The Journal of Systems and Software 94 (2014) 87–97

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

A critical examination of recent industrial surveys on agile method usage

Stavros Stavru*

Sofia University "St. Kliment Ohridski", 5, James Bouchier Str., P.B. 48, 1164 Sofia, Bulgaria

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form
21 December 2013
Accepted 12 March 2014
Available online 8 April 2014

Keywords:
Agile software development
Survey research
Research synthesis

ABSTRACT

Context: Practitioners and researchers often claim that agile methods have moved into the mainstream for the last few years. To support this claim they refer to recent industrial surveys which tend to report high rates of agile method usage. However many of these industrial surveys are conducted by agile consultants, tool vendors, professional societies and independent technology and market research organizations. This raises some important concerns about the possible conflict of interest and the overall trustworthiness of these studies.

Objective: In response to the above concerns, a secondary study was carried out. Its objective was to examine industrial surveys published in 2011 and 2012, determine the extent to which we could trust their reported high rates of agile method usage and provide recommendations on how quality of research could be improved in the future.

Method: Following a rigorous search procedure, nine industrial surveys on agile method usage published in 2011 and 2012 were extracted from both academia and industry. Their thoroughness in reporting and trustworthiness were evaluated using a newly proposed assessment framework based on Guba's four attributes of trustworthiness (truth value, applicability, consistency and neutrality) and a number of methods for assessing survey research in related fields as information, communication and management studies.

Results: The careful examination of the reviewed surveys shows that most of the studies have insufficient thoroughness in reporting and (subsequently) low trustworthiness. Only one (out of nine) study is considered as a scientific contribution in determining the current 2011/2012 rate of agile method usage.

Conclusions: The obtained results support our initial considerations about the trustworthiness of recent industrial surveys on agile method usage and suggest a number of recommendations for increasing the quality and value of future survey research in this regard.

1. Introduction

Agile methods have emerged as an alternative to plan-driven software development methods more than a decade ago (Dingsøyr et al., 2012). Today, they are often *considered as the mainstream in software engineering*. This is often explained with their potential to overcome the *challenges of modern software organizations* which are expected to operate in highly dynamic and competitive environments. In such environments, speed, quality and cost of software development are crucial for organizational survival and agile methods seems to be successfully delivering on all three fronts through their customer focus, responsiveness to change, iterative and incremental delivery of working software and emphasis on individuals and their interactions.[1] To study the popularity of agile methods, many industrial surveys have been carried out. They have reported various *rates of agile method usage* in terms of (1) *the percentage of software professionals/organizations using or moving toward agile methods as compared to alternative methods* like lean methods, plan-driven methods, etc. and (2) *the percentage of software professionals/organizations using specific agile methods* as Scrum, Extreme programming, etc. These rates have being often cited by practitioners and researchers to prove and demonstrate the widespread adoption of agile methods. However, the majority of these surveys are coming from *agile consultants, tool vendors, professional societies and independent technology and market research organizations*

* Tel.: +359 893691583.
E-mail address: stavross@fmi.uni-sofia.bg

[1] http://www.agilemanifesto.org/.

http://dx.doi.org/10.1016/j.jss.2014.03.041
0164-1212/

rather than from academics (Rodríguez et al., 2012). This could be quite *problematic* for at least two reasons: (1) agile consultants/tool vendors/professional societies might not rigorously follow the scientific method while conducting their surveys (e.g. due to time or budget constraints) threatening the trustworthiness of their findings (incl. internal and external validity, reliability, objectivity, etc.); and (2) agile consultants/tool vendors/professional societies might pursue their own private interests in conducting such surveys (e.g. as part of their customer/membership acquisition strategy) that could be in conflict with their findings. This could lead to omitting some important information (e.g. not publishing negative results), altering reported results, etc.

In response to the above concerns a *secondary research* was carried out. Its *objective* was:

> RO: To carefully examine recent industrial surveys on agile method usage, determine the extent to which we could trust their findings in regard to the widespread adoption of agile methods and provide recommendations on how quality of research could be improved in the future.

To achieve this research objective, the study went through a number of *subsequent steps*. First, the *industrial surveys on agile method usage in 2011 and 2012 were extracted* following a rigorous search strategy. Second, the *extracted surveys were assessed for their thoroughness in reporting* – a construct proposed and operationalized by the author for evaluating the available information on how the surveys were carried out (incl. their survey methods, target populations, sampling frames and sizes, response rates, etc.) and used to *select the studies which would be (and were eligible) for further assessment of trustworthiness*. As there was no formal framework for assessing trustworthiness of survey research in software engineering (to the extent of our knowledge), an *assessment framework was proposed* based on Guba's quality model (Guba, 1981) and a number of methods from other related fields as information and communication studies and management studies. The framework was then used to *assess the trustworthiness of the selected studies* and to *provide valuable insights on the quality of survey research in regard to agile method usage*.

The remainder of the paper is *structured* as follows: Section 2 details the search strategy used to identify the industrial surveys on agile method usage in 2011 and 2012. Section 3 presents the findings of the identified studies and discusses some incompatibilities (or inconsistencies) which negatively impact the consolidation (and comparison) of their findings. Section 4 introduces the construct of thoroughness in reporting and applies it in order to select the studies which would (and are eligible to) be further assessed for trustworthiness. Section 5 proposes an assessment procedure for assessing the trustworthiness of the selected studies and uses it to provide insights on the quality of survey research in regard to agile method usage. Section 6 discusses the limitations and threats to validity of the presented study and introduces the actions taken to address them. Section 7 outlines the contributions of the presented study and suggests some important recommendations on how to increase the quality of survey research on agile method usage.

2. Identification of recent industrial surveys on agile method usage

While identifying prior surveys on agile method usage we set *few limitations* (or inclusion/exclusion criteria). First, we focused only on *surveys published between 2011 in 2012*. The reason for this limitation was the actuality and relevance of such surveys in demonstrating and proving the claim that agile methods are currently the mainstream in software engineering. Due to this restriction some widely recognized agile adoption surveys were excluded, including the ones conducted by Forrester Research/Dr. Dobb's (West and Grant, 2010) (probably the first to state that agile methods had joined the mainstream), Gartner (Norton, 2008), Dr. Dobb's and Scott Ambler surveys (Ambler, 2007; Ambler, 2006–2009), and others. Second, following our research objective we included only surveys which investigate *the rate of agile method usage* as compared to alternative methods (e.g. agile vs. lean vs. plan-driven methods) or as compared to each other (e.g. Scrum vs. Extreme programming). Based on this limitation many surveys were excluded although they were covering different aspects of agile usage as practices usage (Kurapati et al., 2012; Ambler, 2012), tools usage (Azizyan et al., 2011), team and developers perceptions of agile usage (Williams, 2012), enablers and barriers to agile usage (Asnawi et al., 2012; Sheffield and Lemétayer, 2012; McHugh et al., 2012; Vijayasarathy and Turk, 2012; Ambler, 2011), effects of agile usage (Rao et al., 2011; Rönkkö et al., 2011) and others. The third and last limitation was that surveys should be *industrial surveys*, meaning that: (1) they are targeted to software professionals and organizations from the software engineering industry (and not to university students for example); and (2) include at least a dozen of organizations (thus excluding single and multicase studies). By applying this third and last limitation few more surveys were excluded as the one conducted at Nokia (Laanti et al., 2011), where the survey population was Nokia's employees and thus was covering only a single (although large and multinational) organization.

Two main *publication sources* where used to extract the surveys. The first one was the *Scopus electronic database*. Scopus is the largest abstract and citation database of academic literature and quality web sources, which ensured the coverage of nearly 20,500 titles from more than 5000 publishers. As such Scopus was mainly used to extract surveys conducted by academics. For surveys conducted by consultants (as Valtech, Xebia, etc.), tool vendors (VersionOne, ThoughtWorks, etc.), professional societies (as Agile Alliance, Scrum Alliance, etc.) and independent technology and market research organizations (e.g. Forrester Research, Gartner, etc.) we used *Google Scholar*. Google scholar allowed us also to cover gray (in the form of theses, technical reports, white papers, etc.) and unpublished literature (as part of web sites, blogs, etc.).

