], and comprises a subset of core software development practices extracted from recognised SPI models.

The ESI survey instrument groups practices under five headings:

- *organisational issues*: project management, change control, training programmes for managers
- *standards and processes*: formal assessment of benefits and risks, management reviews, control of subcontractors, independent audits, coding standards, formal handovers, test planning
- *metrics*: records of actual and estimated resources, error sources, test efficiency, computer performance, project tracking
- *control of the development process*: accountability for estimates and schedules, requirements management, control of code and specification changes, regression testing
- *tools and technology*: use of design notations, automated testing tools, prototyping, data dictionary, project management tools.

The content of the questionnaire has been criticised on two counts by Dutta and Van Wassenhove [44]: firstly, it overlooks important issues related to organisational and customer-supplier management; and secondly, it does not including practices associated

with high maturity organisations (for example, CMM level four and five practices).

However, despite its shortcomings, the ESI study yielded valuable findings from the analysis of the 1,279 responses received over three years. There were 463 responses to the first survey in March 1995 from 17 countries and representing 33 industry sectors [41]. The second survey was conducted in mid 1996 and received 488 responses from 17 countries representing 34 industry sectors [41].

The third and final survey in 1997 generated 397 responses [43] and showed 'wide variation in both awareness and application of process improvement techniques' [42]. Overall, the respondents had adopted 51 percent of all practices. European organisations were urged to become more aware of best practices and process improvement techniques as the European software industry was lagging behind the US in terms of awareness and application of SPI [42].

Many of the organisations which responded to the SBPQ survey were successful in gaining EC funding to improve their software processes. Consequently, there has emerged a valuable body of single-case studies detailing the effectiveness of specific SPI initiatives in small and medium-sized European companies.

3.2. Queensland survey

To overcome constraints of time and cost, and to enable comparison of the adoption of best practice by Queensland organizations with their European counterparts, the ESI was approached for permission to replicate their questionnaire. The survey, which formed part of the Australian National Industry Improvement Program (NIIP), was supported by Software Engineering Australia (SEA) Queensland, and Griffith University's Software Quality Institute (SQI). The intention was to conduct the survey initially in Queensland, then later in other Australian states.

Practice adoption level and organisation adoption level are the dependent variables and care was taken to maintain consistency in terms of scales and items to ensure that comparisons with previous studies could be made. A mail survey was chosen as the data collection method to provide a quick, inexpensive, efficient, and accurate means of assessing information about the population [7].

The unit of analysis for the survey was any Queensland organisation undertaking software development. The target population included all organisations in Queensland which develop software for sale as well as in-house software development

groups within organizations, in effect, a census. Rather than identify a sample of the population, the aim was to reach the entire population of organisations which develop software.

A list of 5,600 likely organisations was compiled from various databases and contact lists from the Queensland Government's Information Industries Bureau (IIB), SEA (Qld), SQI, MIS 3000 [45] and Australia On Disk [46] databases.

The questionnaire was pre-tested using face-to-face interviews with two key informants, and further feedback was received from a SEA Board member. During the pre-test, concerns were raised about the section headings and question groupings of the ESI questionnaire (such as *metrics, standards and procedures, control of development process*). As the more technically complex questions may evoke feelings of anxiety or inadequacy, it was decided to group the questions in the software development lifecycle sequence so that less experienced software developers would not be intimidated by the section headings, and thus respond more readily. The five new headings used were *Requirements and Design; Code and Test; Configuration Management; Estimates and Schedules; and Project Management and Training*.

The first part of the questionnaire sought demographic information. The second part listed 44 questions pertaining to the use of software development practices, and two new questions about programming languages and tools. The third and final part provided space for respondents to record which aspects of software development needed improvement and to suggest how SEA could help.

In January 1999, 5,634 questionnaires were mailed with a cover letter and reply-paid envelope to organisations identified as possible software developers.

