Postprint

When asked about the average accuracy of the effort estimates compared to the actual effort, the results show that " underestimated by 25% or more" was the most common answer. This further solidifies my lesson from SENG302 where my group begun to anticipate worse case scenarios (space for bugs, technology challenges, developer errors etc.), that leads to 30% more in estimations.

Question: Do you think adding 30% more into your estimation helps account for worst case scenarios - or are there more factors (team skill, communication, story points) to take into consideration and its not just this simple? - some paper recommend double it

Question: Any ideas on techniques to account for communication overhead when estimating? - maybe bargining higher at the start

An Empirical Investigation on Effort Estimation in Agile Global Software Development

Ricardo Britto
Department of Software Engineering
Blekinge Institute of Technology
Sweden
ricardo.britto@bth.se

Emilia Mendes
Department of Software Engineering
Blekinge Institute of Technology
Sweden
emilia.mendes@bth.se

Jürgen Börstler
Department of Software Engineering
Blekinge Institute of Technology
Sweden
jurgen.borstler@bth.se

Abstract—Effort estimation is a project management activity that is mandatory for the execution of software projects. Despite its importance, there have been just a few studies published on such activities within the Agile Global Software Development (AGSD) context. Their aggregated results were recently published as part of a secondary study that reported the state of the art on effort estimation in AGSD. This study aims to complement the aforementioned secondary study by means of an empirical investigation on the state of the practice towards effort estimation in AGSD. To do so, a survey was carried out using as instrument an on-line questionnaire and a sample comprising software practitioners experienced in effort estimation within the AGSD context. Results show that the effort estimation techniques used within the AGSD and collocated contexts remained unchanged, with planning poker being the one employed the most. Sourcing strategies were found to have no or a small influence upon the choice of estimation techniques. With regard to effort predictors, global challenges such as cultural and time zone differences were reported, in addition to factors that are commonly considered in the collocated context, such as team experience. Finally, many challenges that impact the accuracy of the effort estimates were reported by the respondents, such as problems with software requirements and the fact that the communication overhead between sites is not properly accounted.

I. INTRODUCTION

Effort estimation, i.e. the process used to predict the effort needed to fulfill a given task [1], is a project management activity that is fundamental to manage resources in an effective way [1]. Reliable effort estimates increase the chances of accomplishing the required work related to a given software project without time or cost overruns [2].

Nowadays, software can be developed in many different contexts, by means of different software development approaches, such as Global Software Development (GSD – software development in a globally distributed manner) [3], Agile Software Development (ASD – software development based on agile methods) [4] and Agile Global Software Development (AGSD) [5], [6] (a combination of GSD and ASD). Given such diversity, we put forth that the way that effort is estimated within the context of a project may vary depending on the software development approach being used.

Despite the fact that there are lots of effort estimation techniques designed for collocated contexts [7]–[9], i.e. software development performed by a team located in a single physical room [10], there is evidence that effort estimation techniques that were originally designed for collocated contexts are not readily applicable in the global context [11]–[14]. To make effort estimation even more challenging in the AGSD context, there is also evidence that agile methods are not readily applicable in the global context neither [15]–[17].

To better understand the state of the art in effort estimation in those three contexts (GSD, ASD and AGSD), three earlier studies were carried out:

- A systematic literature review (SLR) on effort estimation in GSD [18].
- A SLR on effort estimation in ASD [19].
- The results from the two SLRs above were then combined to identify the differences and commonalities and obtain evidence on effort estimation in AGSD [20].

Given the growing number of AGSD projects [16], [17], it is important to complement the evidence on the state of the art towards effort estimation in AGSD with evidence from the state of the practice, which is the main goal of the present work. To do so, we carried out a survey [21] among software practitioners with experience in effort estimation in the AGSD context.

The present paper details that survey and makes five main contributions within the AGSD context:

- It presents how existing effort estimation techniques have been applied in practice.
- It presents the predictors that have been considered in practice.
- It presents the impact of sourcing strategies on the selection of effort estimation techniques.
- It presents the accuracy of the effort estimates reported by practitioners.
- It presents challenges that impact the accuracy of the effort estimates.

The remainder of this paper is organized as follows: Section III details the employed research methodology, followed by the presentation and discussion of the results of the conducted empirical study in Sections IV and V, respectively. Section VI discusses validity threats. Conclusions and future work are described in Section VII.

II. Related Work

There has been only one previous survey within the context of effort estimation in the GSD domain, by Peixoto *et al.* [13]. They investigated how effort estimation techniques have been used by practitioners in the GSD context. Their results did not identify clear criteria used by practitioners when choosing an effort estimation technique.

