A Survey on the Differences of Using User Story and Tasks in the ASD Effort Estimation in Brazil

Diego de Morais[1], Jorge Rady Almeida Jr.[1], Fabio Levy Siqueira[1]

[1]Departamento de Computação e Sistemas Digitais (PCS)
Universidade de São Paulo (USP) – São Paulo – SP – Brazil

{diego.morais, jorgerady, levy.siqueira}@usp.br

Abstract. *This paper investigates the state of the practice of ASD estimation based on User Stories. We conducted a survey with 85 Brazilian professionals experienced in ASD estimating. The survey analyzes what is used in the estimation (User Story, task, or both), its differences, how the estimate is made (especially if there is any segmentation), and the average precision of the effort estimates. The main findings are: 1) Planning Poker is the most used technique and points with a Fibonacci scale as a metric; 2) User Stories are broken down into tasks in the vast majority of teams; 3) Teams that estimate both: User Stories and tasks/subtasks showed greater accuracy compared to the others; 4) At least ¼ of the teams make estimates for the team segmenting by some criteria.*

1. Introduction

Effort estimation is the process by which effort is evaluated in terms of the number of resources needed to deliver an activity that meets the requirements, playing a critical role in any software development project [Trendowicz and Jeffery 2014]. In Agile Software Development (ASD), effort estimation is made during planning, which happens iteratively and at different levels. For example, in Scrum it is done in release planning, sprint planning, and current-day planning [Cohn 2006]. Depending on the agile approach and the project's particularities, it can be performed at different times, with different objects, estimating different activities and using different techniques [Cohn 2006; Trendowicz and Jeffery 2014]. There are several studies on ASD effort estimation investigating these variations [Dantas et al. 2018; Fernandez-Diego et al. 2020; Usman et al. 2014a, 2015]. However, they do not address some specific issues, such as what is used for estimating, the impact of estimating tasks to the detriment of User Stories, and whether estimates are made for the whole team or individually.

The challenges of estimating effort in ASD have been an even more critical theme given the mandatory remote work due to the pandemic (COVID-19), in which distractions, interruptions, and focus have very different characteristics from face-to-face work [Agren and Knoph 2021]. Therefore, this study aims to investigate the state of the practice of ASD estimation based on User Stories, analyzing what is used for estimating (user story, task, or both) and how the estimation is made. In particular, this work investigates the objects and metrics used to estimate, and the existence of segmentation in the teams' estimates and capabilities. Furthermore, following the related works, the average precision of the effort estimates was also evaluated, relating them to the object and metrics used.

The instrument used was a survey, based on a previous survey from Usman, Mendes, and Börstler (2015). The population was limited to Brazilian professionals with experience in estimating ASD and who have used User Stories to represent requirements in a recent project. Our main contribution is presenting the current state in Brazil of effort estimation practice in ASD using User Story. One of the findings is that a significant portion of the teams makes effort estimates for the team segmenting by some criteria. This result will be further investigated as part of our Master's Dissertation.

The remainder of this paper is organized as follows: Section 2 provides the theoretical background for requirements, metrics, estimation, and estimation objects in the context of ASD, while Section 3 summarizes related work. The research methodology is described in Section 4, with its distribution method at the end. The results are discussed in Section 5, along with threats to the validity of this study. Finally, Section 6 describes our findings and future work opportunities.

2. Background

This section provides a theoretical foundation for effort estimation in ASD, covering requirements, size metrics, accuracy metrics, and estimation techniques.

2.1. Requirements and Metrics

There are different representations of requirements that can be used in ASD. A systematic literature review in 2014 on effort estimation [Usman et al. 2014b] found that most works used User Story and Use Case as requirements representation. A survey in 2015 [Usman et al. 2015] found that only 10% of agile projects used Use Case as a requirements specification, with User Story being used in 61% of the projects. These values remained consistent in 2020, according to a systematic literature review [Fernandez-Diego et al. 2020]. Thus, this study is only focused on User Story as requirements representation.

According to Cohn (2004), a User Story describes the functionality valuable to a software user. For industrial professionals, User Story has a crucial aspect: it is the most granular representation of a requirement that developers use to build new features [Lucassen et al. 2016]. User Stories are generally measured in story points [Dantas et al. 2018; Fernandez-Diego et al. 2020], which is a size measurement unit. In most projects, only a predefined set of possible values is used [Cohn 2006]. By definition, story points are relative. Therefore, a User Story that receives two points is assumed to require twice as much effort as a User Story assigned a value of one point. The value can be assigned based on the effort involved, the complexity, and the risk inherent in developing a feature [Coelho and Basu 2012]. A story point usually, albeit not necessarily, corresponds to an ideal workday [Cohn 2006].

Estimating models cannot be evaluated without applying appropriate metrics to measure their accuracy. The most used accuracy metric is Magnitude of Relative Error (MRE) [Fernandez-Diego et al. 2020] – but it is out of our scope to investigate accuracy metrics.