3.3. Summary of results from Queensland survey

The survey responses, representing 203 organisations, indicate that Queensland organisations involved in software development have adopted, on average, almost half of the best practices listed in the questionnaire.

The survey results show that wide variation exists in the adoption of best practice. Firstly, some of the practices, especially those relating to project management planning and customer involvement, are widely adopted, but others, in particular the use of metrics for estimating and testing, are barely used by

the organisations which responded to the survey. Generally, practices of a technical nature are more widely adopted compared to techniques related to support and management. Secondly, considering the level of adoption by organisations, wide variation was recorded as well. Furthermore, higher adoption levels are associated with large development teams, and with organisations which develop commercial-off-the-shelf (COTS) systems. For those not developing COTS systems, the organisations from the utilities, finance and insurance sectors exhibited highest adoption of best practice, with mining and education sectors the lowest.

The survey provided a snapshot of the current state of practice of Queensland development organisations. Comparing the results from this survey with earlier results from Europeans countries, as shown in table 1, the overall adoption of best practice by Queensland organisations is comparable to European firms.

Table 1. Summary of Overall Adoption Levels

Location	Year	N	Adoption
Europe	1995	425	49%
Europe	1997	397	51%
Queensland COTS	1999	87	52%
Queensland non-COTS	1999	116	44%
Queensland All	1999	203	48%

In the next section, challenges and limitations encountered in replicating the ESI survey and data analysis are discussed.

4. Analysis and discussion

The analysis of the Queensland survey highlighted a number of important issues related to the replication of the ESI survey. The following discussion focuses on the content of the questionnaire, interpretation of questions by respondents, reliability of data, generalisability of results, and statistical analysis.

4.1. Currency of best practice techniques

As stated earlier, the content of the ESI questionnaire has been criticised for not covering all possible best practice, specifically that it overlooks important issues related to organisational and customer-supplier management; and it does not include high capability level practices [44]. When the best practice questionnaire was mapped to the ISO/IEC TR

15504-2 processes [47], it became apparent that the best practice questionnaire did not cover the base practices in 13 of the 24 processes included in ISO/IEC TR 15504 [48]. Considering the rigour and effort invested in the formulation of ISO/IEC TR 15504, these processes it includes must be considered to be important, to be *best practice*. The best practice questionnaire heavily emphasises project management, but has no practices relating to risk management, measurement, validation, joint review or audit. There is only one question in the best practice questionnaire which is not covered in ISO/IEC TR 15504: *do procedures exist to ensure that the functionality, strengths, and weaknesses of the system which the software is replacing are formally reviewed?*

As world class standards are dynamic, best practices may be temporary [49], especially in software engineering which has frequently adapted to changes brought about by evolution of technology. With the passage of time, best practice becomes standard practice as other superior practices emerge [50]. For example, agile software engineering methods are now promoted as best practice. Proponents of agile methods would compile a set of best practice techniques focussed on customer satisfaction and early incremental delivery of software; small highly motivated project teams; informal methods; minimal software engineering work products; overall development simplicity stressing delivery over analysis and design; and active and continuous communication between developers and customers [51].

Prior to the Queensland survey execution, one of the members of the SEA National Board questioned the relevance of the techniques in the survey to COTS developers. The point raised was that COTS developers may be using excellent practices which are not included in the ESI questionnaire, for example, management of Beta tests. Jones [52] notes that Beta testing has been used since the 1960s, and Cusumano et al. [53] reported its widespread use at 73 percent. Therefore, it is recognised that the items from the ESI questionnaire may not provide an entirely valid measurement of best practice across the software industry. This point is acknowledged by the ESI: ‘progress in software engineering may not be visible along dimensions measured in the survey’ [54].

Another issue to consider is that software practices may have changed significantly in the six years between the design of the ESI questionnaire and its use in the Queensland survey. For example, reuse is now recognised as one of the most valuable software engineering practices [55], but is not included in the

ESI questionnaire. So while longitudinal studies such as that undertaken by the ESI are valuable in mapping the take-up rate of recommended techniques and practices, the data collection tools need to be kept up-to-date while still providing comparative data. This presents a particular challenge to researchers as modifying the questionnaire may result in findings which are not directly comparable with the findings from the original survey.

