

# Are we teaching UML according to what IT companies need? A survey on the São Carlos-SP region

João Choma Neto

joaochoma@usp.br

Institute of Mathematics and Computer Sciences,  
University of São Paulo  
São Carlos, São Paulo, Brazil

Edson Oliveira Jr

edson@din.uem.br

Informatics Department  
State University of Maringá  
Maringá, Paraná, Brazil

Luiz Henrique Ten Caten Bento

luiz.henrique.bento@usp.br

Institute of Mathematics and Computer Sciences,  
University of São Paulo  
São Carlos, São Paulo, Brazil

Simone R. S. Souza

srocio@icmc.usp.br

Institute of Mathematics and Computer Sciences,  
University of São Paulo  
São Carlos, São Paulo, Brazil

## ABSTRACT

The Unified Modeling Language (UML) arisen intending to unify the many modeling languages and become the universal language for software modeling. Since its creation, more than 25 years have passed. With the growing success of agile methodologies for software development, supported by less modeling and documentation manifest, the use of UML-driven software processes has decreased, thus a specific subset of diagrams has been adopted, such as use case, class, and sequence. On the other hand, in academia, UML is still being taught in Computing courses, most of the time without knowing whether what is taught is what the industry needs for practical use. This paper presents an overview of the UML adoption in IT companies of the São Carlos - SP and region. We analyzed quantitative and qualitative data to support academia at focusing on the most used UML diagrams by practitioners. We constructed a survey composed of 21 questions, distributed to 10 region companies, and we received 24 answers. The results of our study show high usage of UML, including companies adopting agile methods. Certain diagrams are more intensively used, for instance, use case diagrams. The results provide directions to improve UML teaching, focusing on diagrams that best adhere to the development processes employed, in particular, agile processes.

## CCS CONCEPTS

• **Software and its engineering** → **Software development process management**; • **Social and professional topics** → **Software engineering education**.

## KEYWORDS

Software Engineering, UML, Survey, UML Diagrams, Software Modeling Teaching

## 1 INTRODUCTION

Currently, increasingly faster, more complex and better software products are needed to meet the needs of society in general. To do so, the software development process plays a central role to provide such solutions. Development processes are composed of a set of essential activities, such as planning, development, and implementation [28].

As modeling is an essential part of software process activities and it leads to the implementation of a software system [12, 13, 28, 31], in 1997, the *Unified Modeling Language* (UML) emerged to unify existing different modeling techniques at that time. Its purpose of unification has been achieved by becoming an universal language for software modeling, yet being the most used language for modeling software systems [9, 21].

Since then, UML has been taught in the academic grades of courses related to software development, such as Computer Science, Software Engineering, and Computer Engineering. Therefore, due to a large number of diagrams, it is a common practice to teach on more used diagrams by software practitioners. However, academia must be aware of which UML diagrams the software industry really uses, thus improving the impact on taught diagrams in the practitioner's professional routines and, consequently, focusing on what really matters on a modeling-related course.

Existing literature discusses UML learning and teaching in a variety of works, such as, in [3, 4, 27, 30]. However, such works do not compare what they teach to what the software industry needs. Motivated by this theme, in this paper, we concentrate our efforts on demonstrating whether UML teaching is directed to what software industry in the São Carlos - SP region really needs. Thus, this work has a practical and educational motivation, aiming to collect data to allow the elaboration of a regional panorama of the UML use, as well as the identification of the degree of adherence between what is taught and what is actually used by software companies in the region.

This study was divided into two aspects: the educational one, seeking to bring a vision of what is required of IT professionals; and the scientific one, bringing updated data on the use of UML. The results will assist in the current understanding of what is taught in

The author(s) or third-parties are allowed to reproduce or distribute, in part or in whole, the material extracted from this work, in textual form, adapted or remixed, as well as the creation or production based on its content, for non-commercial purposes, since the proper credit is provided to the original creation, under CC BY-NC 4.0 License.  
*EduComp'21, Abril 27–30, 2021, Jataí, Goiás, Brasil (On-line)*

© 2021 Copyright held by the owner/author(s). Publication rights licensed to Brazilian Computing Society (SBC).

academia and what is used in the corporate environment. In addition, the reasons for the use or disuse of UML and how academia has contributed to training professionals according to the expectations of the industry will be showed. The study was inspired by works that used the data collection process called Survey [3, 4, 25, 27, 30]. In addition, we use the guidelines of empirical studies [26].

This paper is organized as follows: Section 3 presents the study's methodology; in Section 4 the results are presented, then discussed in Section 5; Section 6 discusses threats to validity of this work; and Section 7 presents conclusions and directions for future work.

## 2 BACKGROUND

Taking into account the software development ecosystem present in the city of São Carlos - SP and region, we developed a literature review aiming to gather evidence about UML teaching techniques in the São Carlos - SP region. However no information was found on the adhesion of the UML modeling language and compatibility between the study objects brought by the city's educational institutions and what is used in the development context in these companies. But we found similar searches from different regions and periods that inspired us.

The research was done in Europe in 2014 [24] bringing results about science and the use of UML in the professional as well as in the academic context. The survey had an international scope and has an interesting result for comparing and basing the questions for the survey. Another research [18] brings the perception within the academic environment about UML diagrams at the Federal University of Amazonas. In a qualitative view, the study analyzed the perception in the professional environment regarding the impact of using UML [19]. And, more similar to the scope of this study, we have studies, one survey conducted in Europe in 50 different companies and also seeking to understand the use of UML in the professional environment and why the use of UML or not [22], quantitative and qualitative analysis regarding the reasons for the use or not of UML. Research carried out in 2006 [7], and research conducted in the state of Minas Gerais with a focus on relating the type of agile methodology with the use of UML [32].

In addition, we found studies that pointed to the UML disuse: [23] brought as a result that 70 % of respondents did not use UML; [8] observed 73.2 % of disuse; and [33] pointed out 71.4 %, the latter with restrictions on only companies that used agile methodology. Concerned with these data, a recent study [6] the has evaluated multiple attributes of startups and identified gaps in the current state of software engineering education, to collaborate on the development of an industry-focused curriculum.