The included electronic databases were searched using the "*agile AND survey*" string. The total number of hits on Scopus (incl. title, abstract and keywords) was 117. Their titles and abstracts were further reviewed based on the exclusion criteria defined in the previous paragraphs. In result only one publication (RO) was eligible for inclusion. As for Google Scholar, the total number of hits was more than ten thousand (15,500). As it was impossible to review all of them, they were sorted by relevance (to the search string) and limited to the first one thousand on the list. From these publications, only eight were included (AL, AM, BA, SE, TW, VO, VT and XB). In order to further reduce the probability of omitting relevant literature (e.g. as we did not include other popular electronic databases as Engineering Village, Web of Science, etc. and limited the review process for Google Scholar), we did *backward* (using the reference lists of the initial publications) *and forward referencing* (using the cited reference searching functionality provided by Scopus and Google Scholar to retrieve publications citing the initial ones) on the final pool of surveys extracted from Scopus and Google Scholar. However, no additional publications emerged.

A total of *9 surveys* were identified to be relevant to the current 2011/2012 rate of agile method usage (excluding repeated instances of the same industrial survey). These studies are summarized in Table 1 ordered by their key alphabetically (formed by the names of its authors and sponsoring organization).

As seen from Table 1 the majority of the extracted studies (6/9) were coming from industry (and more specifically from agile consultants and tool vendors) and only three of them were conducted

Table 1
Surveys examining the rate of agile method usage in 2011 and 2012.

Key	Authors/Organization	Year	Name	Source	Reference
AL	Ali	2012	Survey on the State of Agile Practices Implementation in Pakistan.	Academia	Ali (2012)
AM	AgileMe.co	2011	New Zealand Agile Survey	Industry	Agileme.co (2011)
BA	Baruah and Ashima	2012	A Survey of the Use of Agile Methodologies in Different Indian SMEs	Academia	Baruah Ashima (2012)
RO	Rodríguez et al.	2012	Survey on Agile and Lean Usage in Finnish Software Industry	Academia	Rodríguez et al. (2012)
SE	Serena	2012	Agile 2012 Survey	Industry	Serena (2012)
TW	ThoughtWorks	2011	Agile Adoption in India	Industry	ThoughtWorks (2011)
VO	VersionOne	2011	State of Agile Development Survey	Industry	VersionOne (2011))
VT	Valtech	2012	Agile India 2012	Industry	Valtech (2012)
XB	Xebia	2011	2nd Dutch Agile Survey	Industry	Xebia (2011)

by academics. This is consistent with previous observations on the distribution of agile usage surveys between academia and industry (Rodríguez et al., 2012) and further supports the claim that *most of what we know about the current rate of agile method usage comes from agile consultants and tool vendors*.

3. Current rate of agile method usage as reported by the identified industrial surveys

Except for the reported rate of agile method usage, some additional information was extracted from the identified studies in order to allow *consolidation (as well as comparisons) of their findings*. Such information included the unit of analysis (e.g. whether the survey was examining individuals, teams, departments, organizations, etc.), whether it was targeted to agile practitioners only (e.g. members of agile professional communities) or to the general public, the sampling frame and sampling size of the study (if provided) and finally the number of respondents and response rate. This information is summarized in Table 2.

Many observations could be made from Table 2. One such observation is that the *unit of analysis of almost all studies* (except BA) *was the individual*. Although this could support the comparison of findings between studies, it leaves unanswered a question which we believe would provide a better insight on the use of agile methods within the industry – *the current rate of agile method usage among organizations?* To argument this claim we could consider a hypothetical case where: (1) the majority of individuals are occupied by a minority of organizations – by applying the Pareto principle (Pareto, 1971) this proportion would be 80% of all individuals are occupied by 20% of all organizations (e.g. currently more than 500,000 people are occupied by the top ten largest organizations as ranked by Forbes Global 2000 under the "Software and Programming" industry); and (2) the majority of organizations is using agile methods while the rest of the organizations are not. Then, if the unit of analysis is the individual and everybody takes part in the survey, the rate of agile method usage would be 20%. However, 80% of the organizations would be actually using agile methods. Thus, by setting the unit of analysis to the individual, we might get potentially confusing and misleading results (20% vs. 80%) when the unit of analysis in not explicitly reported or neglected by the researcher. Other arguments why the unit of analysis should be the organization rather than the individual could be found in Rodríguez et al. (2012). Another observation from Table 2 is that *the majority of studies* (88%) *were targeted to agile practitioners only*. As result, the information on the current rate of agile method usage as compared to alternative methods is much more limited (RO and VO) compared to the information on the current rate of agile method usage as compared to each other. Some observations could be made also in regard *to the individuals who participated in the surveys*. Assuming no duplications, the total number of respondents was 7824 with a mean of 1118 respondents, a median of 249 respondents and a standard deviation of 2185 respondents. Such a big variance (and difference between the mean and the median) indicates that the representativeness of the surveys on agile method usage could significantly vary from one study to another. For example, VO has the greatest number of respondents (total of 6042), while studies as AM and AL have considerably less respondents (31 and 78 respectively). As such (and all else being equal), VO should be considered more representative than AM and AL.

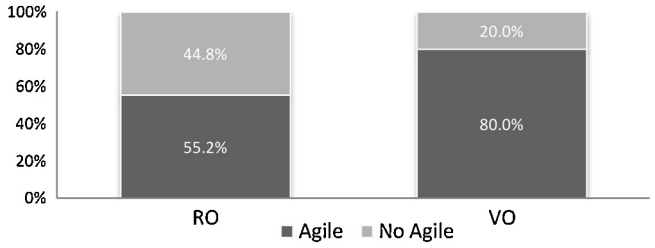


Fig. 1. Current rate of agile method usage as compared to no agile (or alternative) methods.

3.1. Current rate of agile method usage as compared to alternative methods

There were two studies (RO and VO) which provided information on the current rate of agile method usage as compared to alternative methods. According to *RO*, 55.2% of the respondents have reported that *their organizational unit was using agile methods* (more specifically 33.6% were using only agile methods while 21.6% were using a combination between agile and lean methods). As for the use of alternative methods, 2.7% of the respondents were using only lean methods and 42.2% of the respondents were using alternative methods. There was no additional information on the distribution of alternative method usage (e.g. between waterfall, software prototyping, iterative and incremental development, spiral model, RUP, MSF, etc.). As for the VO, 80% of the respondents have reported that *their organizations have adopted agile development practices*. 15% were using alternative methods, while 5% were not sure if their organization was using agile methods or not. Similar to RO, VO also did not provide any information on alternative method usage. The rates of agile method usage as compared to alternative methods are further shown in Fig. 1.

From Fig. 1 we could see that there is a significant difference in the current rate of agile method usage as reported by RO and VO – the rate by VO is almost 25% higher than the one by RO. However, it is *quite difficult to make any comparison between these studies* for a number of reasons. One such reason is the *organizational level at which respondents were asked to give information on agile method usage*. RO was asking its respondents for agile method usage within their organizational unit (which could be their team, department, etc.), while VO was interested in agile method usage within their whole organization. To demonstrate the complications that could result let us consider a case where we have a respondent who is part of an organizational unit (or team) which is not using agile methods. However his organization has organizational units which

Table 2
Additional information about the identified studies.