The present survey differs from Peixoto *et al.*'s survey (S1) in the following ways:

- The context of our study is AGSD, whereas S1 was performed in the GSD context.
- S1 investigated the practices within a single company, whereas our work gathered data from practitioners from many different companies.
- S1 focused solely on effort estimation techniques, whereas ours also focused on the predictors used in effort estimation and the effort estimates' accuracy.

III. Research Methodology

In the following subsections we describe the research methodology employed in this paper.

A. Research Questions

To obtain a detailed understanding on how effort estimation has been performed in the AGSD context, our research questions focused not only on the methods used to estimate effort, but also on the effort predictors used, the accuracy of the effort estimates, and the impact that sourcing strategies may have on the selection of effort estimation methods, as follows:

- **Research question 1 (RQ1)** – Which techniques are used to estimate effort in AGSD?
- **Research question 2 (RQ2)** – What is the impact of the applied sourcing strategy on the selection of an effort estimation method in AGSD?
- **Research question 3 (RQ3)** – Which effort predictors are used to estimate effort in AGSD?
- **Research question 4 (RQ4)** – What are the challenges that impact the accuracy of the effort estimates in AGSD?

B. Survey Design

A survey is a retrospective form of investigation that targets at gathering data from a wider population [21]. To perform surveys, it is mandatory to define its purpose, the analysis unit to be used and a representative sample of the population related to the research problem. In the present work, our survey was defined as follows:

- The *purpose* of the survey is to collect data on the state of the practice in effort estimation in the AGSD context.
- The *analysis unit* is the effort estimation process element (technique, cost driver or size metric).
- The *target population* consists of practitioners who have worked with effort estimation in the AGSD context.
- The *sampling unit* is the practitioner responsible for performing effort estimation in a company.

The instrument of the survey was an on-line semi-structured questionnaire, designed using SurveyMonkey[1] tool. The questionnaire encompassed closed and open-ended questions. The motivation for using an on-line questionnaire was to maximize respondents' participation and coverage.

The questionnaire had three types of questions: demographic questions, AGSD questions and effort estimation questions. Those questions reflected the research questions of this paper (see Subsection III-A).

Demographic questions are related to the respondents' country of residence, their job titles and number of years of experience as a professional in software development industry, teams' sizes and types of applications developed by the respondents' companies. *AGSD* questions are related to aspects of ASD, such as the agile methods and sourcing strategies employed by the respondents' companies. Finally, *effort estimation* questions are related to the effort estimation techniques and predictors (size metrics/cost drivers) employed by the respondents.

Note that the survey detailed herein is part of a wider investigation that has been conducted by the authors focusing on the state of the practice in effort estimation within the agile context (both global and collocated). Therefore, only the questions that are relevant for effort estimation in the AGSD context are presented below. Note that the numbers associated with each of the questions presented below may sometimes differ from the number used in the original questionnaire.

- **Demographic Questions:**
    - **Question 1 (SQ1)** – In which country do you work most of the time?
    - **Question 2 (SQ2)** – For how long have you been working as a professional in software development industry?
    - **Question 3 (SQ3)** – What is your job title?
    - **Question 4 (SQ4)** – What is the size of your team?
    - **Question 5 (SQ5)** – Which types of software systems does your company develop?
- **AGSD Questions:**
    - **Question 6 (SQ6)** – Which agile methods are employed in your company?
    - **Question 7 (SQ7)** – Which sourcing strategies are applied by your company?
    - **Question 8 (SQ8)** – What is the average number of offshore insourced sites when considering the AGSD projects performed by your company?
    - **Question 9 (SQ9)** – What is the average number of offshore outsourced sites when considering the AGSD projects performed by your company?
    - **Question 10 (SQ10)** – What is the most common role of your company in AGSD outsourced projects?
- **Effort Estimation Questions:**
    - **Question 11 (SQ11)** – Which effort estimation

---

techniques for AGSD projects are employed in your company?

- **Question 12 (SQ12)** – What is the average effort estimation accuracy for AGSD projects in your company?
- **Question 13 (SQ13)** – Which of the following factors do you believe are fundamental for estimating effort in AGSD projects?
- **Question 14 (SQ14)** – What are the challenges that impact the accuracy of the effort estimates in AGSD projects?

Questions SQ1–SQ13 used categorical measurement scales (either nominal or ordinal) and were compulsory. Questions SQ3, SQ5, SQ6, SQ8–Q11 and SQ13 also provided free-text space to input an "Other" category answer. Question SQ14 is an open-ended question, so just a free-text space was provided to answer such question. Table I details the scale type and points/categories used for each question.