Based on the literature and existing techniques existing empirical research, we have a practical and educational motivation, aiming to collect data that allow the elaboration of a regional panorama of the use of UML, as well as the identification of the degree of cohesion between what is taught at the academic level and what has used in development process within the process of companies in the region of São Carlos - SP.

## 3 METHODOLOGY

In this section, we present the methodology of our study. We followed the survey process by Kitchenham et al. [16], which is composed of six main activities: defining the survey objectives; survey

design; survey development; survey validation; data validation; and data analysis. We chose a survey as a methodology for this study because of its benefits such as the possibility to remotely collect data, the time-frame independence for data collection, and the ease of distribution [10]. We tried to mitigate problems with surveys by following the recommendations of Ghazi et al. [14] and Linåker et al. [17].

### 3.1 Survey objectives

We defined the survey objectives as follows:

- Address how UML is used in IT companies in the São Carlos - SP region;
- Identify which diagrams are most used by IT companies in the São Carlos - SP region.
- Identify the level of knowledge in UML by employees of IT companies in the São Carlos - SP region?
- Gather inputs on how UML teaching can be improved in the São Carlos - SP region?

### 3.2 Survey design

We designed our survey as a cross-sectional one, in which the participant answers the questionnaire once based on his/her previous experiences [16].

### 3.3 Survey development

The survey development was inspired in similar works [2, 8, 11, 21, 33]. We analyzed the questions used in these works and added suitable ones to our survey. We also incorporated additional questions, thus generating a survey composed of 21 questions. We implemented the survey using *GoogleForms* [15] and it accepted responses for 15 days, from March 27th, 2020 to April 11th, 2020.

The 21 mandatory questions of the survey are divided into:

#### General data

- 1 Name
- 2 Email
- 3 Higher graduation level?
- 4 Are you a student?
- 5 What educational institution did you graduate? (or is graduating)
- 6 What is the city of the educational institution?
- 7 Do you work in the software development area?

#### Professional Data

- 8 Which company?
- 9 What is the company's city?
- 10 What is your position?
- 11 How many years of experience do you have? (College years + Professional)
- 12 Does your company use agile methodology in software development?
- 13 What software development methodology does your company use?
- 14 What adaptations in this development methodology are visible?

#### Experience with UML

- 15 What is your experience with the UML notation and its diagrams?
- 16 Which of the following diagrams are you aware? 12 options of diagrams
- 17 Which of the following diagrams are used in your company? 12 options of diagrams
- 18 In your opinion, what does motivate your company to use or not to use UML?
- 19 If you don't use UML (or use it partially), what other documentation resources for software modeling do you use?
- 20 What difficulties do you, as a professional in the field, observe when using UML?
- 21 In your opinion, what should change/improve in teaching UML modeling?

### 3.4 Survey validation

For this activity, we chose by convenience one of the target companies to test our survey, named Company A, because of the proximity of the authors with such company, thus making it easier to collect answers and feedback. We made the survey available for four days for Company A. Such a company contributed important data to a better survey version. For instance, the initial version asked the participants "Do you work in the industry?" with the idea of the software industry, but this question caused misunderstandings, thus we changed it to "Do you work in the software development area?".

### 3.5 Valid data collection

After the survey distribution and data collection, this activity analyzes the data collected, aiming to verify if the sample corresponds to the interest group. We intend to receive three answers from each of the ten selected companies. Therefore, our set is composed of 30 independent responses, as suggested in [1], ensuring, thus, the statistical validity of the data.

### 3.6 Data analysis

The last activity described in [16] consists of analyzing the collected data to answer the research objectives. Section 4 describes the data analysis. Quantitative questions were measured and presented in infographics. For the descriptive questions, we use the method called Grounded Theory [5, 29], responsible for directing the analysis and interpretation of the responses. The survey package containing the survey and all the results is available at *repository*<sup>1</sup>.

### 3.7 Distribution

We adopted three different strategies for distributing the survey:

- (1) contact known company's employees;
- (2) contact persons in a professional social network, *Linkedin*, who work for a company in the region;
- (3) contact persons by email of company.

The first strategy consisted of a pre-existing contact in the companies which are part of the target audience of the study. Thus, the contact was requested to answer the survey and to ask his/her colleagues to answer it as well. This strategy was the most successful,

thus responsible for the largest amount of answers obtained (80%) from companies: A, B, C, D, and E.

The second strategy consisted of finding target companies' employees in the social network *Linkedin* and contacting them, explaining the study, and requesting them to answer the questionnaire. Also, if the candidate did not answer the message or the survey within three days, other employees from the same company were contacted and, in parallel, the third strategy was implemented.

The third strategy was implemented only in the cases that the first and second strategies failed. After three days of no response, we sent an email to the company, explaining the study, and asking them to solicit their employees to respond to the survey. This strategy was implemented for three cases, of which one answer was obtained, resulting in a 33% of third strategy success.

### 3.8 Sample

After the period of data collection, we processed the obtained data to confirm the validity of the answers. In this processing, answers from companies in cities outside of the defined target region and had incomplete answers were discarded.

The original number of answers was 27. However, after the data processing, we discarded three of these answers: one from a participant that works in São Carlos - SP (outside the defined target region); one for having a generic name and email (impossible to verify the validity); and one incomplete answer (blank answer).

Therefore, the final amount of valid answers was 24, representing all of the ten target companies. All companies chosen as target audience answered the survey. In the analysis, the participants and companies are identified by numbers and letters for identity preservation.

## 4 RESULTS

In this section, we present results from our survey. The analysis follows the survey's questions order. In the end, additional analysis is presented, obtained by treating the collected data and comparing results from different questions.

### 4.1 Demographics

In terms of the participants' education level, the found scenario shows the majority of the participants are graduated (46%) or undergraduates (42%), followed by the percentage of the participants with Master's degree (8%) and the number of participants with a *lato sensu* postgraduate degree(4%) (Figure 1).

The participants' education level distribution endorses one of the study's goals: to evaluate the academic efficiency of preparing its students to the work market, as the distribution shows that the majority of the participants were transitioning from the academic environment to the professional environment at the moment or transitioned recently.

