Adopting Behavior-Driven Development (BDD) in software development: a multivocal review

Shexmo Richarlison Ribeiro dos Santos[1][0000−0003−0287−8055], Guillermo Rodriguez[2][1111−2222−3333−4444], and Fabio Gomes Rocha[3][0000−0002−0512−5406]

[1] Federal University of Sergipe, São Cristóvão SE 49100-000, BRAZIL
shexmor@gmail.com
[2] gomesrocha@gmail.com
[3] ISISTAN Research Institute, Tandil Bs. As. B7000, ARGENTINA
exarodriguez@gmail.com

Abstract. Context: In Software Development, it is essential to use frameworks that can help achieve the goal regarding the behavior expected by the system. Problem: Identify the point of view of researchers and professionals regarding the Behavior-Driven Development (BDD) agile framework. Solution: Characterize the BDD framework through a Multivocal Literature Review (MLR). Method: We performed an MLR in the central databases of white and gray literature. Summarization of Results: We identified the main aspects related to the adoption of BDD to characterize its adoption. Contributions and Impact: Through this study, it was possible to advance the understanding of the adoption of BDD to contribute to the scientific community and industry on the aspects inherent to the application of this framework.

Keywords: Behavior-Driven Development (BDD) · Requirement · Software development · Software Engineering · Systematic review.

1 Introduction

From 2001 onwards, with the Agile Movement, the software development process has suffered significant changes [3]. The Behavior-Driven Development (BDD) framework has recently been discussed and addressed in the Software Engineering (SE) field [15, 16]. The BDD process works through harmonizing requirements elicited by the product owner, developer, and tester. This part of the process is called "3 amigos," it happens at the beginning of the project with the primary goal of making all the parts aware of the elicitation. Therefore, BDD aims to bring out a more assertive process in delivering the final product to the customer, focusing on improving communication and readability in the process [19].

According to Shaw [26] regarding finding better ways for software development and evaluations, the goal of this study was **to analyze** studies on BDD, **with the purpose to** characterize, **respect to** its adoption and application, **from the point of view of** researchers and professionals, **in the context**

of theoretical and applied researches from 2013 to 2023. To achieve the proposed goal, we carried out a Multivocal Literature Review (MLR) aiming to contribute to the scientific community and industry with the findings in white and gray literature related to BDD.

This article shows as follows: **Related work**, where we address related studies; **Behavior-Driven Development**, where we show the importance of understanding BDD based on authors in the area; **Multivocal Literature Review (MLR)**, where we present the methodology used; **Results**, where we present the data obtained; **Discuss**; where we discuss advances in understanding BDD and correlate them with related work; **Threats to validity**, where we present the threats identified throughout the study, as well as how we mitigate them; and, **Conclusion**, where we sum up our contributions obtained through this study and bring suggestions for future works.

2 Related work

Farooq et al. [8] carried out a Systematic Literature Review (SLR), which brought aspects focused on the state-of-the-art BDD to contribute by identifying the benefits of using BDD for the software development process. We also identified the SLR of Arnyndiasari, Ferdiana, and Santosa [1], who investigated software practices for agile developers. In this way, the importance of understanding BDD for software development through an RML could contribute to a greater understanding of the adoption of this framework.

Garousi and Zhi [12], through a survey in Canada, brought as some of their results the fact that BDD is being applied as a testing tool by the industry to highlight the role of testing with more recent methods in software development, pointing out the importance of characterizing the BDD. Furthermore, BDD is a relatively new testing strategy [30], which uses English to describe feature codes. Besides, because it has a structured language, the BDD can deal with the requirements objectively and directly, solving cultural situations, for example, [25]. A better understanding of BDD will contribute to quality software development, reducing rework time and costs.

Recently, BDD has been discussed in academia and applied in industry [28]. According to Arnyndiasari, Ferdiana, and Santosa [1], software development can be successful through agile practices. There are greater possibilities of success when using methods such as BDD. As we can see, characterizing BDD can contribute to better use of this framework and improve the quality of software delivery.

Through our study, we will be able to characterize BDD to present the state-of-the-art of this framework and better understand it from the point of view of researchers and professionals.

