POLITECNICO DI TORINO
Repository ISTITUZIONALE

Six reasons for rejecting an industrial survey paper

Original
Six reasons for rejecting an industrial survey paper / Torchiano, Marco; Filippo, Ricca. - STAMPA. - (2013), pp. 21-26. (Intervento presentato al convegno First Intl. Workshop on Conducting Empirical Studies in Industry CESI 2013 tenutosi a San Francisco (USA) nel 20 May) [10.1109/CESI.2013.6618465].

Availability:
This version is available at: 11583/2506232 since:

Publisher:
IEEE

Published
DOI:10.1109/CESI.2013.6618465

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

Six reasons for rejecting an industrial survey paper

Marco Torchiano
Department of Control and Computer Engineering
Politecnico di Torino, Italy
Email: marco.torchiano@polito.it

Filippo Ricca
DIBRIS
Università di Genova, Italy
Email: filippo.ricca@disi.unige.it

***Abstract*—Context: Despite their importance in any empirically based research program, industrial surveys are not very common in the software engineering literature. In our experience, a possible reason is their difficulty of publication.**

Goal: We would like to understand what are the issues that may prevent the publication of papers reporting industrial surveys.

Method: In this preliminary work, we analyzed the surveys we conducted and extracted the main lessons learned in terms of issues and problems.

Results: Most common critics posed to industrial surveys are: lack of novelty, limitation of the geographic scope and sampling issues.

Conclusions: Most objections that led to reject a survey paper actually are not easy to overcome and others are not so serious. These objections could restrain researchers from conducting this type of studies that represent an important methodological asset. For these reasons, we think that reviewers should be less severe to judge survey papers provided that all the limitations of the study are well explained and highlighted.

I. Introduction

In the software engineering literature two kind of studies provide a *macro* or large-scale view on the field: literature reviews and population surveys. Ciolkowski *et al.* in [1] notice that, while it is quite common to find literature reviews in the Software Engineering field, not many efforts are reported in conducting surveys. The motivation they propose is that while the source of information for reviews are usually papers and books, surveys deal with people. Moreover the process of conducting the survey itself makes its success more difficult.

The general literature, and the above cited paper is no exception, uses the term "survey" with two distinct meanings: (i) literature review, which could be conducted with a rigorous approach an thus qualify as Systematic (SLR) [6], and (ii) population survey [5], which in the software engineering field are often conducted using the questionnaire instrument on and industrial population. In this paper we use the term survey exclusively with the latter meaning.

We strongly believe that industrial surveys in software engineering are really important and useful but they are, at the same time, really difficult to design and conduct/execute (this last point has also been rigorously stated in [11]). In addition to the motivation reported by Ciolkowski *et al.*, and to the fact that surveys are really time-consuming (design of the instrument and execution), we believe that surveys are not very common in (empirical) software engineering literature because they are hard to publish.

In the past years, we conducted several questionnaire-based industrial studies about a few software engineering topics (e.g., model driven development and SOA) [9], [10], [4], [3], [13], [14], [8], [7], [15], [12] with a common denominator: a lot of work (definition of the target population, instrument design, pilot study to validate the questionnaire, data collection, etc.), time invested and "a painful back" in terms of publications. Even in the cases when the story finished well (e.g, in [4], [3], [14] with IEEE Software, TSE and IST publications), it was really difficult to reply to the meticulous reviewers and the earlier conference papers were rejected several times.

The goal of this preliminary work is understanding and discussing what have been the specific issues (and real problems) that prevented the publication of our papers reporting industrial surveys. To try to reach this goal, we have analyzed all the survey specific comments by the reviewers about our articles for the purpose of characterizing the critics with respect to their frequency. We then analyzed the frequency of each specific category of comments and tried to comment about the groundedness. Finally, when suitable, we proposed a generic rebuttal.