C. Survey Execution

Survey respondents were invited primarily from the authors' contact networks. Furthermore, the survey was advertised in two conferences[2], and in many LinkedIn[3] communities on ASD, GSD and software measurement. It was made clear in the invitation that only people who worked with effort estimation in the AGSD context should answer the questionnaire.

The questionnaire was available on-line from August 12th to October 10th 2014 and was answered by a self-selected sample of 51 respondents. Prior to making the questionnaire available on-line, it was validated by another researcher to ensure that the questions were clear and straight forward to answer.

Although the questionnaire was anonymous, respondents could optionally provide contact details (email) if a follow-up clarification was needed.

IV. Results and Analysis

The results herein presented are organized as follows: Sub-section IV-A details the respondents' demographics; subsection IV-B presents the results related to the AGSD questions; subsections IV-C to IV-F present the results for research questions *RQ1* to *RQ4* respectively.

A. Demographics Questions

SQ1 captured the respondents' work countries. A total of 12 different countries were represented in the sample. Their distribution by continents is presented in Table II. Most of the respondents work in Europe or Asia, followed by North America. Such a data distribution complies with the countries/regions where AGSD has been reported more frequently to date [18], [20].

---

[2]XP 2014, the 15th International Conference on Agile Software Development and ICGSE 2014, the 9th International Conference on Global Software Engineering 2014

[3]See http://www.linkedin.com

TABLE I
SURVEY CATEGORIES

Question ID	Scale Type	Question type	Categories' description
SQ1	Nominal	Single answer	All the 206 states recognized by United Nations
SQ2	Ordinal	Single answer	Less than 1 year; More than 1 year, Less than 3 years; More than 3 years, less than 5 years; More than 5 years, less than 10 years; Above 10 years
SQ3	Nominal	Single answer	Program manager; Product owner; Scrum master; Team leader; QA manager; System analyst; Software architect; Designer; Developer; Tester; Other
SQ4	Ordinal	Single answer	1–5; 6–10; 11–15; 16–20; 21–50; 51–100; above 100
SQ5	Nominal	Multiple answers	Telecom/mobile applications; Business applications; Embedded systems; Safety critical systems; E-commerce applications; Financial applications; Data processing applications; Real time systems; System software; Other
SQ6	Nominal	Multiple answers	Extreme programing (XP); Scrum; Kanban; Feature driven development (FDD); crystal; Dynamic system development method (DSDM); Customized XP; Customized Scrum; Combination of XP and Scrum; Other
SQ7	Nominal	Single answer	insourcing; outsourcing
SQ8	Ordinal	Single answer	1; 2; 3; 4; Other
SQ9	Ordinal	Single answer	1; 2; 3; 4; Other
SQ10	Nominal	Single answer	Client; Vendor; Other
SQ11	Nominal	Multiple answers	Planning poker; Use case points; Analogy; Expert judgment; Delphi; COCOMO; Other
SQ12	Ordinal	Single answer	Underestimated by 50% or more; Underestimated by 25% or more; Spot on (0-5)% variation; Overestimated by 5% or more; Overestimated by 25% or more; Overestimated by 50% or more
SQ13	Nominal	Multiple answers	Communication model; Time zone difference; Time zone overlap; Cultural difference; Software process model; Language difference; Geographical distance; Teams prior experience; Teams expertise; Project domain; Non functional requirements; Number of sites; Customer communication; Story points; Use case points; Function points; Object points; Lines of codes; Task size; Other.

Table II
COUNTRIES IN WHICH THE RESPONDENTS WORK MOST OF THE TIME

Continent	Countries	Percentage
Asia	India (13.73%), Pakistan (17.65%), Singapore (1.96%)	33.34%
Oceania	Australia	7.84%
Europe	United Kingdom (1.96%), Sweden (29.41%), Germany (3.92%), Netherlands (1.96%), Italy (1.96%)	39.21%
North America	USA	15.69%
Central and South America	Costa Rica (1.96%), Brazil (1.96%)	3.92%

SQ2 measured the respondent's industrial experience. The results show that 76.47% of the respondents have at least 5 years of experience in the software industry, with close to 51% with 10 years or more of experience (25.49% with "more than 5 years, less than 10 years" and 50.98% with "above 10 years").

SQ3 captured the respondent's job titles (roles). The results show a variety of roles, which also suggests that, at least amongst the respondents, effort estimation is not centralized within a small subset of roles. The results show that in most of the cases effort is estimated by either "developers" (27.45%) or "Other" (21.57%) ("line manager", "information security manager" and "agile coach").