2.2. Estimation Techniques and Objects

According to Usman et al. (2015), most agile teams use estimation techniques that rely on subjective expert evaluation to arrive at a forecast. The most prevalent techniques are Planning Poker, expert opinion, and analogy. It is observed that using these techniques in combination is directly associated to relatively more accurate estimates [Usman et al. 2015]. There is also a notable tendency to study data-intensive techniques using artificial intelligence [Fernandez-Diego et al. 2020].

Planning Poker, expert opinion, and analogy involve subjective evaluations to a greater or lesser extent. Considering studies of human judgment and decision-making by cognitive psychologists [Shepperd et al. 2018], it has been found that human beings generally use heuristics, that is, simple mental strategies that can lead to bad judgments, while sufficient in most circumstances.

In Scrum and Extreme Programming, the team members break down User Stories into individual tasks in the iteration planning [Alyahya et al. 2016]. In this scenario, estimation could be done on only User Stories (requirements), only tasks/subtasks (activities), or both: User Stories and tasks/subtasks. Due to the lack of a generic name for these items, we will use the term estimation object in this study. For example, functional specification, Use Case, User Story, task, and subtask would be estimation objects.

3. Related Works

Three systematic literature reviews (SLR) were conducted to synthesize the state of the art for effort estimation models and practices in agile software development. The first SLR was carried out in 2014 [Usman et al. 2014b], analyzing 25 primary studies from 2001 to 2013. Afterwards, two new SLR updates were performed, the first in 2018 [Dantas et al. 2018], analyzing 15 primary studies between 2014 and 2017, and the other in 2020 [Fernandez-Diego et al. 2020], analyzing 73 primary studies between 2014 and 2020. The results show that effort estimation methods were used in six agile methods (Scrum, XP, TDD, Agile Unified Process, Kanban, and Distributed Agile Software Development). It also shows that Planning Poker has become the most used estimation method. These results also show that the most frequently used size metrics (story points) is closely related to the requirements representation (User Story). Another result is that expert-based estimation methods continue to play an important role and that there is a prominent tendency to study data-intensive techniques. These works conclude that the accuracy of effort estimation models in ASD remains inconsistent and is still a challenge in most analyzed works, although some significant improvements have been identified, such as an increasing number of works reporting adequate ranges of accuracy values [Fernandez-Diego et al. 2020].

Another related work studied the state of practice in 2015 through an online questionnaire with 60 ASD practitioners from 16 countries [Usman et al. 2015]. The study concludes that most agile teams use estimation techniques that rely on subjective expert assessment (Planning Poker, analogy, and expert judgment) and use story points as size metrics. Another result is that the dominant tendency is to underestimate the effort. The main reasons perceived for inaccurate estimates are requirements and management issues.

These researches did not address a specific issue: breaking User Stories into tasks. This theme has already been identified as present in industry practice [Liskin et al. 2014]. Furthermore, the related works do not address the possibility of segmenting (splitting it based on some criteria) the team's estimate and capacity, either.

4. Research Method

This section presents the survey planning based on Kitchenham and Pfleeger (2002).

4.1. Research Questions

Our goal is to obtain the state of practice on how effort estimates are made using User Story in agile software development. The research goes deeper into the issue of breaking User Stories into tasks, the possible impacts of estimating tasks as opposed to User Stories, the average accuracy of estimates, and segmentation in team's effort and capacity estimates. The following research questions were formulated:

- **RQ1:** What objects are used for effort estimation in ASD?
  - **RQ1.1:** What effort estimation techniques are used for each object?
  - **RQ1.2:** What metrics of effort size and scale are used for each estimation object?
- **RQ2:** What is the average accuracy of effort estimation in ASD?
  - **RQ2.1:** How do estimation objects affect the effort estimation accuracy?
  - **RQ2.2:** How do estimation metrics affect the effort estimation accuracy?
- **RQ3:** Are the capacities and effort size estimates segmented during iteration planning in ASD?

RQ1 seeks to understand the objects used to estimate (RQ1.1 and RQ1.2). This study focuses on comparing the estimation of User Stories and tasks/subtasks.

RQ2 focuses on the accuracy of estimates in ASD and the impacts of estimation objects (RQ2.1) and estimation metrics (RQ2.2) used. The intention is to obtain this information from the respondent's perception, without assessing the accuracy metric used, in the same way as in a related survey [Usman et al. 2015].

Finally, RQ3 investigates whether the estimates are made for the whole team or segmented by any criteria, such as specialty (e.g., front-end, back-end, QA).

4.2. Target Audience

The target audience of this study is Brazilian professionals with practical experience in estimating effort in agile software development projects and who work in teams that use User Story to represent requirements. To select experienced professionals, we filtered respondents who participated in the effort estimation process for at least two iterations/sprints. User Stories were chosen as requirement representation because they are the most used according to the related works [Fernandez-Diego et al. 2020; Usman et al. 2015].

4.3. Survey and Planning

The survey was conducted using a self-administered questionnaire written in Portuguese[1]. One concern was to keep it as concise as possible.

Two pilot evaluations were executed with experienced professionals, the first being a business expert (product owner) and the second a senior developer. Based on their feedbacks, we simplified some questions and improved the text, question order, and answer options.