Study	Unit of analysis	Target population	Sampling frame	Sample size	Respondents	Organizations	Response rate
AL	Individual	Agile practitioners	–	–	78	–	–
AM	Individual	Agile practitioners	–	–	31	–	–
BA	Organizational	Agile organizations	–	–	–	18	–
RO	Individual	SE practitioners	16,000	4450	408	200	9%
SE	Individual	Agile practitioners	–	–	–	–	–
TW	Individual	Agile practitioners	–	–	770	330	–
VO	Individual	SE practitioners	–	–	6042	–	–
VT	Individual	Agile practitioners	–	–	249	105	–
XB	Individual	Agile practitioners	–	–	246	–	–

are using agile methods. Then, if asked, the respondent would give a negative answer for RO (as his organizational unit is not using agile methods) and positive for VO (as his organization is using agile methods). This difference in the organizational level at which information was obtained could be one possible explanation for the big difference of 25% in the rates reported by the two studies. Comparisons are difficult also because of *the lack of consistency in the way agile method usage was defined*. VO defined agile method usage as the adoption of agile development practices within the organization. Such a definition includes even organizations which are using only a single agile practice (as Refactoring from Extreme programming) and promises high rates of agile method usage (as the one reported by VersionOne). As for RO, agile method usage was not explicitly defined, leaving respondents to understand it as they wish. This significantly complicates the interpretation of its findings and makes comparisons with other studies very difficult (and even impossible).

3.2. Current rate of agile method usage as compared to each other

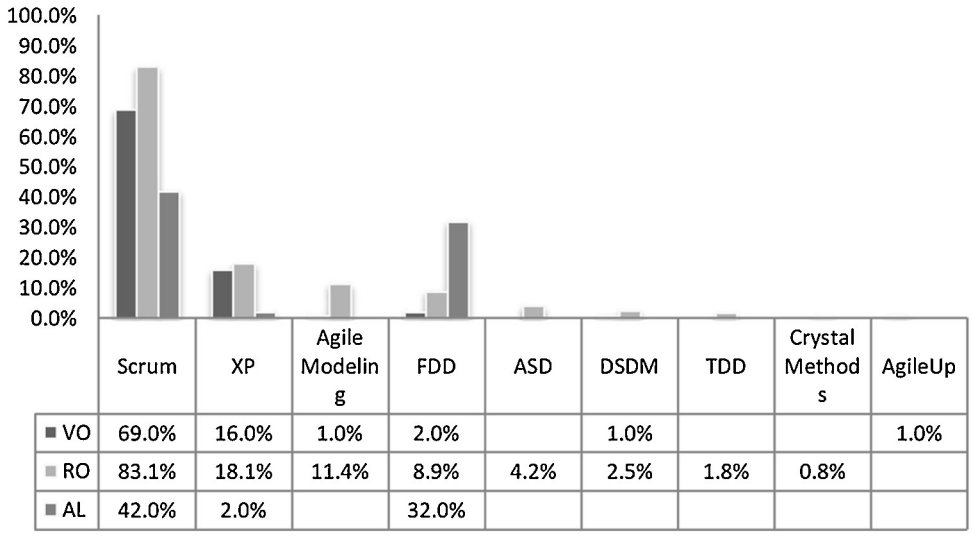
All studies were examining the current rate of agile method usage as compared to each other. However, only three studies (RO, VO and AL) were included in the subsequent analysis as they provided sufficient information on the distribution of agile methods among agile organizations. According to RO, the most popular agile method was Scrum which was used by 83.1% of all respondents. Similar results were obtained by VO and AL, although their percentage was lower – 69% and 42% respectively. The second most popular agile method according to RO and VO was Extreme Programming with 18.1% and 16% respectively. This percentage for AL was 2% – much lower than the one reported by RO and VO. Other popular agile methods were Agile Modeling (11.4% in RO and 1% in VO) and Feature Driven Development (8.9% in RO, 2% in VO and 32% in AL). Rarely used agile methods included Adaptive Software Development (4.2% in RO), Dynamic System Development Method (2.5% in RO and 1% in VO), Crystal Methods (0.8% in RO) and AgileUP (1% in VO). The rates of agile method usage as compared to each other are further shown in Fig. 2.

As seen from Fig. 2 there is no big difference in the rates reported by RO and VO. However, the rates in AL are quite different. One possible explanation for this could be the low number of respondents in AL (only 78 respondents) which could bias the results. Another explanation could be the specific context in which these surveys were carried out – AL examined the agile method usage within the Pakistanian software industry, RO within the Finish software industry and VO worldwide.

To conclude this section we could say that the findings of the identified studies in regard to the current rate of agile method usage agree on the widespread adoption of agile methods and indeed support the claim that they have moved into the mainstream. However, the extent to which we could trust these findings has to be considered as well.

4. Thoroughness in reporting of the identified industrial surveys

Before we could continue with the careful evaluation of trustworthiness of the identified studies, a *certain level of thoroughness in reporting* was required (e.g. describing the method used to conduct the survey, its sampling frame, sample size, response rate, etc.). Verifying that such information exists was a prerequisite for



	Scrum	XP	Agile Modeling	FDD	ASD	DSDM	TDD	Crystal Methods	AgileUp
VO	69.0%	16.0%	1.0%	2.0%		1.0%			1.0%
RO	83.1%	18.1%	11.4%	8.9%	4.2%	2.5%	1.8%	0.8%	
AL	42.0%	2.0%		32.0%					

Fig. 2. Current rate of agile method usage as compared to each other.

selecting the studies that should be further analyzed in regard to their trustworthiness.

In order to evaluate the *thoroughness in reporting* (or just *thoroughness*) we looked at the existing knowledge on how surveys should be conducted in software engineering. We found three works which were relevant to the topic–the series of articles on survey research by Kitchenham and Pfleeger (Kitchenham and Pfleeger, 2001–2003), a handbook for designing effective surveys from Kasunic (Kasunic, 2005), and a process for preparing, conducting and analyzing surveys proposed by Ciolkowski et al. (2003). However, we found no framework for evaluating survey research in software engineering. For that reason we referred to highly cited frameworks from other related fields including information and communication studies (Pinsonneault and Kraemer, 1993; Shenton, 2004) and management studies (Malhotra and Grover, 1998). We used these frameworks, together with the three relevant works from software engineering (Kitchenham and Pfleeger, 2001–2003; Kasunic, 2005; Ciolkowski et al., 2003) as a theoretical background for extracting the *minimum information required* for evaluating trustworthiness of industrial surveys. Also, some *common characteristics of industrial surveys* were taken into account, including that: (1) industrial surveys identify certain characteristics of the situation under study (or describe the situation *as it is*) and did not involve changing/modifying the situation or determining cause-and-effect relationships–therefore they represent *descriptive and exploratory* (rather than explanatory) *type of surveys*; and (2) industrial surveys are targeted to a large sample of the population and often use self-administrated online questionnaires with close-ended multiple choice questions which are easy to quantify through statistical, mathematical or computational techniques–and therefore they represent *quantitative* (rather than qualitative) *type of surveys*. These two specific characteristics had a significant impact on how thoroughness in reporting (and trustworthiness) was defined and operationalized.

The information requirements extracted from the reviewed frameworks were further examined in regard to the specific characteristics of industrial surveys and finally consolidated into a list of *21 relevant criteria*. These criteria were then *weighted* from 1 (being the least important) to 5 (being the most important). The weights were not arbitrary set by the author but were based on the frameworks that were used to extract the criteria. These frameworks emphasized the importance of specific information for the evaluation of trustworthiness and the weights were needed in order to reflect that difference in importance. For example all criteria with weight 5 (as sampling frame, sampling method, sampling size, response rate, etc.) were defined by these frameworks as a must in order to even consider the evaluation of trustworthiness for a given study. On the other side, criteria with weight 1 (incl. sponsorship, objectives, media, response burden, etc.) were considered "good to have". The criteria, their weights and the studies which satisfy them are shown in Table 3 (grouped by their relevant phase of survey research).