4.2. Interpretation by respondents

A further issue concerns how the respondents may have interpreted the questions. Many of the survey questions were worded ‘does a procedure exist ...’ and some respondents may have interpreted this in different ways. For example, is there a *documented* procedure? If so, is it used in *all* projects? Some may have responded *yes* if they occasionally perform the practice, although the procedure may not be documented at all. For each of the 44 questions about best practice, possible responses were *yes*, *no*, *not applicable* or *blank*. Perhaps, rather than *yes/no* options for the adoption questions, the provision of a scale—for example, *all projects; for some projects; no/not applicable; don’t know*—may have improved the precision of the data collected.

4.3. Response rate

To determine the effective response rate, de Vaus [56] advises basing the calculation on the proportion of eligible organisations who received the questionnaire. From the 5,634 organisations to whom questionnaires were mailed, 408 were returned to the researcher by Australia Post as undeliverable, therefore, 5,226 were received by organisations. Of the 563 questionnaires completed and returned, 63 percent (354) were out of scope (i.e. non-developers). Applying this proportion to the number of organisations who received questionnaires, it is estimated that 3,544 would have been non-developers, leaving 2,090 in-scope sample elements. Of these, 209 organisations responded, giving an effective response rate of 10 percent.

The Queensland response rate was disappointing, but within the typical range of 10-20 percent for mail surveys to business establishments [57]. Recent large scale surveys of software development practices received response rates of 8.4 percent and 13.3 percent respectively [58, 59].

In contrast to the wide coverage of the Queensland survey, the ESI questionnaire was administered only to

European firms seeking European Union (EU) funding, in effect, a self-selected sample. The response rates to the ESI surveys have not been published.

4.4. External validity issues

External validity refers to the ability to generalise the results of the research to the external environment [7], and can be improved through replication of the study in other places, with different people, at another time. The ESI results can be generalised to European software development organisations requesting EU funding for software process improvement.

However, for the Queensland study, the extent to which the findings can be generalised to Queensland software development organisations hinges on whether the respondents are representative of software development organisations in Queensland. By using a census, rather than a random sample, sampling frame errors were avoided. The extent of representativeness was determined by comparing characteristics of the responses against the Queensland software industry profile. The responses included in-house, as well as commercial, developers; the proportion of small organisations was consistent with the industry proportion of small firms; and regional and remote organisations were represented in the responses. Therefore, the responses appear to present a balanced coverage of a wide range of software development organisations, with a possible bias towards those interested enough to take the time to complete the questionnaire.

Therefore, as the ESI survey represented firms applying for EU funding for SPI, and the Queensland questionnaires were mass mailed to a large number of organisations, consideration needs to be given to how the difference in motivation of the respondents may have affected the comparability and generalisability of the results.

4.5. Reliability issues

In any survey, respondents may knowingly give untrue or misleading answers [60]. It is difficult to obtain reliable data about current practice because developers are reluctant to admit that they are producing poor software [61]. It is also possible that some survey respondents experienced *evaluation apprehension*. Fearing failure, respondents may have wished to show their organisation in a more positive light, or may have provided the ‘correct’ answers—in

other words, the responses they thought the researcher was seeking.

Although the Queensland best practice questionnaire was addressed to the Software Manager by name or by title, the researcher had no control over who actually completed the form. Also, it is possible that some respondents may not be aware of the actual extent of use of various software engineering practices within their organisations. Prior studies have found that staff at different levels in software organisations reported inconsistent results when surveyed about actual use of software development practices [31, 62].