Table 1. Questions of the survey form, indicating its type (T). (C: Closed; O: Open; P: Partially open) and related research questions (goal)

#	Question	T	Goal
Q1	How long is your industry experience with estimating effort in ASD?	C	Filter
Q2	How are user needs and requirements represented?	P	Filter
Q3	What company do you work for?	O	RQ1
Q4	What is the approximate size of the company you work for?	P	RQ1
Q5	What is your main role within the team?	P	RQ1
Q6	Which agile methods does your team practice?	P	RQ1
Q7	How long have most of the team members worked together?	P	RQ1
Q8	What is the nature of the team's main demands?	P	RQ1
Q9	At what moments are effort estimates made?	P	RQ1
Q10	At what moment does the team commit to what will be done?	P	RQ1
Q11	What development activities are taken into consideration when estimating during iteration/sprint planning?	P	RQ1
Q12	What other activities are estimated during iteration/sprint planning?	P	RQ1
Q13	Are user needs and requirements broken down into tasks?	C	RQ1
Q14	What object are estimates made from?	P	RQ1,2,3
Q15	Which estimation technique is most frequently practiced by your team in iteration/sprint planning?	P	RQ1
Q16	In sprint/iteration planning, the capacities and estimates are made for the entire team or are they separated by any criteria?	C	RQ3
Q17	What is the effort size metric used to estimate?	C	RQ1,2
Q18	If the estimate is made in points, which scale is used?	P	RQ1
Q19	What is the average accuracy of the estimates compared to the efforts made by your team?	C	RQ2

The questions formulated are presented in Table 1. The questionnaire starts by filtering the target audience (Q1, Q2), followed by the professional, team, and company context questions (Q3 to Q8), questions about effort estimation (Q9 to Q18), and ends with a question on the accuracy of the estimates (Q19). In the question "Which object are

[1] The survey form in Portuguese is available at https://doi.org/10.6084/m9.figshare.17014184. This survey form file is in its original format, including tool control questions and other questions that will be covered in another paper. The question numbers do not correspond to the numbers in Table 1.

the estimates made from?" (Q14), when answering "both: User Stories and tasks/subtasks", an option was enabled to answer the following questions related to the object (Q15, Q16, Q17, and Q18) for both: User story and task/subtask. Four other questions were asked with complementary objectives to this study, which are subjects for future works.

4.4. Distribution of the Questionnaire

The questionnaire was available between April and May 2021. We recruited participants via: 1) Publications on the authors' profiles in LinkedIn (www.linkedin.com); 2) Emails and messages to the authors' contacts network; 3) Message to professionals working in agile software development teams in a working group of a non-profit independent private law foundation located in the city of São Paulo/Brazil; 4) Snowball effect.

The questionnaire was hosted in Google Forms. The text contained a brief introduction with basic information about the study' objective, justification for the choice, and the importance of the respondent's participation. Participants were also informed about the study privacy policies in a clear and detailed manner.

5. Results and Discussion

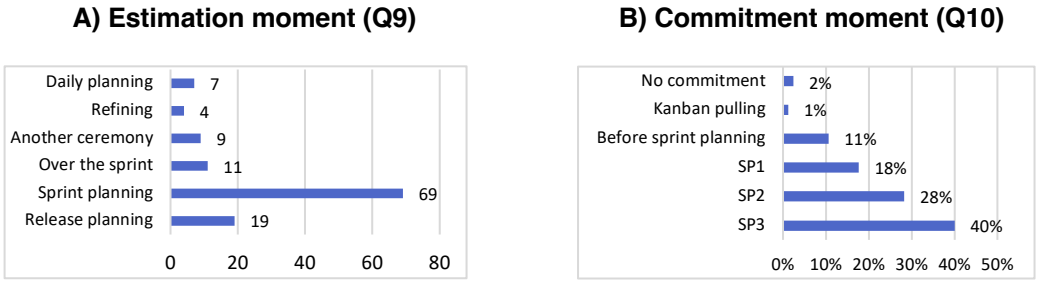
We obtained 139 responses[2]. Considering the target audience (Section 4.2), we excluded responses from people without experience in ASD estimation (Q1), who used a requirements representation different from User Story (Q2), or from professionals from teams that did not make estimations (Q14). Therefore, 54 responses (6, 47, and 1 respectively) were removed, resulting in 85 valid responses for this analysis.

In this sample, 5% of responses are from a single company (Q3). The authors infer that most companies are in Brazil due to convenience sampling and because the question form is in Portuguese.

The size of the companies the respondents work for (Q4) is mostly large – more than 100 employees (85%) – and most professionals perform the role (Q5) of developer (45%). Most of them use a combination of Scrum and Kanban (63%) as agile methods (Q6), followed by only Scrum (28%). Most of their team members have worked together (Q7) for 1 to 3 years (35%). The nature of their teams' main demands (Q8) is mainly new developments (90%) versus only 10% for support/maintenance.

The effort estimation can happen at several moments in the same team (Q9 – multiple selections), being the sprint planning ceremony the predominant one, as shown in Figure 1 – A. The sprint planning is also the most common ceremony in which the team commits to what will be done (i.e., the backlog items that the team commits to completing during the sprint - Q10). More specifically, this is typically done after estimating and after calculating the team's capacity, as shown in Figure 1 – B.

