Journal of Universal Computer Science, vol. 30, no. 13 (2024), 1829-1848
submitted: 4/12/2023, accepted: 4/7/2024, appeared: 28/12/2024

Guided Reading: A Case Study in Evaluating Scientific Reports on Quantitative Research

Alessander Osorio
(Universidade Federal de Pelotas, Pelotas, RS, Brazil
https://orcid.org/0000-0002-7576-0958, alessander.osorio@inf.ufpel.edu.br)

Paulo Roberto Ferreira Jr.
(Universidade Federal de Pelotas, Pelotas, RS, Brazil
https://orcid.org/0000-0001-5631-6766, paulo.ferreira@inf.ufpel.edu.br)

Gerson Geraldo H. Cavalheiro
(Universidade Federal de Pelotas, Pelotas, RS, Brazil
https://orcid.org/0000-0002-4314-3429, gerson.cavalheiro@inf.ufpel.edu.br)

Abstract: Researchers have access to a vast repository of scientific papers reporting research results through the Internet. In this context, fostering collective knowledge growth, researchers may encounter situations where the content of a paper is not effectively utilized due to flaws in the presentation of conducted research. Various fields of knowledge have developed research protocols to assist in the production of scientific reports. In the field of Computer Science, proposals for such protocols are still in development. This paper describes a case study conducted with undergraduate and graduate students in Computer Science, involving guided reading of scientific papers, and recording students' perceptions regarding adherence to a set of guidelines for the presentation of scientific results. The case study focused on papers presenting research results with a quantitative focus, such as performance evaluations. The exercise took place over two academic semesters, where reading instructions were provided, and impressions about the observed content were collected every two or three weeks. The analysis results indicated that, despite being well-written, the papers had gaps in relation to the reading guidelines. The study's conclusions suggest that authors underestimate the importance of providing a comprehensive account of their experience, highlighting the need for protocols to report scientific results in the field of Computer Science.

Keywords: Scientific development, Research education, Research guidelines, Quantitative research, Reporting guidelines
Categories: A.2, L.0.0
DOI: 10.3897/jucs.116830

1 Introduction

For centuries, scientists have relied on trust, self-correction mechanisms, and community traditions to ensure research integrity – this approach succeeds because it's widely acknowledged that science relies on these principles, and high standards and reputation are crucial for scientists [Institute 1992]. Computer Science is a relatively new science field [Ghezzi 2020, Hassani 2017], yet the practice of sharing knowledge through scientific research reports in the form of scientific papers has been adopted.

In the context of scientific software development within academic and research environments related to Computer Science (CS), the challenges faced are not significantly different from those encountered in industrial settings, even though in the former the focus is on disseminating knowledge and in the latter, on product generation. Some of the reasons for this similarity, such as time constraints and competing priorities, have been cited by researchers [Nguyen-Hoan 2010, Kelly 2008, Mourão et al. 2024, Cico et al. 2021]. These factors often lead scientists engaged in software development to allocate limited time and resources elsewhere. However, it is noteworthy that theoretical aspects related to the developed software often find their way into scientific papers [Pawlik and Petre 2012, Hermann and Fehr 2022]. Despite this, several studies, e.g., [Tichy 1995], [Wainer et al. 2009], [Tedre 2014], and [Osorio et al. 2020], have reported shortcomings in result documentation within scientific papers. In particular, [Osorio et al. 2020] strongly recommends the adoption of research and reporting guidelines to enhance the quality of communication within scientific papers, addressing the deficiencies in result documentation.

The present paper builds upon the conclusions presented in [Osorio et al. 2020]. In this context, the paper introduces a proposal for a protocol for presenting results in scientific papers. With this protocol, we aim to contribute to the discussion on how to address the weaknesses in substantiating and consolidating research results in CS in scientific papers, as documented in [Wainer et al. 2009], [Adler et al. 2015], and [Osorio et al. 2020]. These problems, among others, are related to the authors' lack of knowledge in evaluative techniques or gaps in their research education. Other research areas, such as Health, Chemistry, and Biology, have experienced similar situations over time [Jedlitschka et al. 2014], and the adoption of guides or protocols has proven to be greatly helpful in solving these issues [Malta et al. 2010] [Munafo et al. 2017]. In CS, efforts to improve the quality of research presentation have not yet yielded a mature research protocol, although the literature presents a few sporadic initiatives concentrated on specific domains of CS in the last 14 years. The problems in verifying results and the absence of guidelines or protocols in CS that target academic educational purposes are the motivations for this work.