3 Behavior-Driven Development (BDD)

BDD is an offshoot of the Test-Driven Development (TDD) traditional process [14], so where the latter has the test as its primary focus, the former is focused on the behavior expected by the system. BDD was created from the problems found in TDD, being a way to mitigate such situations, according to North [19]. Because it targets behavior, BDD brings greater interactivity between teams and stakeholders, improving communication between all parties involved [5].

BDD uses a ubiquitous language (Gherkin) that works through commands written in daily English that provides a better understanding for all those involved in the process [19], in addition to being able to significantly improve the results aimed at writing tests [29, 7].

Some investigations show benefits to adopting BDD, such as improved communication, collaboration, and living documentation [22, 27, 18]. Living documentation refers to how code can be updated quickly. The code must be updated since it is part of the system documentation.

Moe [14] and Binamungu, Embury, and Konstantinou [4] highlighted some of the purposes of adopting BDD, such as validating the product delivered to match what was requested by the customer and the attention given to data handling and validation, respectively. Therefore, another relevant aspect related to BDD is the richness of detail and focus when eliciting requirements [13].

4 Multivocal Literature Review (MLR)

Multivocal Literature Review (MLR) is a type of Systematic Literature Review that seeks to cross-reference information found in white and gray literature [11]. White literature comprehends articles published on traditional scientific bases, while gray literature is the open Internet, such as websites and blogs. We used the guidelines of Garousi, Felderer, and Mantyla [10] adopting 2nd tier for gray literature. According to these authors, there are three tiers of gray literature, as shown in Table 1.

Table 1. Tiers of Gray Literature

Tiers	Examples
1st (High credibility)	Books, magazines
2nd (Moderate credibility)	Annual reports, sites
3rd (Low credibility)	Blogs, emails, tweets

According to Rodriguez [24], one of the benefits of an MLR is improving the understanding of the topic addressed in terms of researchers and practitioners through the gain obtained by gray literature. Furthermore, it is essential to emphasize that information from gray literature is discarded when only it uses white literature. [9].

We carried out an MLR regarding BDD, showing its relevance in software development because BDD has recently gained more notoriety, becoming a new and stimulating topic to learn.

To carry out the MLR, we used the Parsifal tool[4], which aims to structure the entire research, from the definition of the title to the analysis of the results collected. Thereby, we carried out this investigation from the Goal-Question-Metric approach according to Basili [2], where we have: **To analyze** studies on BDD, **with the purpose to** characterize, **respect to** its adoption and application, **from the point of view of** researchers and professionals, **in the context of** theoretical and applied researches from 2013 to 2023.

As justification for the cut from 2013 to 2023, we used as a starting point the book "Cucumber Recipes: Automate Anything with BDD Tools and Techniques" [6] since Cucumber is one of the most important frameworks for BDD. Also, it is worth mentioning Dan North's article "JBehave. A framework for Behavior-Driven Development (BDD)" [20] BDD's creator. Through these frameworks, it was possible to advance with BDD. In addition, in 2014, the company Sucumbe was created. This company may also have impacted the development of BDD.

Based on such information as Pai et al. [21], we used the PICOC criteria to define the terms: **Population**: BDD; **Intervention**: Requirements, Test; **Outcome**: Tools; **Context**: Software development. We discarded the **C**omparison criteria because we have no intent related to this term. According to the presented keywords and their synonyms, we created the following string to use in the selected databases:

("**BDD**" OR "**Behavior-Driven Development**") AND ("**requirements**" OR "**documentation**" OR "**quality**" OR "**test**") AND ("**tools**" OR "**methods**" OR "**models**" OR "**techniques**" OR "**software development**")

We used the search string in 8 databases, 5 of which were white literature and 3 of gray literature, with the period between 2013 and 2023 as a cut.

Studies identification

The chosen white literature databases were: ACM Digital Library[5], IEEEXplore Digital Library[6], Web of Science[7] and Scopus[8]. Because they are the most relevant bases for the computing field, we selected them as the path to follow in this investigation. In addition, the science@direct[9] base was initially also selected. However, when we entered the string we defined, the base did not return any articles, justifying that the string was too large, so we deleted it.