As seen from Table 3 the *information provided by most of the identified studies is scarce*. The sampling method, which was weighted as highly important information, was not (explicitly) specified by any of the identified studies. The same is valid for the assessment of trustworthiness and how the trustworthiness of the studies was secured, as well as for specifying and describing the conceptual model of the survey. In regard to other highly important information as the sampling frame, sample size and response rate, there was only one study providing such information. This is quite problematic as without this information it is quite difficult (and even impossible) to draw any conclusions on the trustworthiness of the rest of the identified studies (or 8/9 of all studies). On the other hand the information that was most often available was the number of responses received (by 7/9 of the surveys), the method used to collect data (6/9), how long the survey was available to respondents (7/9) and its sponsorship (7/9). Alhough this information was considered important, it was very insufficient in terms of carefully assessing the trustworthiness of the identified studies.

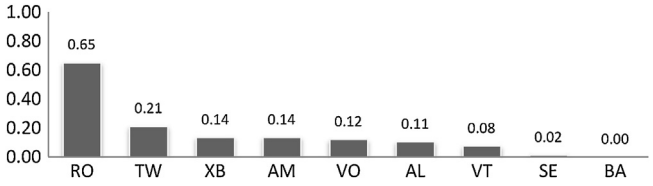


Fig. 3. Thoroughness of each of the identified studies.

Using the criteria and weights from Table 3, the thoroughness for each of the identified studies was *calculated* using the following *procedure*: (1) all criteria satisfied by a study were marked (using Table 3 as a checklist); (2) the weights of these marked criteria were summed up; and (3) the sum was divided by the total sum of weights of all criteria. The resulted number was between 0 and 1. The closer the number was to 1 the higher the thoroughness was. In order to initially evaluate the proposed procedure for calculating thoroughness, it was applied on the studies presented by Ciolkowski et al. (2003). These studies were following a formal methodology for conducting survey research and could be considered as good examples on how surveys should be reported in software engineering. Therefore, they were expected to get high scores in terms of thoroughness. The first survey (the *ISERN Survey*) scored 0.92 (information on the follow-up procedures and the questionnaire itself was missing), while the second survey (the *ViSEK Survey*) scored 0.95 (only the questionnaire was not provided). These results were consistent with our expectations and partially supported its predictive validity. However more rigorous procedures are needed in order to further assess its reliability (e.g. test–retest reliability, inter-rater reliability, etc.) and validity (e.g. content validity, criterion validity, experimental validity, etc.). The same is valid for the construct of thoroughness as defined by the criteria in Table 3.

The *thoroughness of each of the identified studies* is shown in Fig. 3.

As seen from Fig. 3, the study with the greatest thoroughness was RO with a score of 0.65. The rest of the reviewed studies scored less than 0.25 with three studies scoring less than 0.10. This data was further examined depending on whether the studies were coming from academia or industry. The studies conducted by academics scored 0.65 (Rodríguez et al., 2012), 0.11 (Ali, 2012) and 0.00 (Baruah Ashima, 2012), resulting in a mean of 0.25, a median of 0.11 and a standard deviation of 0.35. The same numbers for industry were 0.12, 0.13 and 0.07 respectively, with ThoughtWorks scoring the greatest thoroughness of 0.21 (ThoughtWorks, 2011). Comparing these descriptive statistics and taking into consideration outliers as RO, we could say that the thoroughness of the studies coming from industry is greater than the one from academia. This contradicts with our initial expectations (that the thoroughness of academic studies would be greater than industrial studies) and suggests that surveys on agile method usage could be reported by academics in a sufficient detail (RO) or could be poorly reported (AL) or no reported at all (BA). The great variance (0.35) in thoroughness of these studies (ranging from 0.65 to 0.00) could be partially explained with their publication sources and the criteria and procedures used to evaluate them for publication. RO was published (as a conference proceedings) in a reputable ACM/IEEE forum – the International Symposium on Empirical Software Engineering and Measurement (ESEM), which is the premier conference on empirical software engineering, popular for its high quality standards and rigorous procedures in evaluating empirical research (incl. survey research). As for AL and BA, although they were published in

Table 3
Criteria for thoroughness and surveys which satisfy them.

Criteria for thoroughness	Description	Weight	Studies
Survey definition			
Objectives	The study *defines explicitly its objectives*.	1	TW, RO
Sponsorship	The study *clarifies its sponsorship (or the organizations who are funding the survey) and their interests* in conducting the survey.	1	VO, XB, TW, AM, VT, RO, AL
Survey method	The study *specifies and thoroughly describes* how the survey was conducted (in terms of its phases, settings and context, etc.).	4	RO
Survey design			
Conceptual model	The study *specifies and thoroughly describes its conceptual model* (in terms of objects that are investigated, variables and expected relationships between them, etc.)	4	–
Target population	The study *specifies and thoroughly describes its target population* (in terms of unit of analysis, reporting unit, exclusion/inclusion criteria, sources, etc.).	4	TW, RO
Sampling frame	The study *specifies and thoroughly describes its sampling frame* – the actual set of units from the target population from which a sample would be drawn (or lists all those within the target population who can be sampled). For example the target population might be defined as all organizations which are developing, maintaining or integration software products and services in a given region. However the sampling frame might be restricted to these organizations which have an official (as the survey would be mediated by email). The information on the sampling frame should include at least *the number of units to be sampled*.	5	RO
Sampling method	The study *specifies and thoroughly describes its sampling method* (e.g. non-probabilistic sampling methods, probabilistic sampling methods, etc.).	5	–
Sample size	The study *defines its sample* (in terms of sample size).	5	RO
Data collection method	The study *specifies and thoroughly describes its data collection method* (e.g. interviews, self-administrated questionnaires, etc.).	3	VO, XB, TW, AM, RO, AL
Questionnaire design	The study *describes how the questionnaire was designed* (e.g. the number of questions, type and wording of the questions, question sequence and grouping, translations, etc.).	4	RO
Provisions for securing trustworthiness	The study describes the provisions made to secure trustworthiness (e.g. adoption of appropriate, well recognized research methods, examination of previous research findings, etc.)	3	–
Survey implementation			
Questionnaire evaluation	The study provides information on *how the questionnaire was evaluated* (e.g. through piloting, focus groups, in-depth interviews, statistical methods, etc.).	3	RO
Questionnaire	The *questionnaire* of the study is *available* (e.g. attached to the report or included as an appendix, etc.).	3	–
Survey execution			
Media	The study *describes how the survey was mediated* (e.g. through mail, e-mail, telephone, web, etc.).	1	XB, TW, AM, RO
Execution time	The study *specifies how long the survey was available to respondents*.	1	VO, XB, TW, AM, SE, VT, RO
Response burden	The study *specifies how long the survey took to fill out by respondents*.	1	–
Follow-up procedures	The study *specifies the procedures taken in order to encourage response and prevent non-response*.	2	–
Responses	The study *provides information on the number of responses received*.	3	VO, XB, TW, AM, VT, RO, AL
Response rate	The study *provides information on its response rate*.	5	RO
Survey analysis and packaging			
Assessment of trustworthiness	The study *formally assesses its trustworthiness* (e.g. through calculating measurement error, sample frame error, error of selection, non-response error, etc.).	5	–
Limitations	The study *describes its limitations and threats to validity*.	3	RO

international journals, these journals are not specifically focused on empirical research (or agile software development) neither they are among the high-ranked journals in software engineering. Following this line of thought, we would recommend publication sources which are specifically focused on empirical research and are reputable enough to guarantee a greater extent of thoroughness in reporting and trustworthiness of published studies.

The *minimum level of thoroughness* required for each of the identified studies in order to be further analyzed was set to 0.62. This number was calculated by summing up the weights of all criteria from Table 3 with a weight of 4 (important) or 5 (highly important), and dividing the result by the total sum of weights of all criteria. Such level of thoroughness guaranteed that the extracted study could be at least partially assessed for its trustworthiness. From the identified studies, only the work by Rodríguez et al. (2012) had a thoroughness greater than 0.62. As the trustworthiness of the rest of the studies could be hardly assessed objectively, these studies were excluded from the subsequent analysis. We did only one exception by including the survey by VersionOne as it is being conducted yearly since 2006, making it the most longitudinal survey on agile method usage.