Considering the context of our survey (AGSD), it seems reasonable that evidence suggests developers are also responsible for estimating effort, given that within such context teams are cross-functional and most of the team members are called "developers" [4]. Note that the role "project manager" was not selected by any of the respondents. However, such a role is not necessarily required in agile teams [4].

SQ4 measured the respondents' team size. Our results show that more than half of the respondents (56.86%) reported teams with up to 10 members. Such a size range complies with common agile best practices, where agile teams range from 5 to 9 members in general [4].

There is also a large percentage of respondents (33%) who reported team sizes with "more than 16" and even "more than 100 people". Such numbers do not support the most commonly reported team sizes in the agile literature. However, it is possible to scale the team size up in ASD projects, in order to perform more resource-consuming software projects[4]. Considering the context of this work (AGSD), in our view it is also reasonable to assume that a project carried out globally can demand more human resources and therefore larger teams, when compared to collocated agile projects.

SQ5 captured the software types developed by the respondents. Results show that "Telecom/Mobile" (52.94%) and "Business Applications" (56.86%) were the most frequent answers, followed by "Financial" (39.22%) and "E-

[4]www.scaledagileframework.com/

Commerce" (37.25%) applications. For the free-text "Other" option (17.65%), "CAD" and "Health Care Applications" were provided most frequently.

B. AGSD Questions

SQ6 captured the agile methods employed by the respondents' companies. Results show that the agile method selected most often was "Scrum (84.31%), followed by "Kanban" [22] (37.25%) and "Extreme programming (XP)" [23] (23.53%). The free-text "Other" option (7.84%) revealed only one additional answer, model-driven development [24], which is not strictly an agile method. Since most respondents did not provide contact details (optional in the questionnaire), it was not possible to follow-up the "Other" answers for clarification.

Note that respondents could select several agile methods, which explains the high percentage for "Scrum", in particular. However, it is important to highlight that the fact that "Scrum" was the most frequent answer is not surprising, since there is evidence that such method is the most employed in the software industry [17], [25], [26]

SQ7 asked for the sourcing strategies employed by the respondents' companies. Results show that "insourcing" was the most frequent answer (47%), followed by "both" (35%) and "outsourced" (18%). This result corroborates our earlier results about the state of the art in effort estimation within the context of GSD [18], [20].

Using SQ8, we captured the average number of insourced sites. This questions was only answered by respondents who answered "insourcing" or "both" in SQ7. The results show small differences between categories. The two answers with the highest percentages were "2" sites (26%) and "1" site (24%), followed by "4" sites (19%), "other" (17%), and "3" sites (14%). The valid free-text answers for option "Other" were "average of 5 insourced sites" and number of "insourced sites between 1 and 10". The remaining 5 free-text answers were unclear or empty. Since contact information was not available for those respondents, a follow-up for clarification was not possible.

Using SQ9, we captured the average number of outsourced sites. This questions was only answered by respondents who answered "Outsourcing" or "Both" in SQ7. The results for outsourced sites differ from those for insourced sites. The most frequent answer was "3" sites (33%), followed by "2" (30%) sites and "Other" (19%). All "Other" respondents provided unclear/empty free-text answers. Since contact information was not available for those respondents, a follow-up for clarification was not possible.

SQ10 captured the roles of the respondents' companies in outsourced projects. Results show an even distribution between "Client" (41%) and "Vendor" (44%) companies. Details regarding the four "Other" answers (15%) were not available.

C. Research Question 1

RQ1 relates to the effort estimation techniques that are employed by AGSD practitioners. SQ6 was used to answer RQ1. Results show that "Planning poker" (72.55%) was the effort

estimation technique selected most frequently, followed by "Expert judgment" (47.06%). "COSMIC" (which is actually a size metric), "Multiple expert judgment" and "Delphi-PERT approach" were the methods reported using the "Other" option (15.69%).

Although "Planning poker" and "Expert judgment" are also frequently used in ASD [27], it is interesting to note that practitioners did not add any additional technique as a result of a context change from collocated to global software development. Note that more than 40% of the respondents work in teams larger than 10 people, which could perhaps bring some challenges when relying solely on techniques such as "Planning poker" and "Expert judgment".

Also note that the number of respondents using a single effort estimation technique (51%) was quite similar to the number of respondents using more than one technique (49%).

Table III shows small differences between the two ways in which "Planning poker" was reported: as a single technique (48.65%), or together with other techniques (51.35%). Other techniques were mostly reported together with other options. COCOMO and Delphi were reported only together with other techniques.