We chose the bases for the gray literature regarding the 2nd tier as mentioned in Table 1: Google Scholar[10] (top five tabs) because it is a website with an academic purpose, InfoQ[11] (top five tabs) which is a portal with a variety of

[4] https://parsif.al/
[5] https://dl.acm.org/
[6] https://ieeexplore.ieee.org
[7] https://www.isiknowledge.com
[8] https://www.scopus.com
[9] https://www.sciencedirect.com/
[10] https://scholar.google.com/
[11] https://www.infoq.com/

content related to SE, and Dan North[12] who is the BDD's creator. We defined the top five tabs of Google Scholar and InfoQ to investigate the main results with the developed string. Also, Dan North's website was selected because BDD is the main focus of this investigation, where we investigated in the tab "**tag**" subject "**bdd**" found on the website.

As previously mentioned, the MLR crosses the data found in the white and gray literature to observe what has been researched by the academy and published by relevant authors regarding BDD on a non-scientific basis. In Table 2, we can identify the number of articles that returned from each base when we performed with our search string.

Table 2. Number of articles

Base	Quantity
ACM Digital Library	800
IEEEXplore Digital Library	100
Web of Science	293
Scopus	603
Google Scholar	50
InfoQ	50
Dan North	4
Total	1900

We continue in the following Subsection with the definition of the research questions.

Definition of research questions

As a guide for this study, we used the research questions: RQ1 - What type of results? RQ2 - What type of validation? RQ3 - What type of methodology? RQ4 - In which application domains is BDD most used? RQ5 - In which context is BDD most used? RQ6 - Are benefits of adopting BDD presented? Which ones? RQ7 - Are harms of adopting BDD presented? Which ones? RQ8 - For the adoption of BDD, which are tools most used? These research questions will help us achieve the goal proposed for this study. In Table 3 we present the inclusion and exclusion criteria we used to screen the articles.

Using the criteria presented in Table 3, it was possible to screen the articles as per the following Subsection.

Selection of the articles

Three researchers participated in the selection of articles. One researcher initially carried out the screening, while the other two researchers carried out the validation. During the validation step, there was no need for reconciliation among researchers; we all agreed on the same points.

Initially, we used the automatic search of the Parsifal tool to find duplicate articles. Soon after, we read the titles and abstracts to verify if the paper re-

[12] https://dannorth.net/

Table 3. Criteria

Inclusion	The article presents the factors related to BDD
Exclusion	Duplicate articles
	Articles published only as journal abstracts or forewords
	Articles that are not in English
	Articles that do not partially answer at least one of the research questions
	Articles that do not deal with BDD Behavior-Driven Development

sponded, even partially, to at least one of the research questions. Finally, we read the remaining articles thoroughly. We present the numbers of the screening steps in Figure 1.

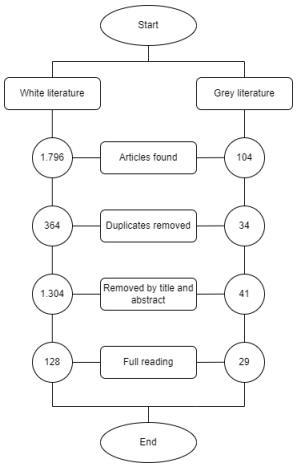

Fig. 1. Selection of the articles

As we can see, we selected 157 articles that formed this research. We selected 18.47% articles from gray literature and 81.53% from white literature. We extracted and grouped the data collected in the graphics presented in the following Section to answer the research questions.

5 Results

We use the definitions mentioned by Shaw [26], we could evaluate each type of result according to the definition provided in her article. We chose her article to guide some of our research questions as she addresses the necessary components

for exemplary research in computing and is a reference in the field. The answers to each of the research questions of this study are below.

RQ1 - What type of results?

According to Figure 2, the type **Report** was the primary type obtained with 90 articles, followed by **Procedure or Technique** with 39, showing cases where it uses BDD. Moreover, the type **Qualitative or descriptive model** also presented a considerable number of 30 articles, followed by **Tool or notation** with 21. The other types not mentioned had a lower quantity than those presented, showing possible types to address in new investigations.