4.6. Statistical validity concerns

Prior to undertaking any correlation analysis to statistically prove the association between adoption of best practice and variables such as organisation size, the characteristics of the Queensland data were explored in order to ensure the correct statistical approach was selected. The assumption of normality is a prerequisite for many inferential statistical techniques. The K-S (Lilliefors) statistic was calculated to assess if the distributions of practice adoption level and organisation adoption level were distributed normally. Although practice adoption level exhibited a normal distribution, the significance level of this test on the variable organization adoption level was less than the required level of .05, indicating that normality cannot be assumed (Kolmogorov-Smirnov (Lilliefors statistic) = .067; df=203; $p=.028$).

Smaller groups of responses were selected to assess the normality of organization adoption level within each group. Groups were selected based on organization size, team size and the organisation’s primary involvement in the industry. The complimentary groups of non-COTS and COTS developers were the only subsets of responses which exhibited a normal distribution for the variable *organisation adoption level*.

As the statistical analysis of the distribution of the variable *organisation adoption level* revealed that the responses fall into two distinct populations (COTS and non-COTS developers), subsequent tests involving the variable *organisation adoption level* were conducted separately for COTS and non-COTS organisations. In the reports of the ESI survey data analysis, parametric tests were used and the characteristics of the distribution of the variables were not reported.

Some researchers claim that useful studies can be conducted even when proscriptions are violated [63], however, others disagree with this approach and assert

that a combination of descriptive and nonparametric techniques are adequate to establish the presence or absence of statistically significant correlations [64].

To provide guidance to software engineering managers, and possibly purchasers, Dutta, Lee and Van Wassenhove [42] assumed normality and used discriminant analysis to identify the practices which distinguish the ESI high level adopters from the low level adopters, and found four practices accounted for the greatest variance.

Although the discriminant analysis requires that the variables are continuous and normally distributed (conditions not met by the Queensland survey data), it was decided to attempt the analysis in order to compare the Queensland survey results with the ESI results. As discriminant analysis cannot be calculated for variables with missing values, seven of the 44 questions which had a high proportion (>20%) of *not applicable* or missing values were excluded from this analysis. The discriminant analysis identified only one question accounting for the variance between high and low overall adoption: *Use of a formal procedure to produce software development effort, schedule, and cost estimates* (Wilks Lambda=.001).

However, due to the fact that the distribution was neither continuous nor normally distributed, both required conditions for use of the discriminant analysis test, and the responses indicated that many of the best practice questions were answered *yes* by the high adoption organisations and *no* by the low adopters, it was decided to conduct the non-parametric Fisher's exact test to compare the responses of the 20 Queensland high level adopters and 20 low adopter organisations for each question. For Fisher's exact tests, all 44 questions were included, and the cases were selected based solely on their overall adoption level (unlike in the discriminant analysis where cases with missing values were excluded). It is recognised that some of the best practice questions are inter-related to some degree, so to take this into account, the *p* value was multiplied by the number of questions (44) (Bonferroni correction as advised in [65]).

In contrast to the discriminant analysis, Fisher's exact test found that a significant difference existed between the high adoption group and the low adoption group for almost all the survey questions (40 of the 44 questions). This important result would not have been found if the analysis had been restricted to a replication of the ESI analysis.

5. Recommendations

After consideration of the issues discussed, recommendations are now formulated to provide guidance to researchers who are undertaking replicated surveys.

It is recommended that, prior to adopting an existing survey instrument, researchers should consider the currency of the question items. It may be necessary to conduct research to identify additional questions to bring the instrument up-to-date with current practice. The addition of more question items does not prevent the comparison of the new results with the original survey findings, but it does require care in reporting such comparisons.

To capture more accurate information about practices used in organisations, researchers are urged to avoid the two-point 'yes/no' scale in favour of a multi-point Likert-type scale such as '*all projects, some projects, no/not applicable, don't know*'. Furthermore, researchers are advised to review the published statistical analysis from prior use of the survey instrument to assess if the analysis is rigorous and if the underlying assumptions are stated and not violated.