Table III
PERCENTAGES OF THE SELECTED EFFORT ESTIMATION TECHNIQUES CONSIDERING THE WAY THEY WERE REPORTED

Technique	Single	Multiple
Planning Poker	48.65%	51.35%
Expert Judgment	20.83%	79.17%
Analogy	6.67%	93.33%
Use Case Points	0.0%	100%
Delphi	0.0%	100%
COCOMO	0.0%	100%

The fact that a respondent selected more than one effort estimation technique can be interpreted in two different ways. Either the effort estimation techniques are used together in the context of a project, or individual effort estimation techniques are singly employed in different projects. As part of our future work, we plan to investigate further this issue via interviews with the survey participants who selected more than one technique and who provided their contact details.

D. Research Question 2

RQ2 was answered by combining the data gathered from SQ5 and SQ6. Results are displayed in Table IV and show that "planning poker" was the most selected effort estimation technique, regardless of the sourcing strategy used (insourcing 43.24%; outsourcing 35.71%; both 36.84%).

The second most reported effort estimation technique changes according to the applied sourcing strategy. When "insourcing" was the employed sourcing strategy, "expert judgment" was as reported as "planning poker" (43.24%). On the other hand, when "outsourcing" was the employed sourcing strategy, "analogy" was the second most reported technique (28.57%).

Table IV
SELECTED METHOD BY SOURCING STRATEGY

Technique	Insourcing	Outsourcing	Both
Planning Poker	43.24%	35.71%	36.84%
Expert Judgment	43.24%	21.42%	31.57%
Analogy	18.91%	28.57%	13.15%
Use Case Points	10.81%	7.14%	7.89%
Delphi	2.70%	7.14%	5.26%
COCOMO	0.0%	0.0%	5.26%

E. Research Question 3

RQ3 relates to the predictors (size metrics/cost drivers) that have been considered by practitioners in AGSD. Survey question SQ7 was used to gather data to answer RQ3.

Results show that "team's skill level" (71%) was the most selected cost driver, followed by "communication model" (65%), "time zone difference" (59%), "team's prior experience" (55%), "cultural difference" (45%) and "software process model" (41%). "Technical dependencies", "difference in work hours between sites" and "uncertainty level" were reported by means of the option "Other" (14%). The results are in-line with the state of the art [18], [20].

Results also show that "task size" (67%) was the most selected size metric, followed by "story points" (63%) and "use case points" (27.45%). "UML points" and "COSMIC function points" were reported by means of the option "Other" (7.8%). Size metrics that are popular in plan-driven software development, e.g. "function points" (20%) and "lines of code" (8%), were selected by fewer respondents.

F. Research Question 4

RQ4 relates to the challenges that impact the accuracy of the effort estimates in AGSD. Survey questions SQ12 and SQ14 gathered data to answer RQ4.

SQ12 asked for the average accuracy of the effort estimates, when compared to the actual effort. The results show that "underestimated by 25% or more" was the most common answer (33.33%), followed by "spot on (0–5)% variation" (19.61%) and "underestimated by 5% or more" (17.65%). In 54.90% of the cases the effort estimates were underestimated (see Total in Table V).

Considering the accuracy values reported in the effort estimation literature, the results suggest that the survey respondents are estimating effort with an acceptable level of accuracy [28]. However, it was puzzling to see close to 20% of the respondents reporting that estimates were "spot-on". We hypothesize that perhaps such respondents work with fixed budget contracts or projects where effort is estimated and later adjusted to comply with actuals.

SQ14 asked for the challenges that impact the accuracy of the effort estimates in the AGSD context. Only 24 respondents (47%) answered this optional open-ended question. The provided answers were categorized as follows:

- **Distribution-related challenges** - Thirteen respondents (25.5%) reported that time, language and cultural dif-

Table V
DISTRIBUTION OF THE RESPONSES BY TYPE OF ERROR

Range	Underestimated	Overestimated	Spot on
By 5% or more	17.65%	9.8%	-
By 25% or more	33.33%	11.76%	-
By 50% or more	3.92%	3.92%	-
Between 0% and 5%	-	-	19.61%
Total	54.90%	25.48%	19.61%

ferences between sites and the distance to the client can affect the accuracy of the effort estimates. In their opinion, such factors affect the communication between sites and this communication overhead is not properly accounted.