In terms of the participants' educational institution, the results show a distribution of 58% of the participants came from *Universidade de São Paulo (USP)*, São Carlos - SP campus, followed by 25% from Institution *Universidade Federal de São Carlos (UFSCar)*, São Carlos - SP campus and, lastly, 17% from *Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)*, São Carlos - SP

<sup>1</sup><https://github.com/JoaoChoma/educomp2021-results-of-survey>

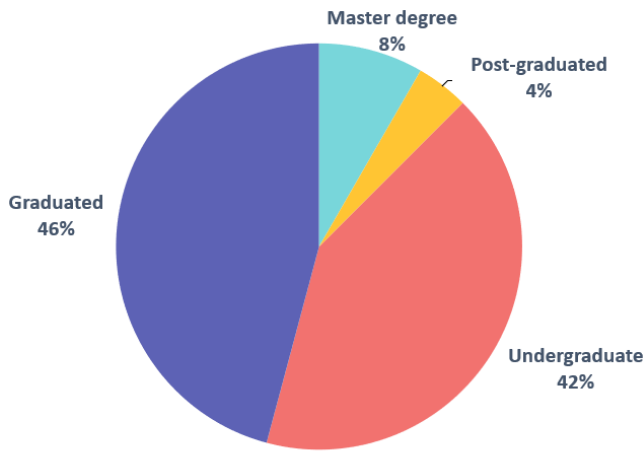


Figure 1: Participants education level.

campus, as represented in Figure 2. This distribution shows all participants were/are enrolled with institutions from the city of São Carlos - SP, evidencing a predominance of professionals from the region's institutions, thus reinforcing one of the study's objectives: to collect inputs to the academic environment, identify how UML was being adopted by IT companies, and thus, improve the teaching of UML in educational institutions in São Carlos - SP.

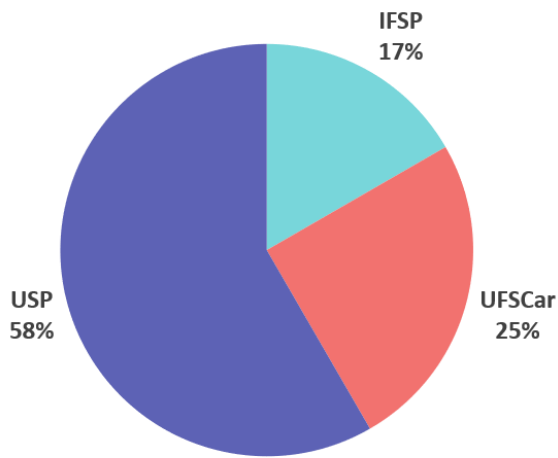


Figure 2: Educational institutions where the participants come from.

For ethical and privacy reasons the name of the companies has been deleted. The obtained distribution for the participants' companies is represented in Figure 3. This distribution shows a predominance of participants from three companies: Company A, with 25% of the participants, Company B, with 21%, and Company C, with 17%. This predominance comes from the preexisting contact between authors and these three companies representatives. Thus, this relationship favored the internal disclosure and, consequently,

a greater adherence from these companies. On the other hand, the companies with less participation were: Companies F, G, H, I, and J, with 4% each. This small percentage does not constitute a threat to data validity due to the fact that the representatives from these companies were very experienced (7 years in average), granting the reliability of the answers.

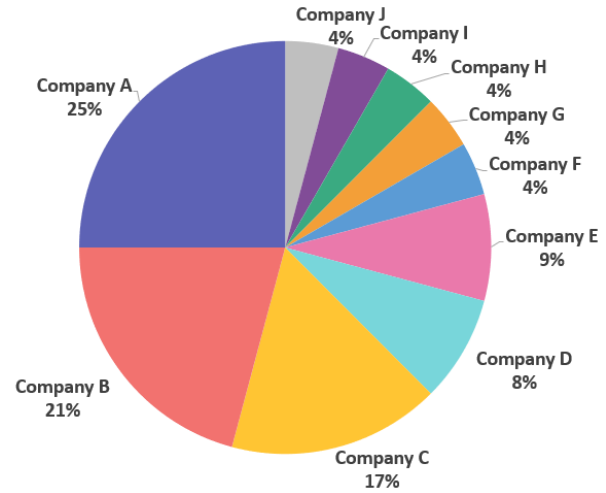


Figure 3: Number of participants by company.

Maintaining the focus on understanding the participants' profile, the participants' role question has its distribution represented in Figure 4. We can observe the majority of participants is Software Developers, with 63% of the answers, followed by Software Engineers, with 17%, Quality Analyst, with 8%, and Project Manager, Business Intelligence, and Software Seller with 4% each. This distribution is interesting for the study as the majority of the participants are directly involved with the development process and, consequently, system modeling and UML. The diversity of roles and levels of experience at work contributed to enrich the discussion.



Figure 4: Participants roles.

The last question focused on the participants' profile is their experience in years. Grouping the results in five-year intervals (to facilitate viewing the results), the distribution obtained can be verified in Figure 5. The chart shows a predominance of participants with experience between 5 and 10 years, on 46% of the sample, then professionals with less than 5 years of experience, 29%, and those with more than 10 years of experience on 25% of the sample. This distribution shows a combined percentage of 71% of participants with more than 5 years of experience, contributing to the study verisimilitude.

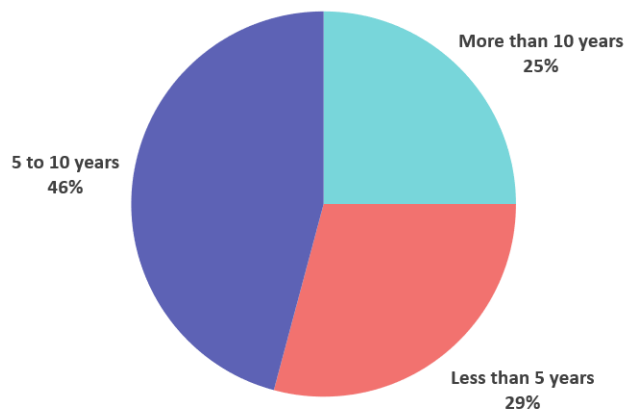


Figure 5: Participants experience

#### 4.2 Companies developing methodology

This study had no restriction about companies development methodologies, as presented in [33], but the obtained result for the question that aimed to identify if the company uses agile methodologies on their development process was that 100% of the participants' companies make use of agile methodologies, this result allows a temporal comparison with the above-mentioned article, as it was published in 2019.