This paper aims to extend the discourse on the adoption of research protocols in CS, with a particular focus on the development of the RODA (Report Oriented Directions for Academics) research protocol. Scientific publications serve as the primary medium for academics and researchers to document their software development projects. Therefore, the quality of these publications significantly impacts communication among potential stakeholders, including other researchers and companies interested in adopting the presented technology. In this paper, our contribution to the discussion encompasses not only presenting motivational aspects but also documenting the validation process of the RODA protocol within the context of the Computer Science research group at our own university. Additionally, we provide a case study illustrating the protocol's applicability in a classroom setting, involving both graduate and post-graduate students.

The remainder of the text is organized as follows. Section 2 presents some papers found in the literature related to our work, Section 3 highlights requirements aspects considered relevant for a CS research protocol and the methodological framework of present work. Furthermore, Section 4 provides a concise overview of RODA, including its validation process. In Section 5, the case study is developed. Section 6 concludes this paper.

2 Related Work

In this section, we present some works identified in the literature that have approaches similar to the experiment documented in this paper. The survey of related works was conducted through a narrative (or traditional) literature review ([Arshed and Danson 2015]), not applying a specific research string, refusal, or selection method.

The first paper is entitled "Reading a computer science research paper" [Fong 2009]. We begin the presentation of related works with this paper, dated 2009, as its title captures the essence of the addressed problem: how to read and absorb the content of a scientific paper in the field of Computer Science. This paper, aimed at graduate or senior undergraduate students, proposes to be a tutorial for reading scientific papers, dividing the reading process into three stages: comprehension, evaluation, and synthesis. In these stages, the paper's reader should be attentive to identifying answers to a series of questions related to each phase of the proposed process. The author of this paper mentions applying their reading process to guide students in the activity of paper reviewing. The form used to facilitate this educational activity is outlined in the paper, although there is no documentation of an application.

The second paper is "How to Write and Read a Scientific Evaluation Paper", also dated 2009. In this paper, the authors emphasize that it is common to observe, in papers about empirical research (the paper is centered in the field of Software Engineering), a deficiency in the reports about the investigated artifacts and experience narratives. The authors present a checklist to be observed both in the process of writing scientific reports and in the reading of these reports. This checklist is presented in the form of questions to be answered considering the content of the report being written or read. The paper also makes an important caveat by stating that, despite the checklist, the reader's judgment, based on their own experience, is relevant in the process of knowledge absorption.

The third paper we list is "Active versus passive reading: how to read scientific papers?" The author argues that the reading process allows for better absorption of content when the reader engages in active reading. In this active reading, the reader applies the method proposed in the paper. The method, based on four initial questions, encourages the reader to predict the paper's content and analyze how closely their prediction aligns with what is reported in the paper. Although there is no report on the method's application, the author describes its adaptation as a discipline in an undergraduate course.

The documentation of a practice carried out in the classroom is effectively presented in the fourth paper, "Teaching Computer Science Students to Communicate Scientific Findings More Effectively" [Wyrich 2023], which we list as related work. In this paper, the authors describe the organization and implementation of a discipline for graduate students. Students were provided with a list of papers to choose for reading (10 papers) and a set of guidelines for constructing a portfolio to present the selected paper. The description of the items required for this portfolio was the only formal guidance on the data to be extracted from the paper to be read. No quantitative or qualitative analysis was produced after the reading processes, with most analyses focused on the application of the discipline itself.

With [Peng et al. 2022] and [Yuan et al. 2023], we present other types of work in which computational tools are used to support reading.

In [Peng et al. 2022], the presented tool aims to provide assistance in understanding a paper by proposing a set of questions to stimulate critical thinking regarding the presented content. This paper is also interesting for confronting the perception of using guidelines (or protocols) for reading scientific reports. The results of two experiments conducted with readers are documented in this paper. The first experiment was carried out with a group of volunteers, all of whom self-identified as regular readers of scientific papers. This heterogeneous group, characterized by varying levels of education, indicated that the tool, by interacting with readers, demonstrated greater value than guidelines presented in a static form. On the other hand, another experiment conducted with a uniform group, junior undergraduates with little or no experience reading papers, indicated that for readers in this situation, the reading assistance tool added little or nothing compared to static guides. As a note in this paper, all participants in both experiments declared English to be their native language.

The approach adopted in the CriTrainer tool, presented in [Yuan et al. 2023], aims to promote the training of readers in the critical reading of scientific papers. The approach provides adaptive support, where summaries of the main ideas of the text being read are presented, along with a series of questions generated from templates to help the reader understand the main ideas and develop critical thoughts. The paper also reports two experiments conducted by the authors. In the first experiment, conducted with volunteers who self-declared their experience level at one of three possible levels (considered in the analysis of results), the utility of features offered by a reading training tool was evaluated. The tool was modeled based on the results obtained in this first experiment. The second experiment assessed the use of the tool by undergraduate students with little research experience. In this second experiment, the usability of the proposed tool is compared with a static tool. The reading comprehension of a paper using both tools was compared in this experiment. One conclusion drawn from the qualitative evaluation performed on the analyses produced by the study subjects is that the interactivity of the tool makes the reading process more fluid. The authors also present, as a conclusion, that the reading training process is quite lengthy, and such a skill set involves other abilities like answering critical thinking questions. The subjects in both experiments have English as their second language.