- **Requirements-related challenges** - Twenty respondents (39.2%) informed that unclear, unstable and mis-documented requirements affect the accuracy of the effort estimates. In their opinion, those challenges lead to unpredicted activities during the development process, increasing the real effort of software projects.
- **Team-related challenges** - Fourteen respondents (27.5%) reported that the lack of domain knowledge, lack of knowledge on the required technologies, lack of experience on the used effort estimation technique and lack of team cohesion (teams with people that never worked or worked little time together) lead to wrong assumptions, which affect the accuracy of the effort estimates. Another challenge presented by the respondents was the fact that in many cases the effort is not estimated by the same team responsible for constructing the software, which also leads to wrong assumptions, jeopardizing the accuracy of the effort estimates.
- **Client-related challenges** - Two respondents (4%) reported that poor participation and mis-prioritization of the tasks by the clients affect the accuracy of the effort estimates. It means that the effort is estimated considering an active participation of the clients; however development teams find a different scenario when a project starts, which in general leads to bigger effort.

V. DISCUSSION

The findings reported herein are discussed as follows (organized by research questions).

A. RQ1 - Which techniques have been used to estimate effort in AGSD?

The survey results suggest that the same techniques used in the collocated context have been used in the AGSD context, without any adaptations. Considering that there are few studies about effort estimation in GSD [18] and AGSD [20], the absence of customized effort estimation techniques is not surprising. Nevertheless, this does not mean that such customization would not be important to improve estimation accuracy and even the prediction process. Thus, we believe that more research should be carried out to investigate this issue further.

B. RQ2 - What is the impact of the applied sourcing strategy on the selection of an effort estimation method in AGSD?

In relation to the impact of the sourcing strategy on the selection of effort estimation techniques, the results suggest the following:

- The applied sourcing strategy has no or small influence on the selection of the technique when just one technique is selected to estimate effort. The same finding appears to be true regarding the selection of the first technique when two or more techniques are selected to estimate effort in AGSD projects. In both cases, "planning poker" was the most frequently reported method.
- The applied sourcing strategy appears to have a bearing on the selection of additional methods; when two or more methods are employed to estimate effort. In such situations, "expert judgment" appears to be related to an "insourcing" strategy and estimation by "analogy" appears to be related to an "outsourcing" strategy.

We suspect that estimation by "analogy" is the second most frequently selected technique because a company would not necessarily have access to the internal information of outsourced sites. Rather, the client company would be obliged to compare a new project with similar past finished projects. This could explain the choice of effort estimation techniques. We also aim to investigate further this issue in collaboration with some of the respondents who answered this survey and provided their contact details.

C. RQ3 - Which effort predictors have been used to estimate effort in AGSD?

The results from this survey suggest that most of the well-known GSD challenges/issues (e.g. "communication" and "cultural difference" [3]) have been considered as cost drivers. However, "team's skill level", which is a relevant cost driver in both global and collocated context, was the most frequently reported cost driver.

The survey did not ask more detailed questions on how the reported cost drivers were measured by the respondents and used in the effort estimation process. Further work will be carried out to obtain such data.

Regarding the reported size metrics, the results suggest that "task size" and "story points" are the size metric that fit properly for estimating effort in AGSD. Not surprisingly, they are the most frequent size metric in ASD literature [4], [23].

D. RQ4 - What are the challenges that impact the accuracy of the effort estimates in AGSD?

A considerable number of the respondents reported reasonable effort estimates in AGSD (Table V). The effort estimates' accuracy reported in the literature on effort estimation is slightly worse [11], [29]. Maybe the respondents were not comfortable to report actual accuracy, even when answering an anonymous questionnaire. Maybe the answers are legitimate. Some of our future work aims to investigate further this issue.

According to the participants of our survey, there are many different challenges that impact the accuracy of the effort

estimates in the AGSD context. Most of the respondents reported that the main problem lies on the requirements, which is also a challenge in collocated agile software projects.

The participants of the survey also pointed out that the communication overhead, which is bigger in globally distributed projects, is not properly accounted most of the time by practitioners. It affects the accuracy of the effort estimates, leading to underestimated effort. The impact reported for this challenge is in-line with the accuracy values reported in the survey (See Subsection IV-F). We believe that further research should be conducted to identify or develop approaches to calculate the impact of the communication overhead on the overall effort of AGSE projects.