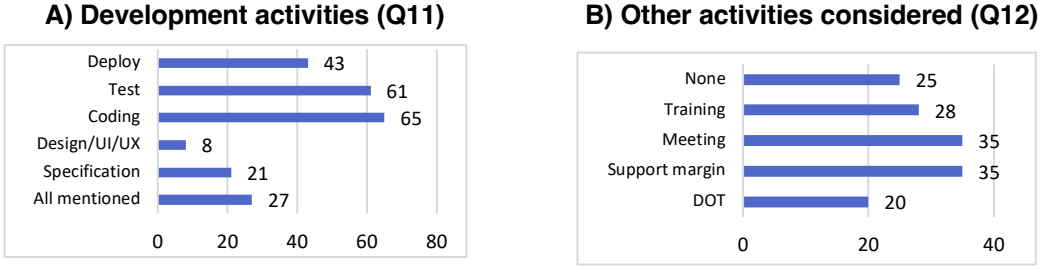
[2] The survey dataset (in Portuguese) is available in the following link: https://doi.org/10.6084/m9.figshare.17014214 - The answers for Q3 were removed to preserve anonymity.



Caption:
SP1: During the sprint planning – before estimating
SP2: During the sprint planning – after estimating and after calculating the team's capacity
SP3: During the sprint planning – right after estimating

Figure 1. Estimation and commitment moments

Coding and testing are the most frequent development activities taken into account when estimating (Q11 – multiple selections), followed by deploy, as shown in Figure 2 – A. Other activities are also considered when estimating (Q12 – multiple choice), with meetings and support margin (i.e., time retention to troubleshoot production bugs or other maintenance activities) being the most frequently chosen, as presented in Figure 2 – B.



Caption:
DOT: Demands from other teams

Figure 2. Activities estimated

5.1. What Objects are Used for Effort Estimation in ASD? (RQ1)

User Stories are broken down into tasks (Q13) by the teams of 73% of respondents. For 21% it varies – some are broken, and some are not. The User Stories are not broken for only 6% of respondents, as seen in Figure 3 – A.

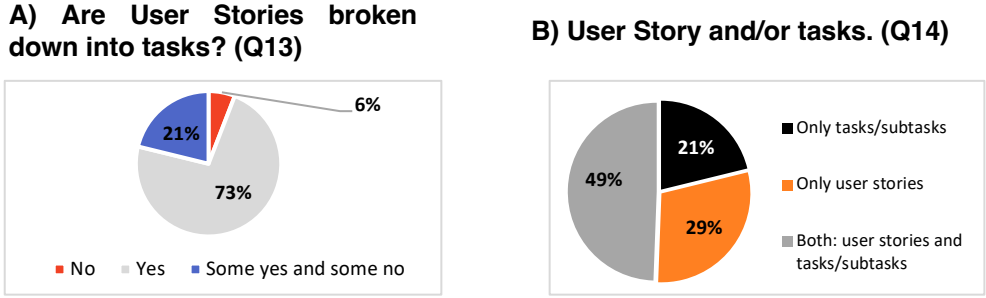


Figure 3. Estimation objects used

Almost half (49%) of teams estimate effort on "both: User Stories and tasks/subtasks" (Q14), although a little more than ¼ (29%) estimate "only User Stories" and the rest (21%) estimate "only tasks/subtasks", as shown in Figure 3 – B.

RQ 1.1: What effort estimation techniques are used for each object?

In the question about the effort estimation technique (Q15), when the respondent selected that he estimated "both: User Story and task/subtask", the option for answering an equal or different technique for each object was enabled. Therefore, each respondent of "both: User Story and task/subtask" has two responses counted, one in the object user story (identified as "User story - in both option") and another in the object task/subtask (identified as "Task - in both option"). The other respondents have a single response in the corresponding object ("only User Story" or "only task/subtask").

In this question, we included the answer option "By who will implement" even though this is not an estimation technique per se.

Planning Poker was the predominant technique for all estimation objects. When the object is "only User Story", Planning Poker is even more prevalent with 72%, and the other techniques are less used. The same does not occur when the object is "User story - in both option" or "Task - in both option", there is a more even distribution of the techniques, as shown in Figure 4.

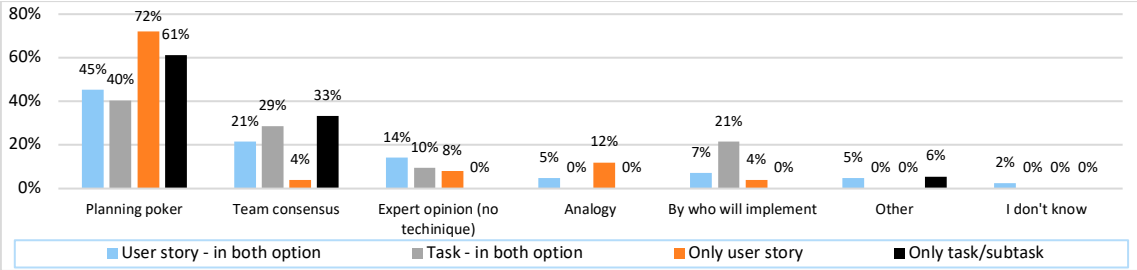


Figure 4. Estimation technique by the object used

The values indicate differences in the technique used concerning the estimation object, with a clear distinction of technique when the use of the object is exclusive (only an estimation object, i.e. "only User Story" or "only task/subtask") or combined ("both: User Story and task/subtask").