The following question asked which methodologies were used by the participants in their work environment, originating the distribution presented in Figure 6. It shows the usage of Scrum only by half of the participants, followed by the combined usage of Scrum and Kanban for 46% of the sample, and lastly the usage of Kanban only by 4% of the participants. Looking for these results in terms of the methodology used, 96% of the participants claimed to use Scrum and 50% claimed to use Kanban in their processes.

#### 4.3 UML modeling experience

Taking a step forward to the main objective of the study, the first question that approaches the UML theme sought to understand the participant's UML modeling familiarity. The participants were asked to classify themselves as a person with no experience with UML modeling, basic experience, moderate experience, or advanced experience. The obtained results of this question can be observed in Figure 7, which shows a distribution of 54% of the participants that classify themselves as a person with basic experience, followed

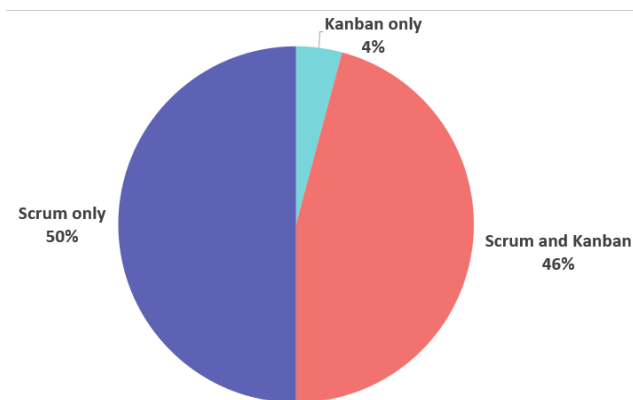


Figure 6: Company's agile methodologies usage.

by 25% with moderate experience, 13% with no experience, and 8% with advanced experience.

Grouping the results on people with and without UML modeling experience, the amount of participants that have had experience with UML is 87% of the sample, and, in this 87%, a third affirm to have moderate or advanced experience, contributing to the proposed discussion and bringing better-grounded answers.

#### 4.4 Knowledge of UML diagrams

In this question, a list of the most known UML diagrams was presented to the participants and they checked all the options of UML diagrams that they learned. The results are shown in Figure 8. The most known diagrams are: use case Diagram and class diagram, both with 96% knowledge by the participants, followed by the Object diagram and the Sequence diagram, both with 58% of knowledge. The lesser-known diagrams are Time diagram (4%) and composite structure diagram (8%).

#### 4.5 UML diagrams usage

In the last quantitative question, similarly to the previous question, the participant was invited to indicate, among the listed diagrams, which ones are used in the company where he works. Figure 9 is shown, for each UML diagram, the number of companies that use the diagram, and the number of participants that know each diagram. It was considered that the company uses a diagram if one of its employees signaled to use it in the company.

In Figure 10, it is possible to observe the diagrams with the highest degree of use in companies: diagram of use cases, used by 70% of companies and 42% of participants; class diagram, used in 70% of companies and reported by 42% of participants; activity diagram, with 40% of companies; object, sequence, communication, components, and implementation diagrams, all with presence in 30% of companies. The diagrams with the least use are the composite structure, which was not mentioned by any respondent, and that of time, mentioned by 4% of the participants.

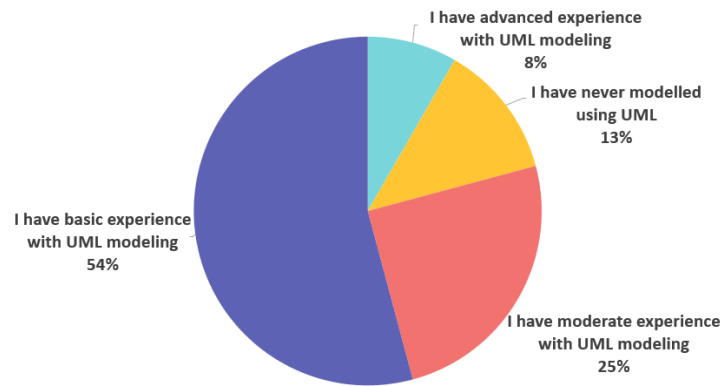


Figure 7: UML modeling participants experience.

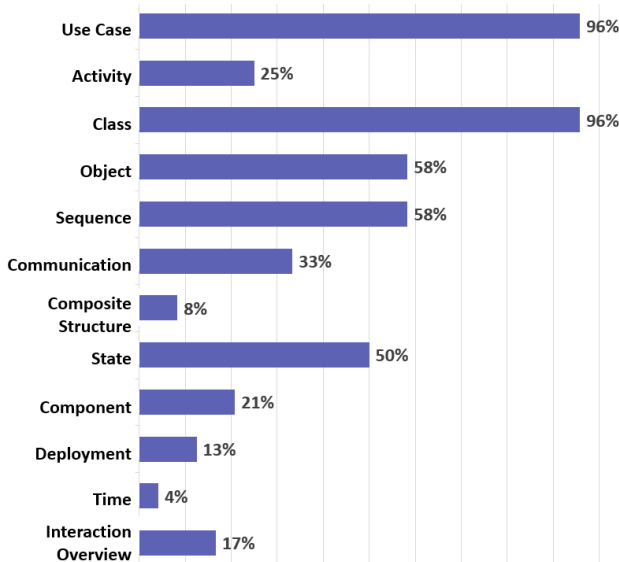


Figure 8: Participants' knowledge in the UML diagrams.

#### 4.6 Motivation for the use or disuse of UML

In this question, the participants answered the possible motivations for using or not using UML. The opinions presented were both positive and negative, showing that the participants sincerely expressed themselves without bias. Concerning the positioning of the participants in this answer, 37.5% of the participants were obtained exposing positive points of the use of UML diagrams, 42% exposing negative points and the rest either remained neutral in the answers or did not know how to answer. It is not possible, therefore, to find a polarization in the participants' perception regarding the use of UML in the context of development.