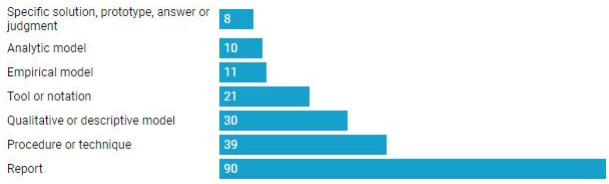

Fig. 2. RQ1 - What type of results?

RQ2 - What type of validation?

Figure 3 answers the second research question. We obtained the types of validation of the selected articles, with the primary type being **Analysis** mentioned in 148 articles, followed by **Evaluation** with 98. We obtained 80 results for **Example** to infer a type of validation widely used when dealing with BDD. In addition, we found 26 articles with **Experience** validation, so it is possible to carry out even more investigations with this type of validation in future works.

Fig. 3. RQ2 - What type of validation?

RQ3 - What type of methodology?

Figure 4 we identify the methodologies used in the articles, answering the third research question. We noticed that the **Descriptive** methodology stood out, being mentioned in 80 articles, denoting the need for a good description of how to carry out an investigation, followed by the **Case Study** with 55 and

Experiment with 28, so that we could observe that case studies, while carried out with a more abstract perception, are less solid than experiments, which aim at more assertive data pointed at their realization.

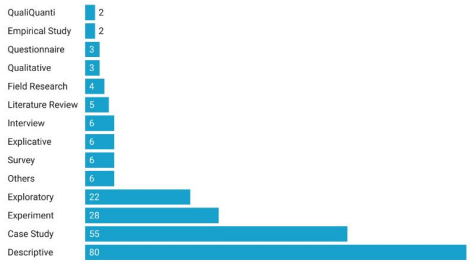

Fig. 4. RQ3 - What type of methodology?

We also identified 22 articles with the **Exploratory** type, which infers the importance of exploring actions aimed at BDD. Finally, it is worth mentioning that we found 6 articles for the **Survey** type of methodology so that there could be more applications of this type aimed at better understanding the behavior of those who use BDD, whether in research at the academy or everyday life in the industry.

RQ4 - In which application domains is BDD most used?

Figure 5 we identify which domains apply BDD, the fourth research question. 72 articles **Not mentioned** which domain, we can understand it as a point to improve in future investigations. In addition, we had 36 articles with the domain **Information Systems** and 29 with **Workflow**, thus denoting the main domains found. It is necessary to carry out studies in specific domains to validate the framework.

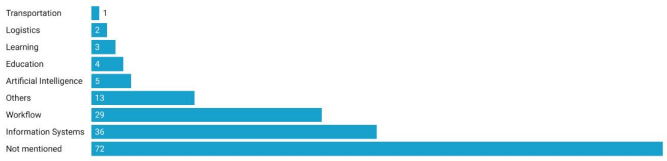

Fig. 5. RQ4 - In which application domains is BDD most used?

To better explain the use of BDD, academia, and industry would benefit from such aspects, as they would be able to identify in which domains the framework would be most effective.

RQ5 - In which context is BDD most used?

Figure 6 we identify the context where BDD applies, answering the fifth research question. 77 articles mentioned that BDD applies as **Test**, followed by **Requirement** with 43 and **Not mentioned** with 27, detonating, in the latter, the need to mention the context in which adopts BDD, aiming to contribute to clearer methodological elaborations.

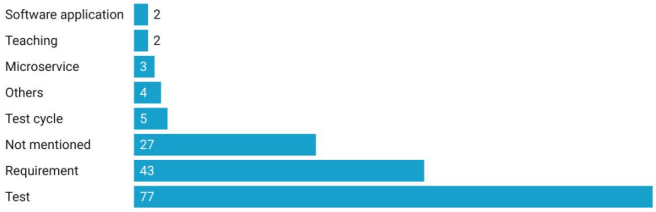

Fig. 6. RQ5 - In which context is BDD most used?

According to North [19], BDD is a tool aimed at behavioral testing of the system. The BDD framework is present in requirements elicitation and test automation throughout the software lifecycle. The context that uses BDD confirms what was intended by Dan North when he created the framework.

RQ6 - Are benefits of adopting BDD presented? Which ones?