To boost response rates, careful targeted sampling is recommended as businesses have built up a considerable resistance to answering mail surveys due to constant inundation with both mail and telephone surveys [66]. The costs incurred with printing and mailing to a large number of organisations may be reduced by use of web technology. Web-based or emailed surveys are less expensive to distribute; it is easier to follow-up for non-respondents; responses do not need to be edited or keyed; and permission can be sought to use e-mail contact lists. The Western Australian branch of SEA used this approach to distribute the Best Practice survey to its members. To counter poor response rates, researchers should also consider designing their studies to complete the survey instrument in face-to-face meetings with developers. Although this approach incurs greater costs compared to mailed surveys, the data collected may be more reliable.

Echoing advice provided by Kitchenham et al. [1], researchers are urged to include full details in their publications to facilitate replication by others. Research publications should explain how the hypotheses were derived from theory, define the variables and state how the sample was selected. The statistical tests and their underlying assumptions, and also the confidence level of each statistical test, should

be reported [67]. In practice, however, limitations on the length of research papers for conferences and journals make it difficult for researchers to include all those elements. Perhaps researchers could provide detailed explanations on a web page, and provide the URL for interested readers. The ESI publications include a copy of the survey instrument, a practice often overlooked by other researchers.

Finally, researchers should be encouraged to collaborate with the researchers who reported the original findings [3]. Such communication will provide a clearer understanding of the motivation, context and limitations of the original study to the researcher undertaking the replication. Lessons learned from the previous surveys can be incorporated into the replication to strengthen the instrument and analysis. This type of participation benefits the software engineering research community by helping less experienced researchers to network with the established experts in the SE field.

6. Conclusions

Surveys are an essential tool for software engineering research and should be promoted to gather information about what software engineers actually do, and to evaluate prescribed practices, methods, tools and standards.

Despite the potential pitfalls and challenges, such as those discussed in this paper, researchers should be encouraged to replicate surveys as replication and refinement provides stronger constructs and variables, and adds to the body of knowledge by building on previous studies. Also, replicating an existing study can reduce the time needed to define the variables, to create the instrument, and to design the statistical tests. Efficiencies result in the presentation of more timely results, an important consideration in the fast-changing SE field.

This paper has analysed problems encountered in replicating a European survey in an Australian context. From the analysis, recommendations have been formulated regarding the questionnaire content, response scale, sample selection, and statistical analysis. These recommendations will help overcome some of the inherent problems in survey replication, and encourage researchers to replicate existing surveys to extend and strengthen the SE discipline.