The paper is structured as follows. Sect. II briefly summarizes the industrial surveys we conducted. Sect. III shows the applied experimental method while Sect. IV lists, partitions in categories and comments the received critics. Finally, conclusions close the paper in Sect. V.

II. Experience gained

The authors of this paper were involved in first person in the conduction of several industrial surveys in the last ten years. In particular we can count six distinct questionnaire-based surveys where either of us were involved.

The main features of the studies are summarized in Table I in terms of:

1) ID code used to refer to the study throughout the paper
2) the year when the study was conducted
3) the year of the first publication
4) the number of respondents (N)
5) the geographical scope of the study
6) the topic,

TABLE I
SUMMARY OF CONDUCTED INDUSTRIAL SURVEYS

ID	Conducted	First Pub.	*N*	Scope	Topic	Sampling
OTS	2004	2005	133	Norway, Germany, Italy	Development with Off-The-Shelf components	Stratified random from CC Db
SF3C	2007	2009	32+40	Sweden, Australia, North Italy	Software projects success factors	random from CC db
MIG	2007	2008	59	Italy	Software migration (Web, SOA, smartphones)	random from CC db + NC
SF	2008	2010	33	Turin area	Perception of software projects success factors	Stratified random from CC db
SOA	2008, 2011	2012	40+119	North Italy	SOA knowledge, adoption and trend	non probabilistic (NC + ML + WG)
MDE	2011	2011	155	North Italy	Model-driven engineering	random from CC + NC + ML + WG

7) the method used for the sampling and the source of contacts, which typically come from one or more among: the Commerce Chamber database (CC db)[1], list from the industrial relationship networks (NC), public mailing lists (ML), and web groups (WG).

More in detail the studies we conducted, in chronological order, are:

OTS The focus of this study was investigating the state of the practice in OTS-based development (i.e., development based on the usage of components such as COTS or open source software) and the reasons for applying these practices [9], [10].

SF3C Different stakeholders involved in the software development may attribute the success of a project to different factors (e.g., realistic expectations and good relation between personnel). This industrial survey explored how different stakeholders perceive project success and what they deem the effect of specific factors on the project outcome [4].

MIG This study aimed at achieving an accurate picture of the state-of-the-practice of software migration activities in the Italian industry, focusing in particular on projects dealing with migration towards the Web, distributed, service-oriented and mobile architectures [13], [14].

SF The main goal of this survey was to assess the perceived effects of several factors (e.g., customer was involved in the project, complete and accurate requirements from the start and software process was monitored using metrics) on the success or failure of the software projects [3].

SOA The main aim of this work was investigating the level of knowledge and diffusion of SOA (Service Oriented Architecture) in the Italian industry. We were also interested to understand what the trend of SOA was (positive or negative?) and what the methods, technologies and tools really used in the industry were [7], [8].

MDE This industrial survey has been conducted with the aim of investigating to what extent and how Italian companies deal with software models and Model driven engineering. The goals of the MDE survey were the following: (1) to understand the real penetration of software modelling in the Italian industry, (2) to understand and document if and how MD* is applied in industry, (3) to understand which motivations led to or prevented modelling and MD* adoption and (4) to understand which processes, modelling languages and tools are used in this context [12], [15].

[1]According to Italian law, any company must sign to a Commerce Chamber, thus the Commerce Chamber database contains an exhaustive list of the Italian IT companies

III. METHOD

The goal of our exercise is to analyze the survey specific comments about our articles for the purpose of characterizing the main issues with respect to their frequency and groundedness from the point of view of the actual or prospective survey conductor in the context of software engineering research.