RQ 1.2: What metrics of effort size and scales are used for each estimation object?

The most commonly used effort size metrics (Q17) are Points for User Story, either when the object is exclusive ("Only User Story) or combined ("User story – in both option"), representing 80% and 69%, respectively. When the objects used are tasks/subtasks, the metrics were well distributed between points and hours, either when it is "only tasks/subtasks" (39% in hours and 61% in points) or the object is combined "Task – in both option" (48% in hours and 38% in points). Few teams use "days" as a metric. The results are shown in Figure 5 – A.

The scale used for the points was complemented with another question (Q18), whose results are shown in Figure 5 – B. When the estimation object was "both: User

Stories and tasks/subtasks", the scale was asked once. It was not asked for each object separately (which can be considered a problem in the survey protocol). The most used scale was Fibonacci, with dominant 95% for "only User Stories", 69% for "only tasks/subtasks", and 50% for "both: User Stories and tasks/subtasks". The object "only tasks/subtasks" obtained a relevant use of T-shirt size (33%).

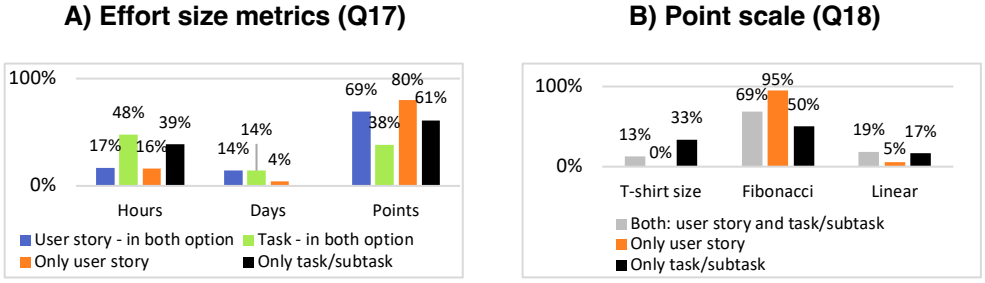


Figure 5. Effort size metrics by the object used

Considering the answers that use Planning Poker as the estimation technique (Q15) and User Story ("both: User story" and "only User Stories" from Q14), we analized the answers of size metrics (Q17). The answer "points" got the vast majority of responses (32 – 86%). In comparison, "hours" got 3 responses (8%) and "days" only 2 responses (5%). There is a clear relationship between the most used effort size metric (Points), the primary estimation technique (Planning Poker), and how requirements are specified in ASD (User Story). This relationship and result are consistent with related works [Fernandez-Diego et al. 2020].

5.2. What is the Average Accuracy of Effort Estimation in ASD? (RQ2)

Of the 85 valid respondents in this study, 13 (15%) responded that they did not know the mean accuracy of their team's estimates, 32 (38%) said they did not measure (Q19), and only 40 (47%) respondents evaluated the estimates. More than ⅓ of the responses do not measure the estimates accuracy. Similarly, the related RSLs [Fernandez-Diego et al. 2020; Usman et al. 2014a] also reported some primary studies not presenting the accuracy of the metric used, 26% and 33%, respectively.

It is not the focus of this study to analyze how accuracy is measured. Therefore, the answers only reflect the respondent's perception, similarly to Usman et al. (2015). We presented fixed value ranges as answer options, converted to the average for presentation in the graphs of Figures 6, 7, and 8.

The highest number of answers (11) was the best average accuracy ("between underestimated 5% and overestimated 5%"), followed by the following two ranges ("underestimated between 5% and 25%", and "overestimated between 5% and 25%") in the second place – both with 8 responses each, as shown in Figure 6. These results differ from those in Usman et al. (2015), whereby the range with most responses (35% of the total) was "underestimated between 25% and 50%", suggesting a greater tendency to underestimate than to overestimate.

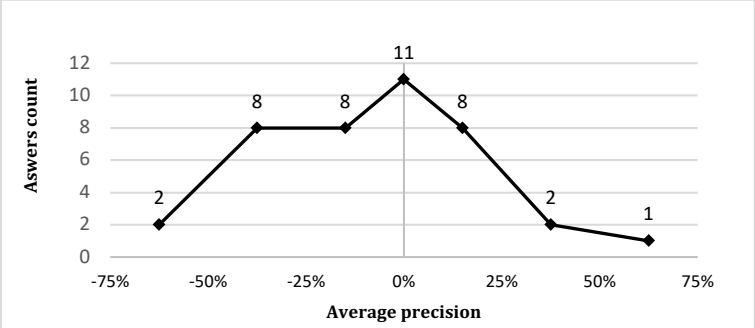


Figure 6. Average accuracy by answers count

RQ 2.1: How do estimation objects affect the effort estimation accuracy?

Of the 40 respondents who measure the estimate, 19 (47%) use "both: User Stories and tasks/subtasks", 12 (30%) use "only User Stories", and 9 (23%) "only tasks/subtasks".