10. References

- [1] B. Kitchenham, S.L. Pfleeger, L.M. Pickard, P.W. Jones, D.C. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering", *IEEE Transactions on Software Engineering*, vol. 28, no. 8, 2002, pp. 721-734.
- [2] ABS, "8126.0 Information and Communication Technology, Australia 2002-03", accessed 15 Oct, <http://www.abs.gov.au/Ausstats/abs@.nsf/1020492cfd63696ca2568a1002477b5/60c99311996ef774ca256f070079c995!OpenDocument>, 2004.
- [3] R.F. Bornstein, "Publication politics, experimenter bias and the replication process in social science research", in *Replication Research in the Social Sciences*, J.W. Neuliep, Ed. Sage, Newbury Park, Ca, USA, 1991, pp. 71-81.
- [4] K.E. Stanovich, *How to think straight about psychology*, 2nd ed., Scott Foresman & Co., Glenview IL, 1989.
- [5] A. Brooks, J. Daly, J. Miller, M. Roper, and M. Wood, "Replication of experimental results in software engineering", accessed 16 January, 1996, <http://www2.umassd.edu/SWPI/ISERN/ISERN-96-10.pdf>.
- [6] E. Fechak, "Summary and analysis of 18th international conference on software engineering", DACS, accessed 19 January, 1996, <http://www.softwaretechnews.com/summer96/icse18.html>.
- [7] W.G. Zikmund, *Business Research Methods*, Dryden Press, Fort Worth, 1994.
- [8] W.S. Humphrey and W.L. Sweet, "A Method for Assessing the Software Engineering Capability of Contractors", Software Engineering Institute, CMU, Pittsburgh CMU/SEI-87-TR-23, September, 1987.
- [9] D. Zubrow, W. Hayes, and J. Siegel, "A Method for Assessing the Software Engineering Capability of Contractors", Software Engineering Institute, CMU, Pittsburgh SEI-94-SR-007, 1994.
- [10] P. Fusaro, K. El Emam, and B. Smith, "The internal consistencies of the 1987 SEI maturity questionnaire and the SPICE capability dimension", *Empirical Software Engineering*, vol. 3, no. 2, 1997, pp. 179-201.
- [11] W.S. Humphrey, D.H. Kitson, and J. Gale, "A comparison of U.S. and Japanese software process maturity", presented at 13th International Conference on Software Engineering, Austin, Texas, 1991.
- [12] M. Tan and C.Y. Yap, "Impact of organisational maturity on software quality", in *Software Quality and Productivity: Theory Practice, Education and Training*, M. Lee, B. Barta, and P. Juliff, Eds., Chapman and Hall for IFIP, London, 1995, pp. 231-234.
- [13] A.P. Cater-Steel and E.P. Fitzgerald, "Quality Assurance Certification: Adoption by Australian Software Developers and its Association with Capability Maturity", presented at Asia Pacific Software Engineering Conference (APSEC), Hong Kong, 1997.
- [14] D.R. Goldenson and J.D. Herbsleb, "After the Appraisal: A Systematic Survey of Process Improvement, its