Figure 7 we present the benefits related to the adoption of BDD, obtaining 54 mentions for **Readability**, followed by 52 for **Communication**, corroborating what North [19] mentioned when characterizing BDD, in addition to what Purkayastha et al. [23] found. Moreover, 51 articles **Not mentioned** if there are benefits related to the adoption of BDD so that there is a need to improve the characterization of BDD in the investigations carried out.

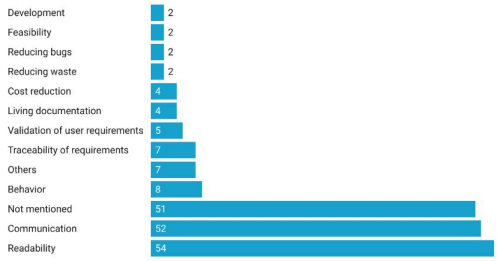

Fig. 7. RQ6 - Are benefits of adopting BDD presented? Which ones?

BDD has positive aspects of Readability and Communication because it uses the Gherkin language. Features written using BDD follow the Given-When-Then pattern, which improves everyone involved in the process's understanding, as well as the execution of more precise and assertive functionality.

RQ7 - Are harms of adopting BDD presented? Which ones?

Figure 8 presents some negative points related to the adoption of BDD. As we can see, the most significant number found were articles that **Not mentioned** harms in its adoption with 131 responses, followed by **Others** with 52, with this number relatively high due to the small mentions focused on details such as aspects related to large-scale projects or delayed return on investment applied to their adoption. We also got 6 mentions for **Cost**, initially associated with the cost of implementation concerning other tools, 3 for **Lack of experience**, and 3 for **Productivity**, so the last two are linked, where we infer that if there is no experience the probability of Productivity becomes low.

Fig. 8. RQ7 - Are harms of adopting BDD presented? Which ones?

As the most significant number found in the articles analyzed were **Not mentioned** related to negative aspects inherent to the use of BDD in 8, it can be understood from this point that the framework has a good resourcefulness in what was proposed by Dan North [19], as its use continues to be adopted by the industry.

RQ8 - For the adoption of BDD, which are tools most used?

Figure 9 we identified the main tools used for the application of BDD, resulting in the **Cucumber** tool in 71 articles, in which it is worth mentioning that this tool is the oldest, followed by 67 articles **Not mentioned**, highlighting, once again, the lack of clarity regarding the characterization of BDD. Also, we got 32 mentions for **Jbehaviour**, a tool created by Dan North.

Fig. 9. RQ8 - For the adoption of BDD, which are tools most used?

These results identified the main points related to the adoption of BDD. We, therefore, proceed with the Discussion of the results found in the next Section.

6 Discussion

We identified the main aspects related to the adoption of BDD, for example, the types of validation related to its adoption, as well as tools that support the execution of this framework, in addition to positive points aimed at its implementation. Related to Farooq et al. [8], Solis [28] and Arnyndiasari, Ferdiana, and Santosa [1], we could advance on the state-of-the-art by characterizing BDD through a Multivocal Literature Review.

We could identify Dan North's primary goal when he created BDD: the focus on communication and readability [19] being achieved within the literature through the validation found in Figure 4. We also identified a need for experimental investigations to identify cases better individually to obtain more accurate results in the studies carried out. However, due to the lack of experimental studies, results can be generalized regarding the same situation.

Regarding the negative aspects related to the adoption of BDD, it was inferred from Figure 8 that most of the articles did not mention negative aspects related to its adoption. We could infer that BDD is a good framework used by the industry, mainly when it comes to requirements elicitation and test, as shown in the Figure 6. Moreover, as mentioned by Garousi and Zhi [12], who surveyed the industry and found the BDD uses as a testing tool, besides Yang, Costa and Zou [30] and Scandaroli [25], in mentioning how BDD works in describing features, contributing to the clarity process.

Through Figure 7, we conclude the importance of people adopting BDD. Having presented the improvements using this framework, Communication, and Readability in delivering the final product will have been more assertive, helping with less cost and rework, as long as the product developed will have more quality.

Through Figure 9, we could realize the Cucumber and Jbehaviour frameworks' impact on BDD. We could identify these most used frameworks related to BDD, inferring their importance to adopting BDD better.