The estimation object "both: User Stories and tasks/subtasks" was the only one to present a high number of responses in the best range (between underestimated 5% and overestimated 5%). In contrast, the others have a more uniform distribution between the ranges, as shown in the graph of Figure 7.

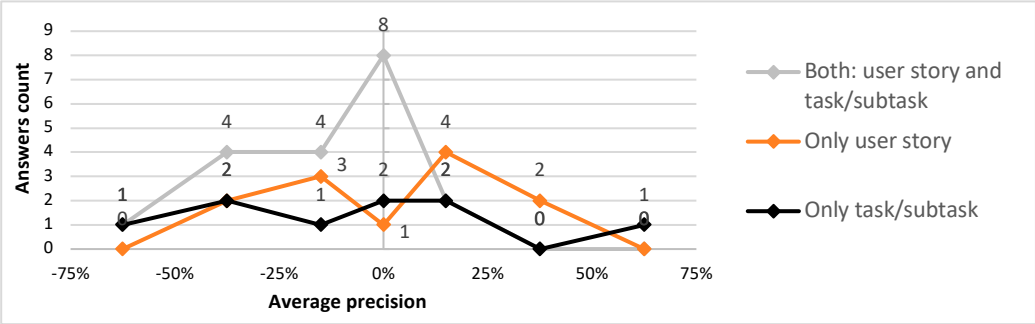


Figure 7. Average accuracy by the estimation object used

RQ 2.2: How do estimation metrics affect the effort estimation accuracy?

When the respondent answered "both: User Stories and tasks/subtasks" as the estimation object in Q14, the effort size metric was asked separately in Q17 – for User Stories and tasks/subtasks – but only one for accuracy was asked in Q19. To avoid inferring an answer and to make sure that the same answer was not counted twice, only answers that have the same effort size metric were considered when the person selected "both: User Stories and tasks/subtasks". Table 2 presents how we considered the data. With this process, 14 respondents were excluded from this analysis. Respondents who answered "only User Stories" and "only tasks/subtasks" were considered normally.

Table 2. Exclusion filter example for insufficient data (RQ 2.2)

	Estimate object (Q14)	Size metrics (Q17)		Result
		User Story	Tasks/subtasks	
Respondent A	Both: User Stories and tasks	Points	Points	considered
Respondent B	Both: User Stories and tasks	Points	Hours	not considered
Respondent C	Only tasks/subtasks	-	Hours	considered
Respondent D	Only User Stories	Days	-	considered

The metric "Points" was the only one to present the highest concentration of responses in the best ranges ("between underestimated 5% and overestimated 5%", "underestimated between 5% and 25%", and "overestimated between 5% and 25%"). In contrast, the others have a more uniform distribution between the intervals, as shown in Figure 8.

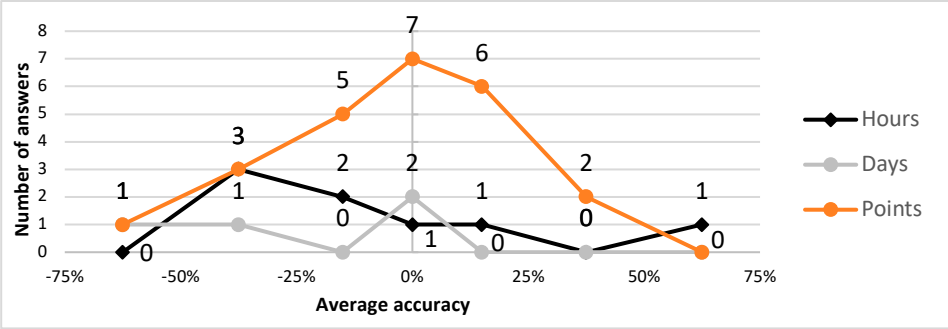


Figure 8. Average accuracy by effort size metric

5.3. Are the Capacities and Effort Size Estimates Segmented During Iteration Planning in ASD? (RQ3)

According to Stray et al. (2018), the first introduction of autonomous teams in software engineering was made by the agile manifesto [Beck et al. 2001], referring to the self-organization of teams as the origin of the best architectures, requirements, and designs. Stray et al. complements saying that perhaps the most straightforward definition of autonomous teams comes from a knowledge management perspective, as those exhibit three conditions: autonomy, cross-fertilization, and self-transcendence [Stray et al. 2018]. For the context of this study, the definition of cross-fertilization becomes particularly important: "refers to the team being formed from individuals with different specializations, behavior, and thinking so that regular interaction improves their understanding of each other's perspectives." [Lundene and Mohagheghi 2018 p. 2]

With teams having different technical specializations and, consequently, different skills in each specialization, we suspect that some teams have difficulties performing a single capacity estimation and planning for the whole team. This suspicion is based on the observation of Brazilian companies' ASD practice, in which specializations such as front-end, back-end, and quality assurance (QA) are specific to different professionals. The questionnaire shows that some teams segment the effort estimation (Q16) by different criteria, as presented in Figure 9. For example, a User Story or task may be estimated to be 2 points for the front-end, 2 points for the back-end, and 1 point for QA – instead of

simply 5 points (this example was presented in the questionnaire to explain what segmentation is).