- Benefits, and Factors that Influence Success", Software Engineering Institute, CMU CMU/SEI-95-TR-009, August, 1995.
- [15] K. El Emam and A. Birk, "Validating the ISO/IEC 15504 measure of software requirements analysis process capability", *IEEE Transactions on Software Engineering*, vol. 26, no. 6, 2000, pp. 541-566.
- [16] F. Brooks, "No silver bullet: essence and accidents of software engineering", *IEEE Computer*, vol. 20, no. 4, 1987, pp. 10-19.
- [17] S. McConnell, "Closing the gap", *IEEE Software*, vol. 19, no. 1, 2002, pp. 3-5.
- [18] J. Ludewig, "Software engineering in the year 2000 minus and plus ten", in *Informatics: 10 years back, 10 years ahead.*, R. Wilhelm, Ed., Springer-Verlag, Heidelberg, Berlin, 2001, pp. 102-111.
- [19] R.L. Glass, "The state of the practice of software engineering", *IEEE Software*, vol. 20, no. 6, 2003, pp. 20-21.
- [20] B. Fitzgerald, "The use of systems development methodologies in practice: a field study", *Information Systems Journal*, vol. 7, no. 3, pp. 201-212, 1997.
- [21] T. McBride, "Standards need more rigour", in *Information Age*, 2004, pp. 65-66.
- [22] N. Fenton, S.L. Pfleeger, and R.L. Glass, "Science and substance: a challenge to software engineers", *IEEE Software*, vol. 11, no. 4, 1994, pp. 86-95.
- [23] I. Richardson and K. Ryan, "Software Process Improvements in a Very Small Company", *Software Quality Professional*, vol. 3, no. 2, 2001, pp. 23-35.
- [24] F.G. Wilkie, D. McFall, and F. McCaffery, "The centre for software process technologies: a model for process improvement in geographical regions with small software industries", presented at Software Engineering Process Group, Orlando, Florida, USA, 2004.
- [25] G. Bucci, M. Campanai, and G. Cignoni, "Rapid assessment to solicit process improvement in small and medium-sized organisations", *Software Quality Professional*, vol. 4, no. 1, 2001, pp. 33-41.
- [26] S.L. Pfleeger, D.R. Jeffery, B. Curtis, and B. Kitchenham, "Status report on software management", *IEEE Software*, vol. 14, no. 2, 1997, pp. 33-43.
- [27] M.C. Paulk, D.R. Goldenson, and D.M. White, "The 1999 Survey of High Maturity Organizations", Software Engineering Institute, CMU, Pittsburgh CMU/SEI-2000-SR-002, February, 2000.
- [28] S. Bandinelli, A. Fuggetta, L. Lavazza, M. Loi, and G.P. Picco, "Modeling and improving an industrial software process", *IEEE Transactions on Software Engineering*, vol. 21, no. 5, 1995, pp. 440-454.
- [29] S. Dutta, L.N. Van Wassenhove, and S. Kulandaiswamy, "Benchmarking European software management practices", *Communications of the ACM*, vol. 41, no. 6, 1998, pp. 77-86.
- [30] A. Goldsworthy, "The Global Information Economy: The Way Ahead", DIST, Commonwealth of Australia, Canberra July, 1997.
- [31] D.N. Wilson, P. Petocz, and K. Roiter, "A study of software quality assurance in practice", in *Software Quality Management III*, vol. 1, Ross, Brebbia, Staples, and Stapleton, Eds., 1995, pp. 159-169.
- [32] M. Azuma and D. Mole, "Software management practice and metrics in the European Community and Japan: some results from a survey", *Journal of Systems Software*, vol. 26, pp. 5-18, 1994.
- [33] J. Blackburn, G. Scudder, and L. Van Wassenhove, "Improving speed and productivity of software development: a global survey of software developers", *IEEE Transactions on Software Engineering*, vol. 22, no. 12, 1996, pp. 875-885.
- [34] B. Curtis, H. Krasner, and N. Iscoe, "A field study of the software design process for large systems", *Communications of the ACM*, vol. 31, no. 11, 1988, pp. 1268-1287.
- [35] M. Zelkowitz, R. Yeh, R. Hamlet, J. Gannon, and V. Basili, "Software Engineering Practices in the U.S. and Japan", *IEEE Computer*, vol. 17, no. 6, 1984, pp. 57-66.
- [36] C.J. Davis, J.B. Thompson, P. Smith, and A. Gillies, "A Survey of Approaches to Software Quality within the United Kingdom", University of Sunderland, 1992.
- [37] C. McPhee and A. Eberlein, "Requirements Engineering for Time-to-Market Projects", presented at 9th IEEE International Conference on the Engineering of Computer Based Systems, Lund, Sweden, 2002.
- [38] A. Rouse, D. Watson, A. Gilbert, and R. Dilnutt, "A Comparison of the System Development Practices in Australian and United States' Organizations", *Monash Information Systems Working Papers Series*, no. 24/95 June, 1995.
- [39] D.R. Jeffery and B. Zucker, "The state of practice in software metrics", School of Info Systems, UNSW, Sydney Tech Report 97/1, 1997.
- [40] J.M. Verner and W.M. Evanco, "In-house software development: what software project management practices lead to success?", *IEEE Software*, vol. Jan/Feb, pp. 86-93, 2005.
- [41] ESI, "Software Best Practice Questionnaire Comparative Analysis of 1995 and 1996", European Software Institute, Bilboa ESI-PIA/96303, November, 1996.
- [42] S. Dutta, M. Lee, and L. Van Wassenhove, "Software engineering in Europe: a study of best practices", *IEEE Software*, vol. 16, no. 3, 1999, pp. 82-90.
- [43] ESI, "1997 Software Best Practice Questionnaire: Analysis of Results", European Software Institute, Bilboa 1997.
- [44] S. Dutta and L. N. Van Wassenhove, "An empirical study of adoption levels of software management practices within European firms", INSEAD, Fontainebleau, France 97/118/TM, 1997.
- [45] Gartner Group, "MIS 3000 (database) ", S.P. Group, Ed., Sydney, 1995.
- [46] Dependable Database Data, "Australia on Disc," 1998.
- [47] ISO/IEC TR 15504, "Information Technology—Software process assessment, Parts 1-9", ISO/IEC TR 15504:1998(E), 1998.