For each paper in the examined set:

1) we selected all the reviews received independently from the outcome (i.e., accept or reject)
2) we categorized the reviews for venue (i.e., conference or journal)
3) we browsed all the reviews received and extracted the sentences concerning issues specific to population surveys and discarded those concerning other aspects. For instance the sentence "*The largest threat to validity to me seems to be self-selection of respondents*" has been considered while we discarded "*Box plots are not enough to describe the data. It would be preferable to include the raw data*"
4) we then analyzed the sentences and identified common themes that appeared in one or more sentences

TABLE II
FREQUENCY OF SURVEY-RELATED COMMENTS PER STUDY AND BY PUBLICATION TYPE

		Type		Study					
	Freq	Journal	Confer.	OTS	SF	SF3C	SOA	Migration	MDE
No Practical usefullness	16	2	14	1	0	2	5	4	4
Sampling bias	16	5	11	2	0	1	6	3	4
Obvious conclusions	7	1	6	1	0	0	4	0	2
People perceptions	7	3	4	2	1	1	3	0	0
Non respondents	6	3	3	1	0	1	2	1	1
Limited geographical scope	5	1	4	1	0	1	1	1	1
Provide package	3	3	0	2	0	1	0	0	0
Limited constructs, factors, or RQs	3	3	0	1	1	1	0	0	0
Sample size	2	0	2	0	0	0	1	1	0
Too much emphasis on method	1	1	0	1	0	0	0	0	0
Imputation missing values	1	1	0	1	0	0	0	0	0
Too broad questions	1	0	1	0	0	0	1	0	0

5) we grouped together the sentences belonging to common themes (i.e., categories)
6) we then analyzed the frequency of each specific category of comments (i.e., we counted the number of comments contained in each category) and tried to comment about their groundedness
7) finally, when suitable, we proposed a generic rebuttal.

IV. CRITICISMS

The result of the procedure outlined in section III are summarized in Table II as frequency of comments. The frequency is reported overall and divided by type of publication and per study.

We first report a few consideration about the general distribution of comments frequency and then we analyze in details the most frequent ones (in this preliminary work, we decided to consider only the first six categories).

First of all, we observe a statistically significant difference of the comments distribution among journal and conference papers (χ^2-test p-value = 0.017). The differences can be graphically appreciated in Figure 1. In particular, we observe markedly higher density of the *No practical usefulness*, *Obvious conclusions*, and *Limited geographical scope* related to the Conference papers than the Journal ones. On the other hand, *Non respondents analysis is missing*, *Provide experimental package* and *Limited constructs, factors, or RQs*[2] are much more common relative to Journals than conferences.

As we will see in a while, the Conference specific comments seem to be the least grounded criticisms. Probably, this is due to the lower expertise about surveys on the part of conferences reviewers w.r.t. journal reviewers.

A. No practical usefulness

An extremely frequent comment we received was that the results presented bear no practical usefulness. For instance the reviewers told us: *"it is not clear how other researchers or practitioners can benefit from the outcomes of this study"*, *"The whole paper, even the conclusion, is strictly descriptive"* and that the *"After reading the paper the question remains: interesting study but who are the results useful for? and how can they be used/applied by others? Given the study, what are you addressing/suggesting?"*. Similar comments targeted almost all our studies.

We believe that a common goal of several surveys in our field is to take a snapshot of some software engineering aspects. Such a snapshot defines the framework for setting up any grounded research plan. Therefore surveys, far from being not practical, have the potential to represent the basis, whatever is the result, to conduct research that has practical relevance.

B. Sampling bias

Gathered under this category we can find at least three distinct issues the survey sample.

- self-selection bias (e.g. *"The largest threat to validity to me seems to be self-selection of respondents."*)
- sampling frame, that is the instrument or method used to select the participants in order to obtain a representative sample (e.g. *"The sample is convenience-based or probably strongly biased.*)
- the sample is not representative of whatever the general population is (e.g. *"one of the most critical issues in surveys is to obtain a valid sample of the target population"*

Concerning the first point at issue, we must make it clear that any survey ought to balance two distinct (and opposing) needs: provide some background introduction so as to motivate prospective respondents, though not too much to intimidate or dissuade those not deeply versed in the topic. The possibility of self-exclusion bias is hardly avoidable, that should be reported among the threats to validity by any honest survey researcher.