Most teams perform a single effort estimation for the whole team. User Stories have a higher rate for this response, either in exclusive form ("only user stories") or combined form ("User story – in both option"), representing 80% and 69%, respectively. Segmented by specialty (e.g., back-end, front-end, and QA) occupies second place. When the estimation object is tasks/subtasks, it is even more representative, 38% for the combined form ("Task – in both option", and 28% for the exclusive form ("Only tasks/subtasks").

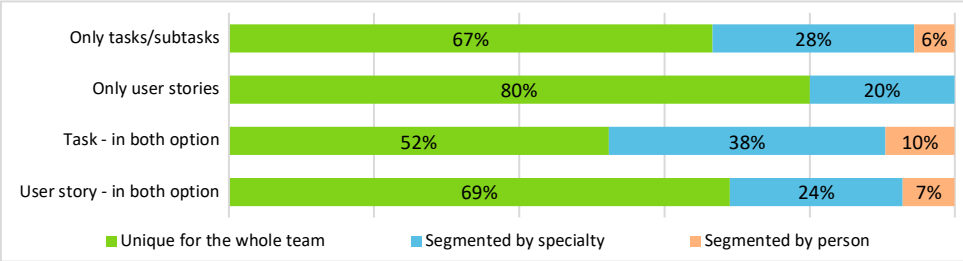


Figure 9. Segmentation of estimates by the object used

A similar distribution of responses among all objects indicates that the separation by estimation objects is not decisive for segmentation.

5.4. Threats to Validity

This section discusses threats to the validity of this work. We use the organization suggested by Wohlin et al. (2012).

The construct validity concerns the threats that may arise due to the improper construction of the research instrument. One threat is that the questions might not be adequate for what we want to measure. To mitigate this threat, we used questions from a related survey [Usman et al. 2015] (Q11, Q15, Q17 and Q19), as we believe they are already more mature and validated. Another threat is that the questionnaire might not address all the research questions. During the questionnaire construction process, each question was mapped to a research question to mitigate this.

Internal validity is concerned with issues such as relevant respondents, confounding factors, and bias in results. One threat is that the selected respondents were not part of the target audience. The answers to the first two questions in the questionnaire were used as exclusion criteria to ensure that the respondent belonged to the population. This procedure led to the exclusion of many responses (as discussed in Section 5). In addition, the respondent was asked about his role within the team (Q5) to identify biases and possible limitations in the answers for not participating in a team or not having visibility of the process. Another threat is respondents not being honest in responses that show problems in their organization. Respondents were assured of their anonymity to minimize this. Finally, another threat is the respondents' failure to understand the issues. A pilot test was conducted with experienced professionals (Section 4.3) to mitigate it.

Conclusion validity concerns whether the conclusions drawn are reasonable regarding the data collected, such as the relationship between factors or inappropriate statistical measures. This study used only frequencies and percentages to identify

common patterns or practices to point out potential areas or relationships for future research efforts. Furthermore, only complete answers were considered in the result analysis.

External validity is related to generalizing the results. The main threat is that we used a convenience sample. To mitigate the risk of respondents not being representative, questions Q3 to Q8 identify the respondent's context. To obtain a broad population sample, actions such as announcing the survey in the authors' social network and a snowball approach were taken. Analyzing the data, we verified that responses from a single company did not exceed 5%. The authors believe that these steps contributed to obtaining a heterogeneous sample in terms of company, team, and role.

6. Conclusion and Future Work

This work presented a survey to investigate the state of the practice of ASD estimation on projects that use User Stories. We analyzed what is used for estimating (User Story, task, or both), its differences, and how the estimation is made (particularly if there is any segmentation). Following Usman et al. (2015), we also evaluated the average accuracy of the effort estimates. The survey was applied between April and May 2021, collecting 139 responses, with 85 considered valid, given our target audience.

The combination of Kanban and Scrum is the most used agile method among respondents. The most used estimation technique is Planning Poker, similar to the results of related works [Dantas et al. 2018; Fernandez-Diego et al. 2020]. Fibonacci scale points are the most effort size metric that respondent teams use. User Stories are broken down into tasks in most teams, and half of them estimate for both: User Stories and tasks. We investigated the relationship between estimation objects (Q14) and the characteristics of the company, project, and team (Q4-Q12). We did not find any correlation between the variables, considering a two-sample Z test [Zou et al. 2003] and $\alpha = 0.05$.

About ⅓ of the teams do not assess the accuracy of the effort estimate (RQ2). The teams that used both: User Story and task/subtask for estimating presented a better accuracy than the others in this study (RQ2.1). Future work may investigate the reasons for these results. Regarding the greater accuracy, we suspect that estimating both allows for refining requirements and understanding the team's problem and solution more accurately. However, a detailed analysis of each estimation moment needs to be researched, including evaluating this approach positive and negative impacts.

At least ¼ of the teams make effort estimates for the team segmenting by some criteria, with the segmentation by technical specialty (technical functionality) being the most answered. We did not find studies about effort estimation segmentation despite our perception of its existence in the Brazilian scenario. We observe the Brazilian industry considering technical specialization, such as back-end, front-end, and QA, as different competencies in the same team, with professionals working exclusively in one of them.