Among the reports that brought motivations for the use of UML, the most recurring arguments were based on the ease of understanding and the way it impacts on further development. An example of a report that brought motivations in favor of use can be seen in this excerpt from the answer of Respondent 18: "The use is made

when something is very complex and or when a certain piece of software or software will be shared by different teams so that the other team has an easier time understanding the methodology of those who developed it", which also highlights the use for sharing ideas with other stakeholders of the project, a fact that can also be seen in the answer of Respondent 7: "I believe we use it because it provides a better understanding of the software to be developed and also helps in understanding the stakeholders in the proposed system".

On the other hand, in the reports that focused more on points for not using UML, the most present argument was related to the demand for time and human material for the elaboration, as can be seen in the reports: "The use of UML demands time and organization of the team, as well as code documentation. The need and benefits are known in the future, but the constant demands often end up generating simpler documents and with few details." (Respondent 24) and "Projects (especially agile) are very susceptible to change. Having to remodel all the diagrams made for the project is very costly. Despite being ideal, in most software projects there are not enough resources to take this cost into account" (Respondent 13). Another negative point present in this last report and common among the responses refers to the mutability of projects within the agile methodology, as can also be identified in the response of Respondent 12: "Today the UML diagram is not widely used in the *Company U* nor in other companies I worked for, as things change constantly and we are unable to keep the documentation up to date".

#### 4.7 Other documentation resources for modeling

In Question 19, it was asked, in case the UML was not used or used parts, which other resources are commonly used for documentation in software modeling. The vast majority of participants did not mention any other resource used or stated that they did not use any specific software or technique. Among the few answers that brought other resources, it is worth mentioning the use of Google Draw, mentioned in two answers, MER, Zeplin, Postman, and Jira, the latter mentioned in one answer each. As an example of the

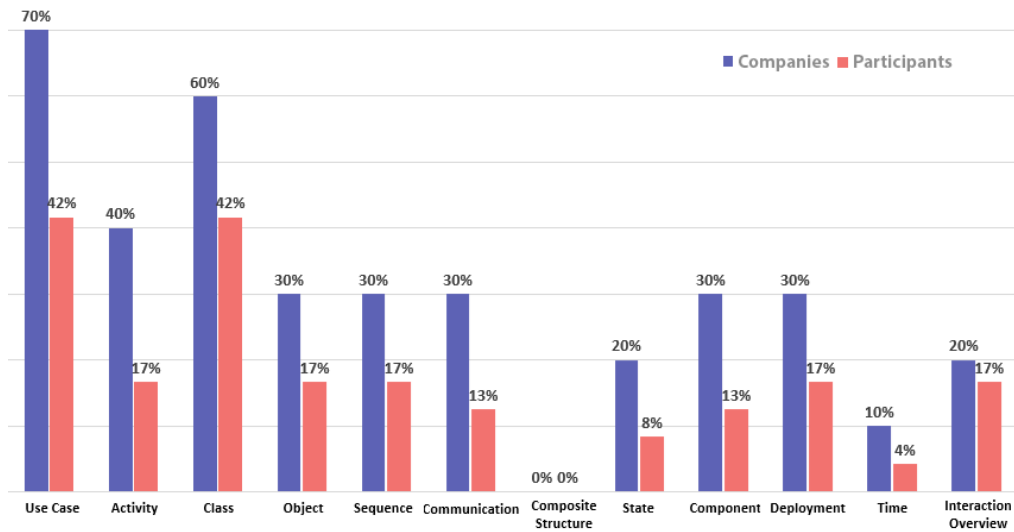


Figure 9: Comparison between knowledge and use of UML diagrams.

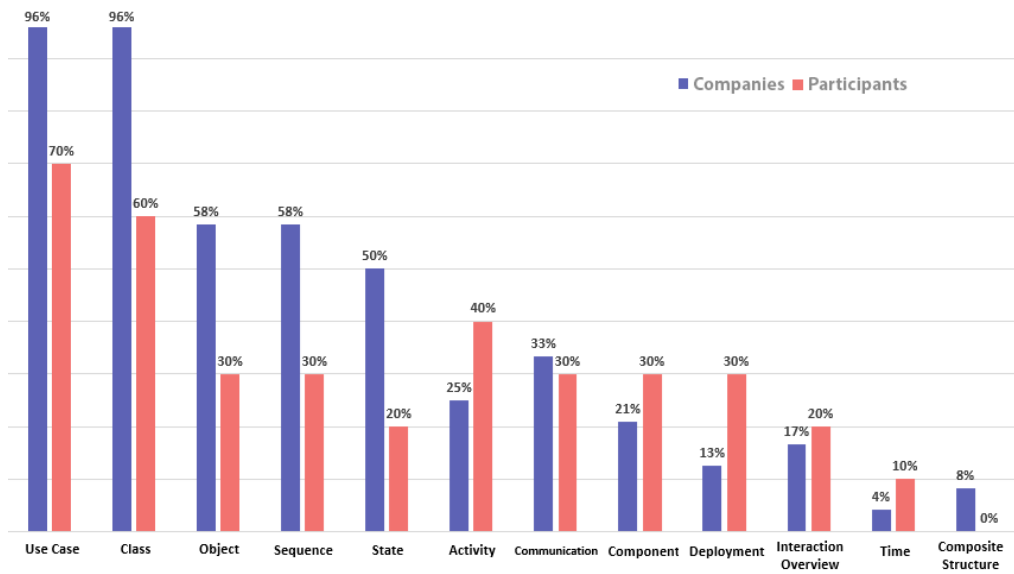


Figure 10: Use of UML diagrams by companies and participants.

majority response, which demonstrates the non-use of any other modeling resource, we have: "The modeling is done in communication with the dev team if necessary momentary diagrams are made on whiteboards only to help understand certain logical. The vast majority of the documentation is made in the code itself and, if necessary, text documents are made explaining critical points of the project" (Respondent 20).

#### 4.8 Difficulties in using UML

Starting with the survey of points that prevent UML from being used, it was asked what were the difficulties seen by the participant, as a professional in the field, in the use of UML. The terms with

the highest incidence among the responses were *diagrams* and *time*, among which the word time stands out as a difficulty, cited in 37.5% of the responses. Other difficulties proved to be recurrent in the set of responses, it is worth highlighting the *confusion* with the diversity of existing diagrams, explained by Respondent 14: "For me, it ends up being a little confusing to use UML due to the wide variety of different diagrams. When I studied the subject at the university, these diagrams were always presented in the same visual form of blocks, with few differences that allowed to differentiate each one, and this made it difficult to clearly understand what each diagram was for and what the process was to build them", as well as the complexity and detail required for the diagrams found in the



response of Respondent 21 in "The notation is very complex many times. There are many details. Which makes it very expensive to make diagrams with precision. Also, by using SCRUM, we often do not have the scope of the project closed at the beginning of it. Drawing diagrams accurately becomes impractical".