Besides reaffirming what Nascimento et al. [17] reported that there are more positive than negative results related to the adoption of BDD, we have advanced in studies on the subject, so that we contribute to the scientific community with the main aspects found in the investigated databases, in addition, to contribute for the industry in terms of understanding the functionality of BDD.

7 Threats to validity

According to Zhou et al., [31], the validity of the research is due to the veracity contained in the paper's information. The authors point out the following aspects as threats to validity: construct, internal, external, and conclusion.

To mitigate threats to construction validity, we made a pilot with the search string created from keywords and their respective synonyms in the databases mentioned in Table 2. We were able to refine the focus of our research to answer the questions mentioned in Section 4.

Among the threats inherent to this investigation are the risks the selection of articles can cause if carried out incorrectly. After searching the databases using the search string, we selected articles that answered at least one of the research questions. Furthermore, there could have been problems in selecting articles if there had been any disagreement among the researchers. However, we solved this threat by clarifying the inclusion and exclusion criteria in Table 3.

As we aimed to identify BDD-related aspects, we could see different results from our search string according to Section 5. We can not generalize the information in the results because each article has different approaches, for example, case studies with large and small projects. However, there is the possibility of benefits such as cost reduction in some instances and not others.

The results of this investigation are replicable following the steps mentioned in Section 4, where we describe every step from protocol creation in the Parsifal tool, passing through the selection of the articles, and ending with the selected ones.

8 Conclusion

Methodologies such as Systematic Literature Reviews are a good fit when there is a relatively new subject in which one seeks to systematize what exists published. We carried out a Multivocal Literature Review (MLR) to synthesize what has been researched and published in the white and gray literature regarding Behavior-Driven Development (BDD).

We could realize the importance of people in eliciting requirements and how much this can impact the delivery of the final product, given that if a requirement is elicited incorrectly or misunderstood, this error will impact the entire software development process.

Through the gherkin language, BDD works in a way focused on the software's behavior, which helps to achieve better-expected results. In this case, using BDD for the requirements elicitation part of the process can help all people involved understand how the software in development must behave.

By characterizing BDD through this MLR, we could contribute to a better understanding of how to use this framework by industry and, on the other hand, how researchers could advance studies related to BDD. We could realize aspects of adopting BDD as mentioned in Section 5 through our eight research questions and answers.

In future work, based on the results obtained in this investigation, we will survey teams that use the BDD to identify whether the aspects related to this framework follow the professionals who use it in their day-to-day work activities. Based on the results of this new investigation, we can carry out a case study with teams that do not use BDD and identify, in practice, whether the identified and validated aspects are replicable to other teams.

9 Acknowledgment

We would to thank the CAPES/BRAZIL master's scholarship.

References

1. Arnyndiasari, D., Ferdiana, R., Santosa, P.I.: Software practices for agile developers: A systematic literature review. In: 2022 1st International Conference on Information System & Information Technology (ICISIT). pp. 238–243. IEEE (2022)
2. Basili, V.R.: Software modeling and measurement: the goal/question/metric paradigm. Tech. rep. (1992)
3. Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., et al.: The agile manifesto (2001)
4. Binamungu, L.P., Embury, S.M., Konstantinou, N.: Maintaining behaviour driven development specifications: Challenges and opportunities. In: 2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER). pp. 175–184. IEEE (2018)
5. Bruschi, S., Xiao, L., Kavatkar, M., et al.: Behavior driven development (bdd): a case study in healthtech. In: Pacific NW Software Quality Conference (2019)
6. Dees, I., Wynne, M., Hellesoy, A.:
7. Dookhun, A.S., Nagowah, L.: Assessing the effectiveness of test-driven development and behavior-driven development in an industry setting. In: 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). pp. 365–370. IEEE (2019)
8. Farooq, M.S., Omer, U., Ramzan, A., Rasheed, M.A., Atal, Z.: Behavior driven development: A systematic literature review. IEEE Access (2023)
9. Garousi, V., Felderer, M., Mäntylä, M.V.: The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In: Proceedings of the 20th international conference on evaluation and assessment in software engineering. pp. 1–6 (2016)
10. Garousi, V., Felderer, M., Mäntylä, M.V.: Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. Information and software technology **106**, 101–121 (2019)
11. Garousi, V., Mäntylä, M.V.: When and what to automate in software testing? a multi-vocal literature review. Information and Software Technology **76**, 92–117 (2016)
12. Garousi, V., Zhi, J.: A survey of software testing practices in canada. Journal of Systems and Software **86**(5), 1354–1376 (2013)
13. Hijriyani, S.N., Widowati, S., Kusumo, D.S.: Application of behavior driven development for validation testing (school information system). In: 2022 1st International Conference on Software Engineering and Information Technology (ICoSEIT). pp. 96–101. IEEE (2022)
14. Moe, M.M.: Comparative study of test-driven development tdd, behavior-driven development bdd and acceptance test–driven development atdd. International Journal of Trend in Scientific Research and Development **3**, 231–234 (2019)
15. Montvelisky, J., Lalit, B.: 2019 state of testing report. https://www.infoq.com/articles/2019-state-testing-report/ (2019)