- [48] A.P. Cater-Steel, "An evaluation of software development practice and assessment-based process improvement in small software development firms", in *EIT*. Griffith University, Brisbane, 2004, pp. 465.
- [49] J.W. Cortada, *Best Practices in Information Technology*, Prentice Hall, Upper Saddle River, NJ, 1997.
- [50] P.B. Cragg, "Benchmarking information technology practices in small firms", *European Journal of Information Systems*, vol. 11, no. 4, 2002, pp. 267-282.
- [51] R.S. Pressman, "Software Engineering Resources", R.S. Pressman & Associates, 2003, accessed 27 July, <http://www.rspa.com/spi/process-agile.html#general>.
- [52] C. Jones, "Variations in software development practices", *IEEE Software*, vol. Nov/Dec, pp. 22-27, 2003.
- [53] M. A. Cusumano, A. MacCormack, C. F. Kemerer, and B. Crandall, "Software development worldwide: the state of the practice", *IEEE Software*, vol. 20, no. 6, 2003, pp. 28-34.
- [54] ESI, "Software Best Practice Questionnaire: Trend Analysis 1995-1997", European Software Institute, Bilbao ESI-1998-ESI-98-TR-001, 1998.
- [55] H. Mili, F. Mili, and A. Mili, "Reusing software: issues and research directions", *IEEE Transactions on Software Engineering*, vol. 21, no. 6, 1995, pp. 528-561.
- [56] D. A. de Vaus, *Response Rate Calculation*, 1996.
- [57] C. M. Paxson, "Follow-up mail survey"s, *Industrial Marketing Management*, vol. 21, no. 3, 1992, pp. 195-201.
- [58] T. Stalhane, P. Borgersen, and K. Arnesen, "In search of the customer's quality view", *Journal of Systems Software*, vol. 38, pp. 85-93, 1997.
- [59] E. A. Larsen and K. Kautz, "Quality assurance and software process improvement in Norway", *Software Process Improvement and Practice*, vol. 3, no. 2, 1997, pp. 71-86.
- [60] G.C. Moore, "An examination of the implementation of information technology for end-users: a diffusion of innovations perspective", University of British Columbia, Vancouver, BC, 1989.
- [61] M. Khurana, Z. He, I. Court, M. Ross, G. Staples, and D. Wilson, "Software quality practices—an empirical study", *Software Quality Journal*, no. 5, 1996, pp. 75-85.
- [62] G.M. Weinberg, "Just say no! Improving the requirements process", *American Programmer*, vol. 8, no. 10, 1995, pp. 19-23.
- [63] L. Briand, K. El Emam, and S. Morasca, "On the Application of Measurement Theory in Software Engineering", 1995, accessed 14 June, <http://www2.umassd.edu/SWPI/ISERN/isern-95-04.pdf>.
- [64] P.K. Lawlis, R.M. Flowe, and J.B. Thordahl, "A correlational study of the CMM and software development performance", in *Crosstalk: The Journal of Defense Software Engineering*, 1995.
- [65] J. Miller, "Statistical significance testing—a panacea for software technology experiments?", *The Journal of Systems and Software*, vol. 73, pp. 183-192, 2004.
- [66] P. Attewell and J.B. Rule, "Survey and other methodologies applied to IT impact research: experiences from a comparative study of business computing", in *The Information Systems Research Challenge: Survey Research Methods*, vol. 3, *Harvard Business School Research Colloquium*, K. Kraemer, Ed., 1991, pp. 299-315.
- [67] D.E. Perry, A.A. Porter, and L.G. Votta, "Empirical studies of software engineering: a roadmap", in *Proceedings of the Conference on the Future of Software Engineering*: ACM Press, 2000, pp. 345-355.