5. Trustworthiness of the selected industrial surveys

The trustworthiness of the selected studies (RO and VO) was assessed using Guba's four attributes of trustworthiness relevant

Table 4
Mapping between the criteria for thoroughness and Guba's four attributes of trustworthiness.

Criteria for thoroughness	Attribute of trustworthiness	RO		VO	
		Value	Effect	Value	Effect
Survey definition					
Objectives	Neutrality	Relevant	+	Not specified	–
Sponsorship	Neutrality	Sponsorship with no conflict of interest	+	Sponsorship with possible conflict of interest	–
Survey method	Consistency	Specified	+	Not specified	–
Survey design					
Conceptual model	Truth value	Not specified	–	Not specified	–
	Consistency	Not specified	–	Not specified	–
Target population	Truth value	Adequate	+	Not specified	–
	Consistency	Specified	+	Not specified	–
Sampling frame	Applicability	Adequate	+	Not specified	–
	Consistency	Specified	+	Not specified	–
Sampling method	Applicability	Non-probabilistic sampling	–	Non-probabilistic sampling	–
	Consistency	Not specified	–	Not specified	–
Sample size	Applicability	Adequate	+	Not specified	–
	Consistency	Specified	+	Not specified	–
Data collection method	Truth value	No triangulation	–	No triangulation	–
	Consistency	Specified	+	Specified	+
	Neutrality	No triangulation	–	No triangulation	–
Questionnaire design	Truth value	Adequate	+	Not specified	–
Provisions to secure trustworthiness	Truth value	Not specified	–	Not specified	–
	Applicability	Not specified	–	Not specified	–
	Consistency	Not specified	–	Not specified	–
	Neutrality	Not specified	–	Not specified	–
Survey implementation					
Questionnaire evaluation	Truth value	Adequate	+	Not specified	–
Questionnaire	Truth value	Not specified	–	Not specified	–
	Consistency	Not specified	–	Not specified	–
Survey execution					
Media	Consistency	Specified	+	Not specified	–
Execution time	Consistency	Specified	+	Specified	+
Response burden	Consistency	Not specified	–	Not specified	–
Follow-up procedures	Consistency	Not specified	–	Not specified	–
Responses	Applicability	Response rate less than 20%	–	Sample size not specified	–
Response rate	Applicability	Response rate less than 20%	–	Not specified	–
Survey analysis and packaging					
Assessment of trustworthiness	Truth value	Not specified	–	Not specified	–
Limitations	Neutrality	Objectively considered	+	Not specified	–

to both quantitative and qualitative studies (Guba, 1981). These attributes are: (1) *truth value* (or internal validity and credibility) – the extent to which the study *measures or tests what is actually intended*; (2) *applicability* (or external validity, generalizability and transferability) – the extent to which the findings of the study *could be applied to other situations* (incl. other contexts, other respondents, etc.); (3) *consistency* (or reliability and consistency) – the extent to which *similar results would be obtained if the study is repeated*; and (4) *neutrality* (or objectivity and confirmability) – the extent to which *the findings of the study are the product of the focus of the inquiry* and not of the biases of the researcher. These four quality attributes were further mapped to their relevant criteria for thoroughness (from Table 3). The reasons for using this mapping rather than using any of the existing assessment frameworks were: (1) achieving *continuation* by reusing the information already extracted and available during the evaluation of thoroughness; (2) achieving *specification* by focusing only on information that is specific to industrial surveys rather than to surveys in general (as in the case of the existing evaluation frameworks); and (3) dealing with *limited information* which is typical for industrial surveys addressed to (and often conducted by) practitioners who are mostly interested in the findings of the survey rather than in how the survey was conducted or how its quality was secured and assessed (Table 3). The mapping was done based on the work of Shenton (2004) and Malhotra and Grover (1998) and evaluated: (1) *truth value* through determining the adequacy of the conceptual model and the target population (incl. its unit of analysis, reporting unit, etc.), the survey questionnaire (incl. its design, evaluation and formulation), the methods used to collect data and to secure and assess its quality, etc.; (2) *applicability* through determining the adequacy of the sampling frame, the sampling method, the sampling size, the number of responses and the response rate; (3) *consistency* through determining the thoroughness of describing the survey method (incl. its phases, settings and context) and various characteristics of survey design (incl. its target population, sampling frame, sampling method, sampling size and questionnaire), survey preparation (incl. its data collection method and media) and survey execution (incl. its execution time, response burden and follow-up procedures); and (4) *neutrality* through determining the relevance of survey objectives, the possible conflict of interest, the use of triangulation (or multiple data collection methods) and the objectivity in considering survey limitations. The mapping between the criteria for thoroughness (Table 3) and Guba's four quality attributes is shown in Table 4. Table 4 also presents some additional information for each of the assessed studies (RO and VO) in terms of their criteria for thoroughness (e.g. the use of probabilistic vs. non-probabilistic sampling for sampling method) and the effect these criteria have on their relevant quality attributes (either positive or negative).

Using the data from Table 4, *each of the four attributes of trustworthiness of the assessed studies was evaluated* using the following *procedure*: (1) all criteria for thoroughness relevant for a given quality attribute were obtained; (2) from the obtained criteria, the criteria which had positive effect on the given quality attribute, were marked; and (3) the marked criteria (from step 2) were

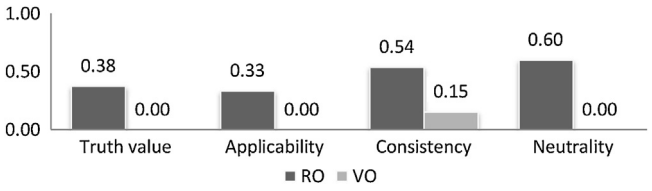


Fig. 4. Trustworthiness of the assessed studies.

count and divided by the total number of criteria relevant for the given quality attribute (from step 1). The result was a number between 0 and 1. The closer the number was to 1 the higher the trustworthiness was in regard to the given quality attribute. The trustworthiness of the assessed studies (RO and VO) is shown in Fig. 4.

As seen from Fig. 4, the trustworthiness of VO is low (with 0 for truth value, applicability and neutrality), while the trustworthiness of RO is medium to high (high for neutrality and consistency and medium for applicability and truth value). Using Fig. 4 we could make a number of interpretations of the reported data:

- *We cannot make any generalization on the reported rate of agile method usage* – The reason for that is the low to medium applicability of the assessed studies. The applicability of VO is zero due to the lack of crucial information as its sampling frame, sampling size and its response rate, as well as the use of non-probabilistic (convenience) sampling method. For RO applicability is 0.33 mostly due to the non-probabilistic (judgmental) sampling method and the low response rate (9%). Thus, by taking into consideration the characteristics of the assessed studies in terms of applicability, we could say nothing about the current rate of agile method usage in general (or beyond the concrete set of respondents/contexts of the assessed surveys).
- *It is highly questionable whether the construct of agile method usage was defined and operationalized (or measured) consistently* – The reason is the low to medium truth value of the assessed studies. For VO truth value is zero again due to missing information in regard to its conceptual model and target population, the questionnaire used (incl. how it was designed, evaluated, etc.), etc., while for RO is 0.38 mostly due to the lack of information on its conceptual model and how the quality of the collected data was secured and assessed. To further show how the lack of information on the definition and operationalization of agile method usage could create confusion and misinterpretation (threatening its truth value) let us consider the following possible interpretations of agile method usage (valid for both studies): (1) strictly following any of the existing agile methods as Scrum, Extreme programming, etc.; (2) applying the majority of the processes and techniques prescribed by a given agile method; (3) applying at least one of the processes and techniques prescribed by a given agile method; (4) following the values and principles of agile software development in a custom development method; and many more. Thus to reduce confusion and misinterpretation, the study is expected to clearly define and operationalize the construct of agile method usage and provide relevant information in this regard as part of its questionnaire or other publically available resources as survey instructions, cover letters, reports, etc. For both surveys (RO and VO) no such information was found (neither their questionnaire was available).
- *Objectivity in conducting and presenting surveys on agile method usage could significantly vary* – RO, the study coming from academia, is highly objective (with neutrality of 0.60). This means that the findings of the study are mostly product of the focus of the inquiry and not biased by the researcher (who could intentionally or unintentionally alter the results, omit important information, etc.). This is not valid for VO, which neutrality is zero. The reason for such a low neutrality is again the lack of information (e.g. what the objectives of the study are, what are its limitations, etc.), the lack of triangulation and the possible conflict of interest.
- *Recent surveys on agile method usage could be difficult to replicate and even if partially replicated, obtaining similar results could be questionable* – This is specifically valid for VO, which consistency is 0.15. Missing important information as the survey method used, its conceptual model, the definition of the target population, the sampling frame/method/size, etc. makes any replication almost impossible. As for RO, its consistency is high enough (0.54) to allow the study to be at least partially replicated. However, additional information as its conceptual model, the questionnaire itself, follow-up procedures used to increase the response rate and address non-response bias, the response burden, etc. could be crucial for obtaining similar results.