#### 4.9 Suggestions for improving UML teaching

As a final question, the respondent was asked to express his/her opinion about the teaching of UML and what improvements he/she suggested. Among the suggestions presented, the theme with the greatest recurrence was that of a more credible approach to the professional context, bringing situations that are more similar to what is seen when working in the field. Examples of answers that bring this theme: "The modeling problems usually arrive with a ready-made text, with the objective of extracting the relationships and behaviors of the application. In the real world, it is rare for such a text to exist, and when it exists it is not so precise. It would be interesting to study requirements analysis techniques with a focus on assembling the diagrams and obtaining this information from the client in the best possible way" (Respondent 9) and "Exercising the student to do two things: 1 - interpret the diagram to know what it means, and 2 - explain this diagram in non-technical language to include the other areas of the company in the development process" (Respondent 2).

#### 4.10 Knowledge of diagrams by educational institution

Stratifying the result of question 16 among educational institutions, it is possible to outline which of the diagrams are most emphasized in the academic environment and which are least remembered. It is possible to observe this stratification in Figure 11, where it is possible to identify similar profiles between institutions for use case, class, and communication diagrams. The diagrams that most diverge between institutions are: sequence diagram, for which 83% of respondents trained at UFSCar said they were aware of them, compared to 50% for USP and IFSP; object diagram, remembered by close portions between UFSCar and IFSP participants (67% and 64%), diverging from the USP portion (25%); state diagram, recognized by all USP respondents and less than half of UFSCar and IFSP respondents (33% and 43%, respectively); component diagram, remembered only by 36% of the IFSP representatives.

#### 4.11 Number of diagrams used by company

A rearrangement of the results obtained in question 7 allows an understanding of the number of diagrams used in each of the companies in the region, as shown in Figure 12. The companies that make the most use of UML diagrams in the development context are Companies C and G, which make use of 9 of the 12 diagrams listed, followed by Company B with 8 diagrams. The companies identified as those that make the least use were Companies E and J, which do not use any diagram. This analysis presents a considerable disparity between companies in the region, even though all of them use agile methodologies, in 3 out of 10 companies there is the use of at least 8 out of 12 diagrams and, in contrast, 4 out of 10 companies do not use more than one of the 12 diagrams. This shows that it is

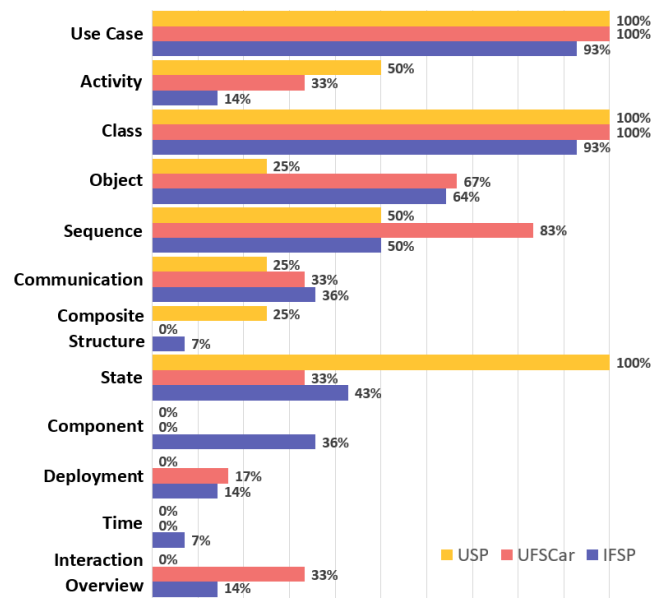


Figure 11: Knowledge of participants by type of UML diagram rearranged between educational institutions.

not possible to draw a standard on the number of diagrams used in companies.

#### 4.12 Comparison between knowledge and use of UML diagrams

One of the main discussions in this paper focuses on understanding the correspondence of what is discussed in the academic environment about UML and its diagrams, with what is used in the professional environment. A perspective that supports this discussion is the comparison between the results of questions sixteen and seventeen, knowledge compared to use. The results of this perspective are presented in Figure 10.

Analyzing Figure 10, it is possible to verify a parity between the two diagrams with greater knowledge and the two with the greatest share of use in companies, the use case, and class diagrams. On the other hand, it is possible to notice a disparity in the case of the activity diagram, which is known by a quarter of the participants and used in 40% of companies in the region. Analyzing quantitatively the use of diagrams in companies in the region, the most used scenario is the diagrams: use case, with 70%; class, with 60%; activity, with 40%; object, sequence, communication, component, and implementation, with 30% each. Analyzing, then, the same scenario for science: use cases and class, with 96%; object and sequence, with 58%; and state, with 50%. It is possible to identify an opportunity to relocate part of the time lost with the object, sequence, and state diagrams for the activity diagram, which is in the third place regarding the use and is located in the half with less science between the diagrams.

### 5 DISCUSSION OF RESULTS

This section discusses the results previously presented. The first relevant discovery alludes to the personal profile of the research



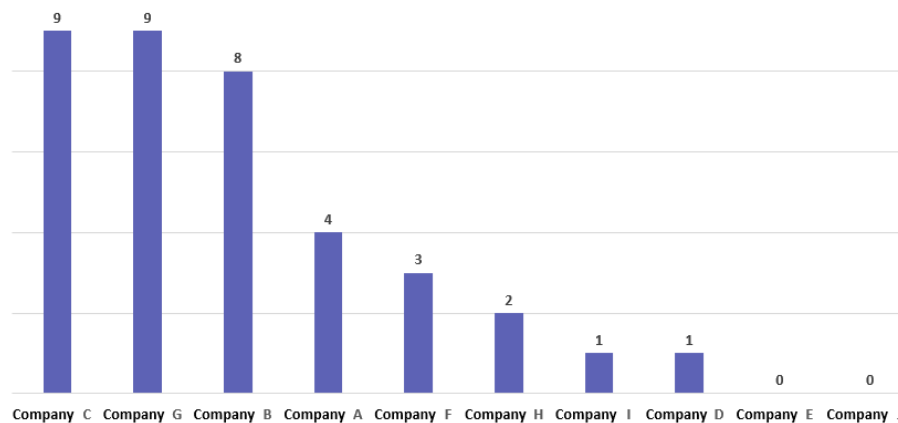


Figure 12: Number of UML diagrams used by company.

participants, who represent the total population of employees of the development companies in the region, in terms of educational institution: 100 % of the participants had the training institutions of the city of São Carlos - SP: USP, UFSCar, and IFSP, which shows a scenario of high absorption of students trained in the region by companies in the sector, as well as low demand for human material from other regions.