Concerning the second type of comments, it is important to understand how relevant companies, i.e. those hosting software development projects, can be located. Typically national administrations in European Community adopt a standard classification of economic activities (NACE[3]). In the latest classi-

[2]We gathered under this item all comments concerning mainly usefulness, validity, relevance of the constructs measured with the questionnaire, the investigated factors, or the research questions

[3]http://circa.europa.eu/irc/dsis/nacecpacon/info/data/en/index.htm - visited on Jan 30, 2013

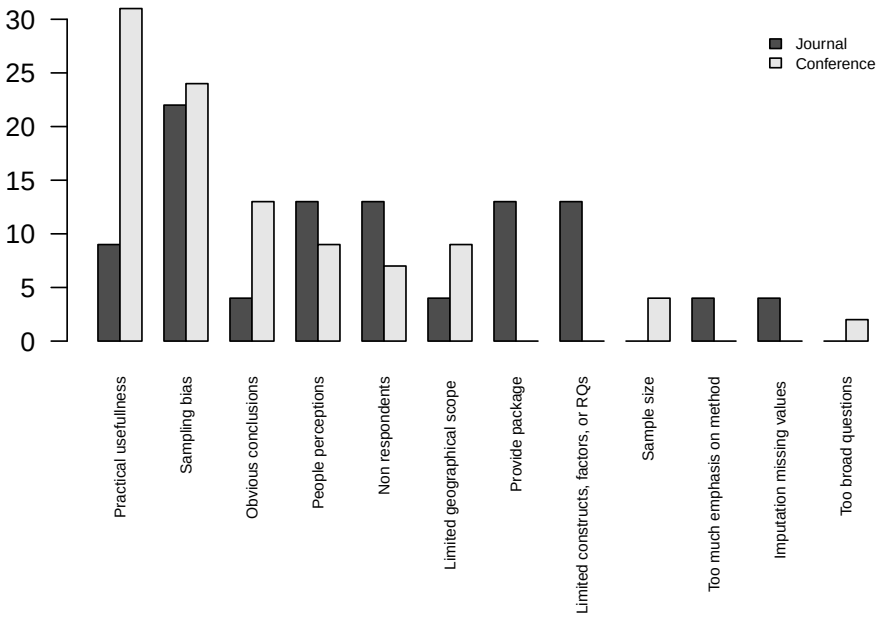

Fig. 1. Distribution of comments by venue type: Journal vs. Conference

fication the software relevant categories are J.62.01 (Computer programming activities) and J.62.02 (Computer consultancy activities). This high-level classification poses two problem [2]: (i) not all companies with those codes do actually develop software and (ii) several companies with other codes – e.g. telecoms – do actually develop software. As a consequence, a pure random sampling based on official records (such as the commerce chamber databases) and standard codes has little chances of sampling relevant companies. As a consequence, supplementing random sampling with convenience sampling is a necessity. So the ensuing threat to validity is unavoidable and the only possible behavior is to accept and report it.

As far as the third sample-related criticism, in the surveys we conducted the sampling unit that constitute the target population are software development projects. Since no directory exists that lists projects, we were forced to perform and indirect sampling – getting to the projects the companies –, which introduce clustering issues [5]. Moreover, often, the respondents – developers and managers – do not correspond to the sampling units – projects – of the study thus a second clustering is introduced.

A typical problem deriving from the first clustering – several eligible projects in the same company – consists in an uneven probability of selecting projects hosted in large companies w.r.t those hosted in smaller ones. A possible mitigation strategy consists in oversampling large companies in order to render the selection probability fairer.

The second clustering – several projects for a given respondent – involve a possible bias when the respondent has to select the project. Here a possible strategy is to ask about the most recent project or the project the respondent is more familiar with. The goal is to ensure high reliability of the information collected through the survey.