The results from Fig. 4 further support our argumentation (from the previous section) that industrial surveys should be considered for their trustworthiness only if they have certain level of thoroughness in reporting. If the thoroughness of a survey is very low then it could be expected that its trustworthiness would be very low as well. This is because any information, relevant to trustworthiness, if missing is considered to have a negative impact on trustworthiness. However, sometimes we might be interested in evaluating only one of the quality attributes proposed by Guba and skip the rest (e.g. we might be interested only in the generalizability of survey findings). In this case we could evaluate thoroughness in reporting only in terms of this specific quality aspect, including these thoroughness criteria which are relevant for it (using Table 3). Nevertheless, we strongly recommend taking into consideration all four attributes of trustworthiness when evaluating the trustworthiness of a given industrial survey.

The results from Figs. 3 and 4 also *support* our *initial considerations* in regard to the trustworthiness *of recent surveys on agile method usage.* From all of the identified studies, only the study by Rodríguez et al. (2012) turned out to have sufficient thoroughness in reporting and trustworthiness in order to be considered (taking also its limitations in terms of applicability and truth value) as a scientific contribution in determining the rate of agile method usage in 2011 and 2012.

6. Limitations

The presented study has its recognized limitations and threats to validity. One such limitation comes from the publication sources and the search strategy used to extract the surveys on agile method usage. It could be that some studies on agile method usage could have been omitted either because there were not indexed by the used electronic databases (Scopus and Google Scholar) or because of the limited coverage of the search algorithms and terms. As result, the findings of this study could be subject to *publication bias* (Song et al., 2010). In order to mitigate the risk of publication bias some additional actions were taken, including: (1) the use of heterogeneous electronic databases (Scopus for extracting surveys conducted by academics and Google Scholar for surveys conducted by consultants, tool vendors, professional societies and independent research and marketing organizations); (2) the use of Google Scholar to cover gray and unpublished literature and thus reduce the gray literature bias (Auger, 1998) and the file drawer effect (Rosenthal, 1979); and (3) the use of first-level backward and forward referencing on the final pool of surveys extracted from Scopus and Google Scholar. These actions should have reduced the probability that the omitted surveys would have contained information

that would critically alter the findings of the presented study and threaten its *generalizability* (or *external validity*).

Another limitation is the *narrowed time scope of 2011 and 2012* which *delimited the selection of potential surveys to analyze* – especially in the case of academic studies, where we had only three publications, representing both ends of the quality scale (and no surveys in between). The motivation for this time scope was: (1) our specific objective to critically and objectively examine recent industrial surveys on agile method usage in order to provide valuable insights on their quality and suggest recommendations for future improvements; and (2) the actuality and relevance of these surveys in determining the rate of agile method usage at the moment of the study (end of 2012). However the limited number of academic studies eligible for analysis was a valid concern and had to be further examined. To do so we applied the search strategy in Scopus for 2008–2010. It returned total of 140 publications with no publication eligible for inclusion. This revealed that relevant academic studies were even more lacking between 2008 and 2011, and further increased our confidence that the used time scope was adequate for the purposes of the presented study.

Another limitation of the presented study comes from the fact that the identification, the selection and the assessment of the surveys on agile method usage were conducted by a single researcher (the author of this paper). This could result in *researcher (experimenter)* and *reporting bias* thus threatening the *internal validity* of the study. Furthermore, for studies coming from industry, the data was extracted mostly from publically available (or official) reports which could be intentionally missing a lot of information (in terms of how the survey was conducted and its quality secured) as they are addressed to the general public. As the organizations were not contacted to provide additional information, the *reported trustworthiness could be less than it actually is*. Other *threats to internal validity* come from the construct of thoroughness used to select the surveys (in terms of the criteria for thoroughness, their weights, etc.), the operationalization of this construct (in terms of used formula, minim level of thoroughness, etc.) and the mapping between the criteria for thoroughness and the quality attributes proposed by Guba (Guba, 1981). As they were not rigorously assessed for validity (as content validity, criterion validity, experimental validity, etc.) and reliability (as test–retest reliability, inter-rater reliability, etc.), the presented study could be subject to *measurement bias*. In order to reduce the effects of measurement bias the following actions were taken: (1) a solid theoretical background was used in defining and operationalizing the construct of thoroughness and its further mapping to the quality attributes proposed by Guba (Guba, 1981); and (2) the predictive validity of thoroughness in reporting was initially (and positively) assessed by applying it on the two studies presented by Ciolkowski et al. (2003). Despite these limitations the thoroughness in reporting and trustworthines fit the purpose of the presented study as they provide a systematic way to assess the quality of survey research on agile method usage and make recommendations based on it (and not constructing, presenting and validating a formal assessment framework for survey research in general).

Although there are some limitations and threats to validity, the presented study objectively and critically examines recent industrial surveys on agile method usage, and further presents and discusses some of their shortcomings in terms of thoroughness in reporting and trustworthiness. As such it lays the ground for increasing the quality of survey research in regard to agile method usage.

7. Conclusions

The research objective of the presented study was to carefully examine recent industrial surveys and determine the extent to which we could trust their reported high rates of agile method usage. In doing so it went through a number of *subsequent steps*: (1) the literature review extracted a total of 9 surveys on agile method usage conducted in 2011 and 2012; (2) by using the newly proposed construct of thoroughness in reporting, the thoroughness for each of the extracted studies was assessed and few studies were selected to be further evaluated for trustworthiness; and (3) by using the proposed mapping between the criteria for thoroughness and the quality attributes by Guba (Guba, 1981), the *trustworthiness of the selected studies was further assessed and discussed*. Except for *critically examining recent industrial surveys on agile method usage* and *providing valuable insights on the quality of survey research in this regard*, *the contributions of the presented paper* include also the *definition and operationalization of the construct of thoroughness in reporting (or thoroughness), the mapping between thoroughness and the four quality attributes by Guba* (Guba, 1981) and the *promotion of this mapping into a new framework for assessing trustworthiness* of industrial surveys. The latter could be used for *assessing industrial surveys in a more general context* as well as a baseline for improving the quality of survey research in regard to agile method usage.