This paper did not aim to exclusively analyze companies that use agile methodologies, however, as a result of question 12, which asked whether the participant's company used agile methodology, 100% of the participants and, consequently, 100% companies use this methodology. This information corroborates one of the purposes of this paper: to bring inputs about the region's professional scenario for improvements in academic education, aiming to train professionals better prepared for the regional market. With the information that all companies use agile methodology, it is important to highlight this issue in the classroom.

Concerning the use of UML in the professional environment, the results obtained in this paper show a divergent scenario from what had been reported in similar works in different periods and regions. As shown in the motivation of this paper, the results found for the use of UML were: 30% of participants in [21], 26.8 % in [8] and 28.6% in [33]. This scenario showed a trend of disuse in the UML, a scenario that is not repeated in the work: 66% of the participants affirmed to use at least one of the diagrams in one or more stages of the software process. Analyzing as to the companies that use it, this number grows to 80% of the companies, in which at least one employee claimed to use at least one diagram in the process.

A possible assumption to justify this relatively high number of UML utilization, when compared with the results of similar works, also refers to the fact explained in the second discovery mentioned in this section: 100% of the participants were or are being trained in the institutions of the city of São Carlos - SP. Better training brings greater control over the tools presented, allowing them to be used in professional environments. This fact can be evidenced by the high science results on the part of certain diagrams, showing that this knowledge was fixed. In its turn, analyzing Figure 9, it can be seen that there was no use by any participant of the composite

structure diagram, thus being interesting to analyze the importance of studying it during teaching it, being able to dedicate more time to other commonly used diagrams.

Comparing the graphs of Figures 11 and 9, some opportunities can be identified, as in the case of the state diagram for the X institution, which presents knowledge for this similar diagram to the one presented for the *use case* and *class* diagrams, which present the greatest uses, a fact that is not repeated for the *state* diagram, used by 8% of the respondents. In this way, it would be possible to reallocate part of the efforts to needy diagrams with greater focus, as is the case with the *component* and *implementation* diagrams that are used in 30% of companies, however, none of the professionals from the institution indicated knowing.

Unlike related studies, there is considerable use of UML in São Carlos - SP IT companies. With this result, we suggest that the UML content should be preserved in the curriculum of educational institutions, of course in an updated and optimized way to meet the trends presented by IT companies. The opportunities in the modeling area, with the mastery of agile methodologies and the trend of continuous acceleration of processes, are vast. One of them would be, in principle, the adequacy of UML modeling for agile methodologies, relieving the most valued asset in these methodologies: time.

## 6 THREATS TO VALIDITY

We recognize the following external and internal threats that could have affected the validity of our results.

A possible threat to external validity is related to data collection carried out online through an electronic questionnaire. Besides that, the choice of participants was through virtual contact, which could increase doubts about the questionnaire. To minimize this risk, we sought to collect at least 30 responses, however, we obtained only 27. Another possible threat to external validity refers to the generalization of the results because our answers were collected in full from the respondents. Therefore, we cannot generalize the results of this study, even though the evidence indicates that IT companies in São Carlos - SP use UML in their development processes, further studies are needed to mitigate this threat.

The threats to internal validity include the construction of the questionnaire used, to mitigate the threat, a pilot study was applied to correct and perfect the questionnaire in order to minimize the doubts that the participant might encounter. Another threat to internal validity is in the developed process used to carry out the study, to mitigate this threat, we inspired on empirical studies that carried out the exploration with survey [20, 25] and we base our exploratory process on the study [26] that defines guidelines for the development of empirical studies.

## 7 CONCLUSION

The main objective of this study was to identify how UML was being adopted by IT companies and, thus, improve the teaching of UML in educational institutions in São Carlos - SP. As a result, we identified the use of UML mainly in the context of agile methods. Based on the quantitative and qualitative analysis carried out in the responses achieved by the survey, it was possible to identify ways to improve the teaching of UML based on the perspectives of software engineers from São Carlos - SP IT companies.

Due to the lack of previous similar studies in the studied region, it is not possible to measure whether UML usage has grown or has decreased. It would be interesting to apply this survey again at different times of time and, therefore, measure whether the utilization rate of UML has changed or remained.

Finally, it might be interesting to replicate this study by collecting data from recently graduated participants and software development areas, professionals. In this way, it would be possible to create an overview of the knowledge obtained from the graduation, without other experiences interference, and to compare with data on the use of UML in IT companies.

## ACKNOWLEDGMENTS

The authors acknowledge the São Paulo Research Funding, FAPESP, for the financial support under process no. 2018/25744-6.