The two strategies mentioned above are just examples. The truth is that despite all the possible efforts to assure a representative sample any assessment is very difficult and subjective. We believe that too much emphasis on the representativeness of the sample looks like discussing about the gender of the Angels.

C. Obvious conclusions

The typical comment we received is something along the line of *"The results are hardly surprising or controversial"*. Generally, we do not understand why surveys should report surprising and/or controversial findings, while this is not in general expected from other forms of empirical studies. Moreover, this sound really strange to us. Elaborating a little about this kind of comments, it seems to us that the reviewers would like to see published only industrial survey reporting surprising and controversial findings. But this is a nonsense: a survey take a picture of the reality, and the reality is rarely surprising or controversial.

Actually, when we read this kind of comments, often comes to our mind the famous song by Leonard Cohen, Everybody Knows

Everybody knows that the dice are loaded
Everybody rolls with their fingers crossed
Everybody knows that the war is over
Everybody knows the good guys lost
Everybody knows the fight was fixed
The poor stay poor, the rich get rich
That's how it goes
Everybody knows
. . . [4]

[4]The whole song can be listened e.g. here: http://www.youtube.com/watch?v=GUfS8LyeUyM

We dismiss this comments in such a unserious and hilarious way because most of the time they mistake anecdotal evidence for empirically based evidence: the former is yet another everybody knows, the latter is a form of science.

D. People perceptions

Sometimes we receive comments along the lines of "*it seems to provide only a 'the general perception is' argument.*". The observation is strictly speaking true, the questionnaire instrument is designed to collect self-reported behaviors, facts, and attitudes.

Two rebuttals are possible for this kind of comments. First, this is a known limitation but it is the price to pay to get a large scale snapshot of a phenomenon that would be otherwise impossible to take. Second, software engineering activities heavily depend on humans and therefore their perceptions matter, possibly even more than any automatic and abstract assessment of a method, technology, or tool.

E. Non respondents

A few comments we received concern the non-respondents (e.g. "*Did you perform a non-respondent analysis? Why did those people not answer the questionnaire, although you have contacted them before?*"). A detailed analysis of the non respondents, e.g. to understand the reasons for their missing response, is clearly quite difficult: if they did not replied in the first place it is unlikely that they will devote some time to explain us why.

As far as the response rate is concerned, it can be computed only when standard sampling methods are used. Though, if mailing lists are used to advertise the survey – as we did in a few cases – the number of contacted people is not known and thus the response rate cannot be computed.

F. Limited geographical scope

A common complaint we received for our surveys concerns the limited geographical coverage ("*Furthermore, all results are of course limited to the Italian industry*") even in cases when several countries were involved (e.g."*I would suggest as future work to extend the survey to other continents. All the companies in the current survey are from Europe and it can be nice to see if the state-of-the-practice is similar in other latitudes*").

While strictly speaking such criticism is generally grounded, unless a very large scale international cooperation is set up to conduct a survey (with a huge effort), the coverage will always be more or less limited. Such limitation, per se, does not represent an issue for rejecting a paper provided the researchers do not attempt to stretch the generalization beyond reasonable limits. For instance IT industry in North of Italy does not differ much from other regions in central Europe, while it may differ quite significantly from North America.

V. Conclusions

In this preliminary work, we have analyzed the comments of the reviewers about our six previously conducted industrial surveys and we have grouped together the sentences belonging to common themes. We discovered that the most common critics were: lack of novelty, limitation of the geographic scope and sampling issues.

Some objections that led to reject a survey paper (or prevented its publication) actually are not easy to overcome (e.g., *Sampling bias*) and others are not so serious to deserve a strong reject or a publishing delay (e.g., *Obvious conclusions* and *Limited geographical scope*). For these reasons, we think that reviewers should be less severe to judge survey papers. Otherwise, this trend could restrain researchers from conducting this type of studies that represent an important methodological asset.