By assessing and discussing the thoroughness in reporting and trustworthiness of recent empirical surveys on agile method usage the present study is part of a broader debate on research challenges in software engineering. In their vision of software engineering research Sjoberg et al. (2007) argument that the quality of existing empirical studies (incl. survey research) should be increased and present a list of common issues that should be targeted in the near future (incl. the lack of sufficient detail when reporting empirical studies and various quality issues related to determining the scope of validity, securing internal and external validity, etc.). To achieve this the authors advise the provision of concrete guidelines and recommendations for conducting and assessing empirical research and emphasize that such are currently lacking for case studies, action research, surveys and theory building. Furthermore they highlight the importance of strengthening the collaboration between academia and industry in order to do so. Similar roadmap for software engineering research is proposed by Perry et al. back in 2000. They have stated that "If we want empirical studies to improve software engineering research and practice, then we need to create better studies and we need to draw more credible conclusions from them" – something which is still valid today as it is shown by recent examinations of existing empirical research in software engineering (Selby, 2007; Parnas and Curtis, 2009; Cruzes and Dyb, 2010; Weyuker, 2011; Wieringa, 2012; Jain et al., 2013). As for the specific case of agile software development, the research challenges are even more acute. Dybå and Dingsøyr (Dybå and Dingsøyr, 2008) in their systematic review of existing empirical studies of agile software development have identified that one of the major research challenges "is to increase the quality of studies". From their initial sample of 270 articles, only 33 primary and 3 secondary studies (or 13% of all articles) fulfilled their 11 criteria for quality (in terms of rigor, credibility and relevance). Further analysis of these studies revealed that they are often "not well described, issues of bias, validity, and reliability were not always addressed; and methods of data collection and analysis were often not explained well". Later on they have included quality as one of the most important areas for improvement in their preliminary roadmap for empirical research on agile software development (Dingsoyr et al., 2008). In a more recent study Dingsøyr et al. (2012) have examined how research on agile has progressed in the last decade and showed that the number of studies has significantly increased in the recent years. This is also valid for studies published in high-ranked journals, which could be a sign of increase in quality. However quality remains as one of the major research challenges (Dingsøyr et al., 2012).

Following the broader debate on research challenges in software engineering and based on the critical examination of recent industrial surveys on agile method usage the presented study suggests a number of recommendations:

- As the unit of analysis of almost all of the identified studies was the individual, more studies are needed to examine the rate of agile method usage on an organizational level;
- *More studies are also needed to examine the rate of agile method usage as compared to alternative methods* (e.g. agile vs. lean vs. plan-driven) as their number is very limited at this moment (only two or 22% of the identified studies);
- *Special attention should be paid when building the conceptual model of future survey research* (incl. the thorough definition and operationalization of the construct of agile method usage – e.g. by using or modifying existing scores as the Scrum-Butt Test, Nokia Test for Scrum (Little, 2007), How Agile Are You – A 42 Point Test (Waters, 2008), etc. or determining various agile levels as the defined in the Agile Adoption Framework (Sidky et al., 2007), Agile Adoption and Improvement Model (Qumer and Henderson-Sellers, 2008), etc.) *in order to allow consolidation (and comparison) of findings* (something which could be hardly done for the identified studies);
- *More studies should be conducted by academics* to uniform the distribution of surveys between academia and industry, and increase the confidence in the widespread adoption of agile method usage;
- *More detailed information should be provided in future reports* (e.g. covering at least the criteria from Table 3 with weight of 4 or 5) in order to increase the confidence in the reported studies and allow the objective assessment of their trustworthiness;
- *Special provisions should be taken to increase the applicability of surveys on agile method usage* (e.g. defining adequate sampling frame and size, using probabilistic sampling, achieving high response rate and minimizing non-response bias, etc.) so generalizations of their findings are possible beyond the specific set of their respondents and contexts;
- *Special provisions should be taken to increase the truth value of surveys on agile method usage* (e.g. through building adequate conceptual model, following appropriate, well recognized research methods, using triangulation to cross validate results, minimizing measurement error, etc.) in order to increase the confidence that they examine what they actually intended;
- *Special provisions should be taken to increase the consistency of surveys on agile method usage* (e.g. through in-depth methodological description) so replications of these studies are possible and obtaining similar results are at least partially guaranteed;
- *Special provisions should be taken to increase the objectivity of surveys on agile method usage* (e.g. through thorough description of their objectives, motivation and sponsorship, recognition of their limitations and threats to validity, in-depth methodological description, etc.) in order to ensure that their findings are not biased by the individuals or organizations conducting them.

The findings of the presented study *supported our initial considerations about the trustworthiness of recent industrial surveys* on agile method usage. However this does not mean that agile methods have or have not moved into the mainstream. There are many other indicators that should be considered as well, including: (1) the *increasing research on agile software development* (evident by the number of publications and specialized scientific conferences, workshops, sessions, etc.) (Dingsøyr et al., 2012); (2) the *significant body of professional literature currently available* (e.g. a simple search of "agile software development" on Amazon.com returns more than 900 results); (3) *active professional communities* which are organizing a lot of agile related events, including group gatherings, conferences, workshops, camps, etc. (e.g. Agile Lean Europe, Agile Australia, Agile Central Europe, Scandinavian Agile, Agile Eastern Europe, Agile India, Agile Open Spain, Italian Agile Day, Turku Agile Day, Agile Cambridge UK, Agile Testing Days, Agile Coach Camp and many more); (4) the *increasing number of successful stories* about agile adoption coming from some of the largest organizations in the software engineering industry (incl. IBM, Microsoft, SAP, Symantec, Adobe Systems, Google, Apple, Siemens, Cisco Systems, etc.); (5) the *increasing number of traditional and very conservative organizations* which are *favoring agile methods* (as the UK government and DoD of USA); (6) the fact that *Project Management Institute is now certifying agile project managers*; and many more. But high quality industrial surveys are still needed in order to provide strong empirical evidence for the widespread adoption of agile methods and to determine the actual rates of their usage.

Acknowledgments

This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014) and the National Research Fund in Bulgaria under contract DMU03-40.

References

Agileme.co, 2011. New Zealand Agile Survey. http://agileme.co/pubs/NZ_Agile_Survey_2011.pdf
Ali, M.A., 2012. Survey on the State of Agile Practices Implementation in Pakistan. Int. J. Inform. Commun. Technol. Res. 2 (5), 459–464.
Ambler, S., 2006–2009. Agile adoption rate survey. Dr. Dobb's J..
Ambler, S., 2007. Survey says. . .agile has crossed the chasm. Dr. Dobb's J. 32 (8).
Ambler, S., 2011. Agile Adoption Strategies.
Ambler, S., 2012. Agile Planning. http://www.ambysoft.com/surveys/agilePlanning2012.html
Asnawi, A.L., Gravell, A.M., Wills, G.B., 2012. Factor analysis: investigating important aspects for agile adoption in Malaysia. In: International Conference on Agile and Lean Software Methods (AgileIndia2012), pp. 60–63.
Auger, C.P., 1998. Information Sources in Grey Literature. K. G. Saur Verlag.
Azizyan, G., Magarian, M.K., Kajko-Mattson, M., 2011. Survey of agile tool usage and needs. In: Agile Conference (AGILE), pp. 29–38.
Baruah Ashima, N., 2012. A survey of the use of agile methodologies in different Indian Small and Medium Scale Enterprises (SMEs). Int. J. Comput. Appl. 47 (20), 38–44.
Ciolkowski, M., Laitenberger, O., Vegas, S., Biffl, S., 2003. Practical experiences in the design and conduct of surveys in empirical software engineering. In: Conradi, R., Wang, A. (Eds.), Empirical Methods and Studies in Software Engineering, vol. 2765. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 104–128, http://dx.doi.org/10.1007/978-3-540-45143-3_7.
Cruzes, D.S., Dyb, T., 2010. #229 Synthesizing evidence in software engineering research. In: Paper presented at the Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Bolzano-Bozen, Italy.
Dingsoyr, T., Dyba, T., Abrahamsson, P., 2008. A Preliminary Roadmap for Empirical Research on Agile Software Development. In: Agile, 2008. AGILE'08. Conference, 4–8 August 2008, pp. 83–94, http://dx.doi.org/10.1109/Agile.2008.50.
Dingsøyr, T., Nerur, S., Balijepally, V., Moe, N.B., 2012. A decade of agile methodologies: towards explaining agile software development. J. Syst. Software 85 (6), 1213–1221, http://dx.doi.org/10.1016/j.jss.2012.02.033.
Dybå, T., Dingsøyr, T., 2008. Empirical studies of agile software development: a systematic review. Inform. Software Technol. 50 (9–10), 833–859, http://dx.doi.org/10.1016/j.infsof.2008.01.006.
Guba, E., 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ. Technol. Res. Dev. 29 (2), 75–91, http://dx.doi.org/10.1007/BF02766777.
Jain, S., Ali Babar, M., Fernandez, J., 2013. Conducting empirical studies in industry: balancing rigor and relevance. In: 2013 1st International Workshop on Conducting Empirical Studies in Industry (CESI), 20 May 2013, pp. 9–14, http://dx.doi.org/10.1109/CESI.2013.6618463.
Kasunic, M., 2005. Designing an Effective Survey. Handbook CMU/SEI-2005-HB-004. Software Engineering Institute, Carnegie Mellon University.
Kitchenham, B.A., Pfleeger, S.L., 2001–2003. Principles of survey research 1–6. In: ACM SIGSOFT Software Engineering Notes.
Kurapati, N., Manyam, V.S.C., Petersen, K., 2012. Agile software development practice adoption survey. In: Wohlin, C (Ed.), Agile Processes in Software Engineering and Extreme Programming, vol. 111. Lecture Notes in Business Information Processing. Springer, Berlin, Heidelberg, pp. 16–30, http://dx.doi.org/10.1007/978-3-642-30350-0_2.
Laanti, M., Salo, O., Abrahamsson, P., 2011. Agile methods rapidly replacing traditional methods at Nokia: a survey of opinions on agile transformation. Inform. Software Technol. 53 (3), 276–290, http://dx.doi.org/10.1016/j.infsof.2010.11.010.