## REFERENCES

- [1] Andrea Arcuri and Lionel Briand. 2014. A hitchhiker's guide to statistical tests for assessing randomized algorithms in software engineering. *Software Testing, Verification and Reliability* 24, 3 (2014), 219–250.
- [2] Darlan Arruda and José Filho. 2014. Software Metrics: A Survey Conducted within Brazilian IT companies [PT]. <https://doi.org/10.5748/9788599693100-11CONTECSI/PS-739>
- [3] Loli Burgueño, Antonio Vallecillo, and Martin Gogolla. 2018. Teaching UML and OCL models and their validation to software engineering students: an experience report. *Computer Science Education* 28, 1 (2018), 23–41. <https://doi.org/10.1080/08993408.2018.1462000>
- [4] Peter J. Clarke and Alfonso Pierantonio. 2018. Teaching modeling: a software perspective. *Computer Science Education* 28, 1 (2018), 1–4. <https://doi.org/10.1080/08993408.2018.1486535>
- [5] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- [6] N. M. Devadiga. 2017. Software Engineering Education: Converging with the Startup Industry. In *2017 IEEE 30th Conference on Software Engineering Education and Training (CSEET)*. 192–196.
- [7] Brian Dobing and Jeffrey Parsons. 2006. How UML is used. *Commun. ACM* 49, 5 (2006), 109–113.
- [8] Kleimner Farias, Lucian Gonçalves, Vinicius Bischoff, Bruno Carreiro da Silva, Everton T Guimarães, and Jacob Nogle. 2018. On the UML use in the Brazilian industry: A state of the practice survey (S).. In *SEKE*. 372–371.
- [9] A. M. Fernández-Sáez, D. Caivano, M. Genero, and M. R. V. Chaudron. 2015. On the use of UML documentation in software maintenance: Results from a survey in industry. In *2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS)*. 292–301.
- [10] Arlene Fink. 2003. *The survey handbook*. Sage.
- [11] Fernanda Chacon Fontoura. 2019. *Uso de Metodologias de Desenvolvimento de Software e de Engenharia de Requisitos em empresas de Tecnologia: um estudo a partir de um Survey*. B.S. thesis. Universidade Federal do Rio Grande do Norte.
- [12] Martin Fowler. 2003. *UML Distilled: A Brief Guide to the Standard Object Modeling Language* (3 ed.). Addison-Wesley Professional.
- [13] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional.
- [14] A. N. Ghazi, K. Petersen, S. S. V. R. Reddy, and H. Nekkanti. 2019. Survey Research in Software Engineering: Problems and Mitigation Strategies. *IEEE Access* 7 (2019), 24703–24718.
- [15] GoogleForms. 2020. Google Forms. [https://docs.google.com/forms/d/e/1FAIpQLSdnt81ZyJKXXrgrYXR1WS9QF\\_ib-JJJg5q9y\\_LYtzUQ7M0fLw/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSdnt81ZyJKXXrgrYXR1WS9QF_ib-JJJg5q9y_LYtzUQ7M0fLw/viewform?usp=sf_link)
- [16] Barbara A. Kitchenham and Shari L. Pfleeger. 2008. *Personal Opinion Surveys*. Springer London, London, 63–92. [https://doi.org/10.1007/978-1-84800-044-5\\_3](https://doi.org/10.1007/978-1-84800-044-5_3)
- [17] Johan Linåker, Sardar Sulaman, Martin Höst, and Rafael de Mello. 2015. *Guidelines for Conducting Surveys in Software Engineering*. Technical Report 1.1. Lund University, Sweden.
- [18] Adriana Lopes, Igor Steinmacher, and Tayana Conte. 2019. UML Acceptance: Analyzing the Students' Perception of UML Diagrams. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*. 264–272.
- [19] Ariadi Nugroho and Michel RV Chaudron. 2008. A survey into the rigor of UML use and its perceived impact on quality and productivity. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. 90–99.
- [20] Leo Natan Paschoal and Simone do Rocio Senger de Souza. 2018. A Survey on Software Testing Education in Brazil. In *Proceedings of the 17th Brazilian Symposium on Software Quality (Curitiba, Brazil) (SBQS)*. Association for Computing Machinery, New York, NY, USA, 334–343. <https://doi.org/10.1145/3275245.3275289>
- [21] M. Petre. 2013. UML in practice. In *2013 35th International Conference on Software Engineering (ICSE)*. 722–731.
- [22] Marian Petre. 2013. UML in practice. In *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, 722–731.
- [23] Marian Petre. 2013. UML in practice. In *2013 35th International Conference on Software Engineering (ICSE)*. IEEE, 722–731.
- [24] Gianna Reggio, Maurizio Leotta, and Filippo Ricca. 2014. Who knows/uses what of the UML: A personal opinion survey. In *International Conference on Model Driven Engineering Languages and Systems*. Springer, 149–165.
- [25] Luiz A. L. Rodrigues and Jacques D. Brancher. 2019. Playing an educational game featuring procedural content generation: which attributes impact players' curiosity?
- [26] Forrest Shull, Jeff Carver, A. Bldg, and Guilherme Travassos. 2001. An Empirical Methodology for Introducing Software Processes. *ACM SIGSOFT Software Engineering Notes* 26 (07 2001). <https://doi.org/10.1145/503271.503248>
- [27] Keng Siau and Poi-Peng Loo. 2006. Identifying difficulties in learning UML. *IS Management* 23 (06 2006), 43–51. <https://doi.org/10.1201/1078.10580530/46108.23.3.20060601/93706.5>
- [28] I. Sommerville. 2019. *Engineering Software Products: An Introduction to Modern Software Engineering*. Pearson.
- [29] I. Steinmacher, A. P. Chaves, T. U. Conte, and M. A. Gerosa. 2014. Preliminary Empirical Identification of Barriers Faced by Newcomers to Open Source Software Projects. In *2014 Brazilian Symposium on Software Engineering*. 51–60.
- [30] Bedir Tekinerdogan. 2011. Experiences in teaching a graduate course on model-driven software development. *Computer Science Education* 21, 4 (2011), 363–387. <https://doi.org/10.1080/08993408.2011.630129>
- [31] Andreas Wortmann, Olivier Barais, Benoit Combemale, and Manuel Wimmer. 2020. Modeling languages in Industry 4.0: an extended systematic mapping study. *Software and Systems Modeling* 19, 1 (Jan. 2020), 67–94. <https://doi.org/10.1007/s10270-019-00757-6>
- [32] Alan Xavier, Fábio Martins, Ricardo Pimentel, and Denis Carvalho. 2019. Aplicação da UML no contexto das metodologias ágeis. In *Anais do VI Encontro Nacional de Computação dos Institutos Federais*. SBC.
- [33] Alan Xavier, Fábio Martins, Ricardo Pimentel, and Denis Carvalho. 2019. Aplicação da UML no contexto das metodologias ágeis. In *Anais do VI Encontro Nacional de Computação dos Institutos Federais (Belém)*. SBC, Porto Alegre, RS, Brasil. <https://doi.org/10.5753/encompif.2019.6353>