VI. THREATS TO VALIDITY

The research reported in this paper is subjected to the following validity threats:

- **Construct validity** is concerned with the relation between the theories behind the research and the observations [30]. This type of threat was mitigated by collecting data from a wide range of respondents, who worked in many different countries and companies.
- **Conclusion validity** is concerned with the possibility of incorrect conclusions about relationships in the observations that could arise from errors in the data sources [30]. To mitigate such kind of threat, the survey was validated by other researchers to ensure that the questions were clear and straight forward to answer. Besides, we made clear for the respondents that just people who had worked with effort estimation should answer the questionnaire. Nonetheless, we cannot claim that all questions were well understood by the respondents.
- **Internal validity** is related to issues that may affect the causal relationship between treatment and outcome [30]. To mitigate this threat, the survey was validated by other researchers. Considering that the questionnaire was to be answered just once per respondent, we believe that the possibility of learning effect was removed. Finally, no fatigue effect was expected to affect the respondents, since the time to answer the questionnaire was short (15 minutes).
- **External validity** is concerned with the ability to generalize the findings of a given study beyond its study context [30]. In general, research in empirical software engineering is hard to generalize, because the research outcomes are highly influenced by the research context. In order to improve the representativeness of the findings of this work, we collected data from many respondents, widening the number of contexts for which the findings of this work could be generalized.

VII. CONCLUSIONS

This paper presented the results of a survey on effort estimation in the context of Agile Global Software Development. The main goal of this paper was to complement the already existent

evidence obtained from the state of the art by obtaining evidence from the state of the practice.

An on-line questionnaire was made accessible from August 12 of 2014 to October 10 of 2014. Respondents from 12 different countries have participated in the survey. A self-selected sample of 51 respondents have fully answered the questionnaire.

The reported effort estimation techniques are well-know in the collocated context. This result was not surprising once it was already known that there have been few studies on effort estimation in AGSD context.

It was surprising the fact that most of the respondents reported accurate effort estimates. They also reported many challenges that impact the accuracy of the effort estimates, such as unclear, unstable and mis-documented requirements. In addition, they reported that it is common to underestimate the impact of the communication overhead between the distributed sites of AGSE projects, which leads to underestimated effort.

Regarding the impact of the sourcing strategies on the selection of effort estimation techniques, the results suggest that when just one technique is applied, the impact of the employed sourcing strategy appears to be insignificant. However, when more than one method is applied, the selection of the second method appears to be affected by the sourcing strategy (i.e. "expert judgment" related to "offshore insourcing" and "estimation by analogy" related to "offshore outsourcing").

Finally, about the predictors, the results suggest that the main GSD challenges have been considered by the practitioners, along with other factors that are also relevant in any other context, such as the experience of the team.

We intend to perform further research in order to confirm and/or clarify some of our findings, such as the accuracy level of the reported effort estimates. Likewise, we intend to understand how the practitioners have been measured and incorporating the considered cost drivers in their effort estimation processes. Finally, additional research should be conducted to identify or develop approaches to calculate the impact of the communication overhead on the overall effort of AGSE projects. We believe that the listed further research could be carried out by means of case studies, which is one of the directions for our future work.

VIII. ACKNOWLEDGMENTS

REFERENCES

[1] E. Mendes, *Cost Estimation Techniques for Web Projects*. IGI Publishing, 2007.

[2] N. Fenton, W. Marsh, M. Neil, P. Cates, S. Forey, and M. Tailor, "Making resource decisions for software projects," in *Proceedings of 26th International Conference on Software Engineering - ICSE'04*, Edinburgh, Scotland, 2004, pp. 397–406.

[3] J. Herbsleb, A. Mockus, T. A. Finholt, and R. E. Grinter, "Distance, dependencies, and delay in a global collaboration," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work - CSCW'00*, New York, USA, 2012, pp. 319–328.

[4] K. Schwaber and M. Beedle, *Agile Software Development with Scrum*. Prentice Hall, 2001.

[5] D. Šmite, N. B. Moe, and P. J. Ågerfalk, Eds., *Agility across time and space: making agile distributed development a success*. Springer Verlag, 2010.

[6] N. Kamaruddin, N. Arshad, and A. Mohamed, "Chaos issues on communication in agile global software development," in *Business Engineering and Industrial Applications Colloquium (BEIAC), 2012 IEEE*, April 2012, pp. 394–398.

[7] E. Mendes, "Using knowledge elicitation to improve web effort estimation: Lessons from six industrial case studies," in *Proceedings of 34th International Conference on Software Engineering - ICSE'12*, Zurich, Switzerland, 2012, pp. 1112–1121.

[8] ——, "Building a web effort estimation model through knowledge elicitation," in *Proceedings of 13th International Conference on Enterprise Information Systems - ICEIS'11*, Beijing, China, 2011, pp. 8–11.