Among the six works that we listed as related and selected to present in this section, we identified that the first four conducted experiments with students to improve their paper reading skills. In these presented examples, the experiments were based on providing students with a set of guidelines, more or less formalized depending on the type of approach reported, to be observed when reading the text. The format of "questions" to be answered based on the content of the paper is the most common. In the cited works, the origin of such guidelines is not actually reported. In our approach, we also document an exercise in directed reading of papers. However, unlike the others, the foundation of the guidelines to be observed was built from a broad study of research protocols existing in the literature.

The next two works were presented as an illustration of how computational tools can assist in the process of reading and understanding scientific papers. The presented reading protocol, in its current state, consists of a set of static guidelines, in the sense that they are presented in the form of a list of requirements. Adapting it to an automated tool can expand the potential adoption of the proposed protocol. However, focusing on the training of new researchers, broadening the scope of the protocol and consolidating it are priorities compared to the creation of such a tool.

3 Contextualization

In this section, we present the premises behind the proposed guide. Initially, we discuss the importance of adopting guidelines for constructing scientific reports, illustrated with examples from the context of software production processes, even in industry, based on works identified in the literature associated with Software Engineering. Subsequently, we provide a brief overview of the findings of a Systematic Literature Review (documented in [Osorio et al 2020]) related to identifying research protocols in CS.

3.1 Positioning guide requirements

Members of the scientific community, at least most of them, may consider that the primary means of disseminating the results of their work is through the publication of scientific papers. However, in comparison to other fields, CS lacks consolidated research protocols. A side effect of the absence of protocols may, in part, explain why [Brings et al. 2018] states that scientific publications are criticized as unsuitable and often seen as difficult to comprehend by industry [Yamashita 2015]. To illustrate this situation, we found in [Franch et al. 2022] the report of a survey conducted with industry developers regarding their perceptions of the relevance of scientific papers. In the study conducted, the respondents were provided with a summary, constructed from a standardized template, of the evaluated paper.

Scientific research, whether it seeks to investigate, explain, or create something, is characterized by its research objective. To maintain credibility, it is essential that every phase, from inception to conclusions, is meticulously detailed in the final report – usually documented as a scientific paper in a journal or in an event. This thorough documentation enables the verification, validation, and reproducibility of its efficiency and effectiveness.

The nature of the research object plays a pivotal role in shaping the methodology, development approach, and evaluation process. A classification of these research objects, presented in [Peffers et al. 2012], includes research objects such as Algorithm, Constructors, Framework, Instantiation, Method, and Model. These categories align with evaluative methods, which, as articulated by the same author, encompass Logical Argumentation, Technical Evaluation, Experimentation, Subject-Based Experimentation, Action Research, Prototype, Case Study, and Illustrative Scenario.

For effective evaluation, it is crucial that the chosen evaluative methods demonstrate statistical significance, a prerequisite that depends on the nature of the measurements undertaken and the appropriate statistical tests applied, as emphasized by Wainer [Wainer et al. 2007]. Statistical tests serve as systematic procedures for scrutinizing measurement results, providing them with significance and reliability. When comparing measurements, these tests ensure that observed differences hold genuine statistical importance. Such statistical details should be explicitly documented in the research report. To support the development of materials for disseminating the results of scientific research, notably scientific papers, both structural models for organizing content in papers and guidelines for selecting the subjects to be covered have been proposed.

IMRaD is a widely accepted model for structuring scientific papers in the scientific community. IMRaD stands for Introduction, Method, Results, and Discussion, indicating the main sections to be included in the proposed structure. A study, documented in [Makowska 2018], provides an insight into the adoption of this structural pattern in papers from technological areas (microelectronics, computer studies, and telecommunication). As a general conclusion, the paper identified IMRaD as being highly useful as a tool for organizing scientific reports, although it identified that its application varies according to the specific needs of the research field discussed in the report. In particular, the author concludes that papers are more likely to follow the IMRaD structure when presenting results of conducted research, and when dealing with other topics, the observed structure tends to be more theoretical.

Research guides are comprehensive documents that offer a set of guidelines, many of which are well-established, intended to facilitate the creation of scientific publications. In the field of Health, notable guides like PRISMA [Moher et al. 2009], STROBE [Ebrahim and Clarke 2007], and CONSORT [Moher et al. 2012] are frequently cited as references for researchers. In CS, the adoption of research protocols can enhance the credibility and acceptance of the presented report, similar to the exemplified areas. Additionally, it can facilitate its uptake by the industry. Furthermore, it is essential to consider the importance of simplifying the reading of these documents.