To summarize, this study contributes to presenting the current state of effort estimation practice in ASD using User Story (in Brazil):

- Confirm that Kanban and Scrum remain the most used agile methods with User Story, as well as the Planning Poker estimation technique;
- Evidence that many teams continue not to measure the accuracy of effort estimates.

Moreover, this study brings new relevant information to the academy (at least for this sample):

- Most teams break down User Stories into tasks, and half of them estimate for both;
- The teams that used both: User Story and task for estimating presented a better accuracy than the others;
- A significant portion of the teams makes effort estimates for the team segmenting by some criteria.

As future work, our Master's Dissertation study will continue investigating the topic of segmentation within the team to estimate effort and effort capacity through the Grounded Theory qualitative analysis method. We hope to understand the motivation, how it works, the difficulties and benefits of the teams that segment, in addition to the differences for the teams that do not segment. We also hope to understand better the relationship of segmentation to user stories and tasks.

References

Agren, P. and Knoph, E. (2021). *COVID-19's Impact on Agile Software Development*. Master thesis.

Alyahya, S., Alqahtani, M. and Maddeh, M. (2016). Evaluation and improvements for agile planning tools. *2016 IEEE/ACIS 14th International Conference on Software Engineering Research, Management and Applications, SERA 2016*, p. 217–224.

Beck, K., Grenning, J., Martin, R., et al. (2001). Manifesto for Agile Software Development. *The Agile Alliance*, p. 12–14.

Coelho, E. and Basu, A. (13 aug 2012). Effort Estimation in Agile Software Development using Story Points. *International Journal of Applied Information Systems*, v. 3, n. 7, p. 7–10.

Cohn, M. (2004). *User stories applied : for agile software development*. 13th. ed. Addison-Wesley.

Cohn, M. (2006). Agile estimating and planning. In *VTT Symposium (Valtion Teknillinen Tutkimuskeskus)*. . , [accessed on Oct 2].

Dantas, E., Perkusich, M., Dilorenzo, E., et al. (2018). Effort Estimation in Agile Software Development: An Updated Review. *International Journal of Software Engineering and Knowledge Engineering*, v. 28, n. 11–12, p. 1811–1831.

Fernandez-Diego, M., Mendez, E. R., Gonzalez-Ladron-De-Guevara, F., Abrahao, S. and Insfran, E. (2020). An Update on Effort Estimation in Agile Software Development: A Systematic Literature Review. *IEEE Access*, v. 8, p. 166768–166800.

Kitchenham, B. A. and Pfleeger, S. L. (2002). Principles of survey research part 2. *ACM SIGSOFT Software Engineering Notes*, v. 27, n. 1, p. 18–20.

Liskin, O., Pham, R., Kiesling, S. and Schneider, K. (2014). Why we need a granularity concept for user stories. *Lecture Notes in Business Information Processing*, v. 179 LNBIP, p. 110–125.

Lucassen, G., Dalpiaz, F., Van der Werf, J. M. E. M. and Brinkkemper, S. (2016). The use and effectiveness of user stories in practice. In *Lecture Notes in Computer Science*. . Springer, Cham.

Lundene, K. and Mohagheghi, P. (21 may 2018). How autonomy emerges as agile cross-functional teams mature. In *Proceedings of the 19th International Conference on Agile Software Development: Companion*. . ACM. https://dl.acm.org/doi/10.1145/3234152.3234184.

Shepperd, M., Mair, C. and Jørgensen, M. (9 apr 2018). An experimental evaluation of a de-biasing intervention for professional software developers. In *Proceedings of the ACM Symposium on Applied Computing*. . Association for Computing Machinery.

Stray, V., Moe, N. B. and Hoda, R. (2018). Autonomous agile teams: Challenges and future directions for research. *ACM International Conference Proceeding Series*, v. Part F1477, p. 1–5.

Trendowicz, A. and Jeffery, R. (2014). *Software Project Effort Estimation*. 1. ed. Cham: Springer International Publishing.

Usman, M., Mendes, E. and Börstler, J. (27 apr 2015). Effort estimation in Agile software development: A survey on the state of the practice. In *ACM International Conference Proceeding Series*. . Association for Computing Machinery.

Usman, M., Mendes, E., Weidt, F. and Britto, R. (17 sep 2014a). Effort estimation in Agile Software Development: A systematic literature review. In *ACM International Conference Proceeding Series*. . Association for Computing Machinery.

Usman, M., Mendes, E., Weidt, F. and Britto, R. (17 sep 2014b). Effort estimation in Agile Software Development: A systematic literature review. In *ACM International Conference Proceeding Series*. . Association for Computing Machinery.

Wohlin, C., Runeson, P., Höst, M., et al. (2012). *Experimentation in software engineering*. 1. ed. Berlin, Heidelberg: Springer Berlin Heidelberg. v. 9783642290

Zou, K. H., Fielding, J. R., Silverman, S. G. and Tempany, C. M. C. (2003). Hypothesis testing I: Proportions. *Radiology*, v. 226, n. 3, p. 609–613.