16. Montvelisky, J., Lalit, B.: 2021 state of testing report: Testing careers, covid-19, agile, coaching training, automation. https://www.infoq.com/articles/2021-state-testing-report/ (2021)
17. Nascimento, N., Santos, A.R., Sales, A., Chanin, R.: Behavior-driven development: A case study on its impacts on agile development teams. In: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops. pp. 109–116 (2020)
18. Nascimento, N., Santos, A.R., Sales, A., Chanin, R.: Behavior-driven development: an expert panel to evaluate benefits and challenges. In: Proceedings of the XXXIV Brazilian Symposium on Software Engineering. pp. 41–46 (2020)
19. North, D.: Introducing bdd. https://dannorth.net/introducing-bdd/ (2006)
20. North, D.: Jbehave. a framework for behaviour driven development (bdd) (2012)
21. Pai, M., McCulloch, M., Gorman, J.D., Pai, N., Enanoria, W., Kennedy, G., Tharyan, P., Colford Jr, J.M.: Systematic reviews and meta-analyses: an illustrated, step-by-step guide. The National medical journal of India **17**(2), 86–95 (2004)
22. Pereira, L., Sharp, H., de Souza, C., Oliveira, G., Marczak, S., Bastos, R.: Behavior-driven development benefits and challenges: reports from an industrial study. In: Proceedings of the 19th International Conference on Agile Software Development: Companion. pp. 1–4 (2018)
23. Purkayastha, S., Goyal, S., Phillips, T., Wu, H., Haakenson, B., Zou, X.: Continuous security through integration testing in an electronic health records system. In: 2020 International Conference on Software Security and Assurance (ICSSA). pp. 26–31. IEEE (2020)
24. Rodríguez, G., González-Caino, P.C., Resett, S.: Serious games for teaching agile methods: A review of multivocal literature. Computer Applications in Engineering Education **29**(6), 1931–1949 (2021)
25. Scandaroli, A., Leite, R., Kiosia, A.H., Coelho, S.A.: Behavior-driven development as an approach to improve software quality and communication across remote business stakeholders, developers and qa: two case studies. In: 2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE). pp. 105–110. IEEE (2019)
26. Shaw, M.: Writing good software engineering research papers. In: 25th International Conference on Software Engineering, 2003. Proceedings. pp. 726–736 (2003)
27. Silva, T.R., Fitzgerald, B.: Empirical findings on bdd story parsing to support consistency assurance between requirements and artifacts. In: Evaluation and Assessment in Software Engineering, pp. 266–271 (2021)
28. Solis, C., Wang, X.: A study of the characteristics of behaviour driven development. In: 2011 37th EUROMICRO conference on software engineering and advanced applications. pp. 383–387. IEEE (2011)
29. Storer, T., Bob, R.: Behave nicely! automatic generation of code for behaviour driven development test suites. In: 2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM). pp. 228–237. IEEE (2019)
30. Yang, A.Z., da Costa, D.A., Zou, Y.: Predicting co-changes between functionality specifications and source code in behavior driven development. In: 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR). pp. 534–544. IEEE (2019)
31. Zhou, X., Jin, Y., Zhang, H., Li, S., Huang, X.: A map of threats to validity of systematic literature reviews in software engineering. In: 2016 23rd Asia-Pacific Software Engineering Conference (APSEC). pp. 153–160. IEEE (2016)