[9] ——, "The use of bayesian networks for web effort estimation: Further investigation," in *Proceedings of 8th International Conference on Web Engineering - ICWE'08*, New York, USA, 2008, p. 203216.

[10] S. Teasley, L. Covi, M. Krishnan, and J. Olson, "Rapid software development through team collocation," *Software Engineering, IEEE Transactions on*, vol. 28, no. 7, pp. 671–683, Jul 2002.

[11] N. Ramasubbu and R. K. Balan, "Overcoming the challenges in cost estimation for distributed software projects," in *Proceedings of 34th International Conference on Software Engineering - ICSE'12*, Zurich, Switzerland, 2012, pp. 91–101.

[12] N. C. Narendra, K. Ponnalagu, N. Zhou, and W. M. Gifford, "Towards a Formal Model for Optimal Task-Site Allocation and Effort Estimation in Global Software Development," in *Proceedings of 2012 Service Research and Innovation Institute Global Conference*, Silicon Valley, USA, 2012, pp. 470–477.

[13] C. E. L. Peixoto, J. L. N. Audy, and R. Prikladnicki, "Effort Estimation in Global Software Development Projects: Preliminary Results from a Survey," in *Proceedings of 5th IEEE International Conference on Global Software Engineering - ICGSE'10*, Princeton, USA, 2010, pp. 123–127.

[14] A. Lamersdorf, J. Munch, A. F. V. Torre, C. R. Sanchez, and D. Rombach, "Estimating the Effort Overhead in Global Software Development," in *Proceedings of 5th IEEE International Conference on Global Software Engineering - ICGSE'10*, Princeton, USA, 2010, pp. 267–276.

[15] D. Batra, "Modified agile practices for outsourced software projects," *Commun. ACM*, vol. 52, no. 9, pp. 143–148, Sep. 2009.

[16] E. Hossain, M. Ali B., and H. Y. Paik, "Using scrum in global software development: A systematic literature review," in *Proceedings of 4th IEEE International Conference on Global Software Engineering - ICGSE '09*, Limerick, Ireland, 2009, pp. 175–184.

[17] S. Jalali and C. Wohlin, "Global software engineering and agile practices: A systematic review," *Journal of Software: Evolution and Process*, vol. 24, no. 6, pp. 643–659, 2012.

[18] R. Britto, V. Freitas, E. Mendes, and M. Usman, "Effort estimation in global software development: A systematic literature review," in *Global Software Engineering (ICGSE), 2014 IEEE 9th International Conference on*, Aug 2014, pp. 135–144.

[19] M. Usman, E. Mendes, F. Weidt, and R. Britto, "Effort estimation in agile software development: A systematic literature review," in *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*, ser. PROMISE '14. New York, NY, USA: ACM, 2014, pp. 82–91.

[20] R. Britto, M. Usman, and E. Mendes, "Effort estimation in agile global software development context," in *1st International Workshop on Estimations in the 21st Century Software Engineering - 15th Conference on Agile Software Development - XP2014*, May 2014.

[21] A. Fink, *The Survey Handbook*, ser. Survey Kit Second Edition 1 1. SAGE Publications, 2003.

[22] D. J. Anderson, *Kanban: Successful Evolutionary Change for Your Technology Business*. Blue Hole Press, 2010.

[23] K. Beck and C. Andres, *Extreme Programming Explained: Embrace Change*. Addison-Wesley, 2004.

[24] T. Stahl and M. Voelter, *Model-Driven Software Development: Technology, Engineering, Management*. Wiley, 2006.

[25] VERSIONONE, "8th annual state of agile survey," VERSIONONE, Tech. Rep. 8, 2014. [Online]. Available: http://www.versionone.com/pdf/2013-state-of-agile-survey.pdf

[26] P. Diebold and M. Dahlem, "Agile practices in practice: A mapping study," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, ser. EASE '14. New York, NY, USA: ACM, 2014, pp. 30:1–30:10.

[27] M. Cohn, *Agile Estimating and Planning*. Prentice Hall, 2005.

[28] S. D. Conte, H. E. Dunsmore, and V. Y. Shen, *Software Engineering Metrics and Models*. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc., 1986.

[29] B. Kitchenham, E. Mendes, and G. H. Travassos, "Cross versus within-company cost estimation studies: A systematic review," *Software Engineering, IEEE Transactions on*, vol. 33, no. 5, pp. 316–329, May 2007.

[30] C. Wohlin, P. Runeson, M. H¨st, M. Ohlsson, B. Regnell, and A. Wesslen, *Experimentation in Software Engineering*, ser. Computer Science. Springer, 2012.