Little, J., 2007. The Nokia Test.
Malhotra, M.K., Grover, V., 1998. An assessment of survey research in POM: from constructs to theory. J. Oper. Manage. 16 (4), 407–425.
McHugh, M., McCaffery, F., Casey, V., 2012. Barriers to adopting agile practices when developing medical device software. In: Mas, A., Mesquida, A., Rout, T., O'Connor, R., Dorling, A. (Eds.), Software Process Improvement and Capability Determination, vol 290. Communications in Computer and Information Science. Springer, Berlin, Heidelberg, pp. 141–147, http://dx.doi.org/10.1007/978-3-642-30439-2_13.
Norton, D., 2008. The Current State of Agile Method Adoption. Gartner, Inc.
Pareto, V., 1971. Manual of Political Economy. Augustus M. Kelley, Publishers.
Parnas, D.L., Curtis, B., 2009. Empirical research in software engineering: a critical view. IEEE Software 26 (6), 56–59, http://dx.doi.org/10.1109/MS.2009.184.
Perry, D.E., Porter, A.A., Votta, L.G., 2000. Empirical studies of software engineering: a roadmap. In: Paper Presented at the Proceedings of the Conference on The Future of Software Engineering, Limerick, Ireland.
Pinsonneault, A., Kraemer, K.L., 1993. Survey research methodology in management information systems: an assessment. JMIS 10 (2), 75–105.
Qumer, A., Henderson-Sellers, B., 2008. A framework to support the evaluation, adoption and improvement of agile methods in practice. J. Syst. Software 81 (11), 1899–1919, http://dx.doi.org/10.1016/j.jss.2007.12.806.
Rao, K.N., Naidu, G.K., Chakka, P., 2011. A study of the Agile software development methods, applicability and implications in industry. Int. J. Software Eng. Appl. 5 (2), 35–46.
Rodríguez, P., Markkula, J., Oivo, M., Turula, K., 2012. Survey on agile and lean usage in finnish software industry. In: Paper Presented at the Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Lund, Sweden.
Rönkkö, M., Peltonen, J., Frühwirth, C., 2011. Examining the effects of agile methods and process maturity on software product development performance. In: Regnell, B., Weerd, I., Troyer, O. (Eds.), Software Business, vol. 80. Lecture Notes in Business Information Processing. Springer, Berlin, Heidelberg, pp. 85–97, http://dx.doi.org/10.1007/978-3-642-21544-5_8.
Rosenthal, R., 1979. The "File Drawer Problem" and the tolerance for null results. Psychol. Bull. 86 (3), 638–641.
The Scrum-Butt Test. http://scrumbutt.me/
Selby, R., 2007. Empirical software engineering research roadmap discussion and summary. In: Basili, V., Rombach, D., Schneider, K., Kitchenham, B., Pfahl, D., Selby, R. (Eds.), Empirical Software Engineering Issues. Critical Assessment and Future Directions, vol. 4336. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 184–187, http://dx.doi.org/10.1007/978-3-540-71301-2_54.
Serena, 2012. Agile 2012 Survey. http://www.serena.com/docs/repository/alm-benchmark/Serena-Agile-2012-Survey-Infographic.pdf
Sheffield, J., Lemétayer, J., 2012. Factors associated with the software development agility of successful projects. Int. J. Project Manage., http://dx.doi.org/10.1016/j.ijproman.2012.09.011.
Shenton, A.K., 2004. Strategies for ensuring trustworthiness in qualitative research projects. Educ. Inform. 22 (2), 63–75.
Sidky, A., Arthur, J., Bohner, S., 2007. A disciplined approach to adopting agile practices: the agile adoption framework. Innovations Syst. Software Eng. 3 (3), 203–216, http://dx.doi.org/10.1007/s11334-007-0026-z.
Sjoberg, D.I.K., Dyba, T., Jorgensen, M., 2007. The Future of Empirical Methods in Software Engineering Research. In: Future of Software Engineering, 2007. FOSE'07, 23–25 May 2007, pp. 358–378, http://dx.doi.org/10.1109/FOSE.2007.30.
Song, F., Parekh, S., Hooper, L., Loke, Y.K., Ryder, J., Sutton, A.J., Hing, C., Shing, C., Pang, C., Harvey, I., 2010. Dissemination and publication of research findings: an updated review of related biases. Health Technol. Assess. 14 (8), 1–220.
ThoughtWorks, 2011. Agile Adoption in India. www.thoughtworks.com/articles/agile-adoption-india-2011
Valtech, 2012. Agile India 2012. http://www.slideshare.net/ivaltech/agile-india-2012-survey-results-final
VersionOne, 2011. State of Agile Development Survey Results. http://www.versionone.com/state_of_agile_development_survey/11/
Vijayasarathy, L., Turk, D., 2012. Drivers of agile software development use: dialectic interplay between benefits and hindrances. Inform. Software Technol. 54 (2), 137–148, http://dx.doi.org/10.1016/j.infsof.2011.08.003.
Waters, K., 2008. How Agile Are You? (Take This 42 Point Test).
West, D., Grant, T., 2010. Agile Development: Mainstream Adoption has Changed Agility. Forrester Research, Inc.
Weyuker, E.J., 2011. Empirical Software Engineering Research – The Good, The Bad, The Ugly. In: Empirical Software Engineering and Measurement (ESEM), 2011 International Symposium on, 22–23 September 2011, pp. 1–9, http://dx.doi.org/10.1109/ESEM.2011.66.
Wieringa, R., 2012. Towards a Unified Checklist for Empirical Research in Software Engineering: first proposal. In: Paper presented at the 16th International Conference on evaluation & Assessment in Software Engineering, EASE 2012, Ciudad Real, Spain.
Williams, L., 2012. What agile teams think of agile principles. CACM 55 (4), 71–76, http://dx.doi.org/10.1145/2133806.2133823.
Xebia, 2011. Second Dutch Agile Survey. http://www.xebia.com/en/agilesurvey2011

Stavros Stavru is a researcher in the Faculty of Mathematics and Informatics at Sofia University St. Kliment Ohridski. His research interests cover a variety of Software engineering and Management issues mostly in the field of Agile software development and Corporate culture/values. He graduated Informatics and Business administration in Sofia University St. Kliment Ohridski and is currnelty pursuing PhD in Software engineering in the same university. Mr. Stavru has a solid industrial background in leading, training and coaching agile teams and is authoring various agile games including "The sandwich retrospective", "Means first, then states" and "Values advocacy".