In conclusion, we believe that some, however shaky, evidence is better than no evidence. Provided the study has been well conducted we are well aware of its limitations. Paraphrasing the famous statement by Pleeger and Kitchenham [11]: *We do not want to give the impression that there is any way of turning a bad survey into a* publication; *if a survey is a lemon, it stays a lemon* and should not be published.

References

[1] M. Ciolkowski, O. Laitenberger, S. Vegas, and S. Biffl. Practical experiences in the design and conduct of surveys in empirical software engineering. In R. Conradi and A. I. Wang, editors, *ESERNET*, volume 2765 of *Lecture Notes in Computer Science*, pages 104–128. Springer, 2003.

[2] R. Conradi, J. Li, O. Slyngstad, V. Kampenes, C. Bunse, M. Morisio, and M. Torchiano. Reflections on conducting an international survey of software engineering. In *Empirical Software Engineering, 2005. 2005 International Symposium on*, page 10 pp., 2005.

[3] E. Egorova, M. Torchiano, and M. Morisio. Actual vs. perceived effect of software engineering practices in the italian industry. *Journal of Systems and Software*, 83(10):1907 – 1916, 2010.

[4] E. Egorova, M. Torchiano, M. Morisio, C. Wohlin, A. Aurum, and R. Svensson. Stakeholders' perception of success: An empirical investigation. In *Software Engineering and Advanced Applications, 2009. SEAA '09. 35th Euromicro Conference on*, pages 210 –216, aug. 2009.

[5] R. M. Groves, F. J. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. *Survey Methodology*. John Wiley and Sons, 2009.

[6] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering - a systematic literature review. *Information and Software Technology*, 51(1):7 – 15, 2009.

[7] M. Leotta, F. Ricca, M. Ribaudo, G. Reggio, E. Astesiano, and T. Vernazza. An Exploratory Survey on SOA Knowledge, Adoption and Trend in the Italian Industry. In *Proceedings of 14th International Symposium on Web Systems Evolution (WSE 2012)*, pages 21–30. IEEE, 2012.

[8] M. Leotta, F. Ricca, M. Ribaudo, G. Reggio, E. Astesiano, and T. Vernazza. Soa adoption in the italian industry. In M. Glinz, G. C. Murphy, and M. PezzË, editors, *ICSE*, pages 1441–1442. IEEE, 2012.

[9] J. Li, R. Conradi, C. Bunse, M. Torchiano, O. Slyngstad, and M. Morisio. Development with off-the-shelf components: 10 facts. *Software, IEEE*, 26(2):80–87, 2009.

[10] J. Li, O. Slyngstad, M. Torchiano, M. Morisio, and C. Bunse. A state-of-the-practice survey of risk management in development with off-the-shelf software components. *Software Engineering, IEEE Transactions on*, 34(2):271–286, 2008.

[11] S. L. Pfleeger and B. A. Kitchenham. Principles of survey research: part 1: turning lemons into lemonade. *SIGSOFT Software Engineering Notes*, 26(6):16–18, 2001.

[12] F. Tomassetti, M. Torchiano, A. Tiso, F. Ricca, and G. Reggio. Maturity of software modelling and model driven engineering: a survey in the italian industry. *IET Seminar Digests*, 2012(1):91–100, 2012.

[13] M. Torchiano, M. Di Penta, F. Ricca, A. De Lucia, and F. Lanubile. Software migration projects in italian industry: Preliminary results from a state of the practice survey. In *ASE Workshops*, pages 35–42. IEEE, 2008.

[14] M. Torchiano, M. Di Penta, F. Ricca, A. De Lucia, and F. Lanubile. Migration of information systems in the italian industry: A state of the practice survey. *Information and Software Technology*, 53:71–86, January 2011.

[15] M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, and G. Reggio. Preliminary findings from a survey on the MD* state of the practice. In *International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 372 –375. IEEE, sept. 2011.