3.2 Research Protocols in CS

A systematic literature review was conducted with the aim of identifying research protocols or guidelines for Computer Science (CS) [Osorio et al 2020], conducted according to the protocol proposed in [Kitchenham 2007]. In this section, we present the proposals that were uncovered during this review. Our research encompassed papers published in journals and conferences indexed by the primary databases in the field of CS[1] over the past 14 years.[2] However, during this period, only five research guides specific to various domains of CS were identified. As of December 2022, we have found no additional publications that would allow us to determine whether these guides have been effectively adopted or if there is an increasing trend in their development. The limited number of research guides available for CS has previously been highlighted in [Runeson 2009]. The complete documentation of this Systematic Literature Review (SLR) can be found in [Osorio et al 2020]. However, we still provide the search string used:

“computer science” AND (“reporting standards” OR “reporting evaluation” OR “reporting guidelines” OR “reporting statements” OR “reporting checklist” OR “research standards” OR “research evaluation” OR “research guidelines” OR “research statements” OR “research checklist”)

The work by Kitchenham et al., as outlined in [Kitchenham 2022], provides a comprehensive set of guidelines for reporting secondary studies within the field of software engineering. These secondary studies typically involve the mapping or systematic organization of findings from primary studies, essentially constituting review studies. It’s important to note that this approach draws inspiration from the field of Health.

Garousi et al. propose a set of 14 guidelines primarily tailored for systematic literature reviews. Notably, these guidelines allow for the inclusion of publications from unconventional sources not typically associated with academia, such as blogs, videos, and internet documents [Garousi et al. 2019]. In contrast, De Moura et al. offer a list of 22 guidelines specifically designed for simulation models [França 2016]. However, it’s important to note that these guidelines are exclusively geared toward experimental research and do not follow the conventional methodological structure often seen in scientific reporting.

Jedlitschka [Jedlitschka et al. 2014] propose a set of guidelines tailored to address organizational needs, categorized into three distinct groups, with a specific focus on facilitating reporting in scientific publications within organizational contexts. In contrast, Budgen [Budgen et al. 2009] provide a list of nine guidelines specifically crafted for scientific publications, targeting observational studies within the software engineering field. These guidelines stem from an extensive literature review that identified a solitary

[1] ACM, IEEE, Science@Direct, and Springer.
[2] From 2009 to 2022.

reference publication and are supplemented by the authors' empirical knowledge. While these guidelines offer a structured approach, their primary emphasis lies in the qualitative analysis of results, which may not comprehensively cover the methodological intricacies typically expected in scientific publications.

In [Vrhovec et al 2020], we find the work presented by Vrhovec et al. in which a guide is proposed to promote improvements in the quality of research reports. This work was not identified in the SLR documented in [Osorio et al 2020]; however, its authors propose a set of guidelines, which they call SAE-CSAR[3], extending the IMRaD structure to make the dissemination of scientific results in scientific papers more effective. The authors delimit their contribution, at its current stage, to reports of case studies and action researches. The validation of the guidelines was conducted through a survey on the adherence of the proposed guide's guidelines in a collection of papers involving this type of research in the field of Cybersecurity.

We also observed that a substantial portion of initiatives in the field of CS draws inspiration from studies conducted in other domains, with the Health sector being particularly prominent due to its extensive adoption within the scientific community. In light of this, we identified the most commonly utilized initiatives within this context and selected three of them as the cornerstone for our work.

3.3 Methodological Framework

The objective of this paper is to present a case study in which a guided reading of scientific papers was conducted by a group of undergraduate and graduate students, master's and doctoral degrees. For this purpose, we first introduce the guide utilized as a tool for guiding the students through the reading process. This guide is presented in Section 4. The guide is presented in a summarized form, with emphasis given to its validation process, which was carried out through two instruments. The first one considers the perspective of compliance, according to the proposal presented by [Kitchenham 2008], of the proposed guide in relation to addressing the expectations of different categories of readers when reading scientific papers. The second involved conducting a survey involving researchers from different areas of Computing working in postgraduate studies.

The main contribution of the paper, the case study of a guided reading, is found in Section 5. This case study unfolded over two academic semesters. The students participating in the study analyzed different scientific papers and noted their perceptions in response to meeting the recommendations proposed by the guide. These perceptions were documented in closed-ended multiple-choice questions, with more than one question for each guidance. The result documented in this paper consists of a qualitative analysis of the set of analyses obtained at the end of the semester in which the guided reading was conducted.

4 The *RODA*

During the development of RODA [Osorio et al 2020], we conducted two case studies to validate the arguments presented in the literature regarding the validation of scientific research results. Our meta-analyses, encompassing 1013 papers from 2000 to 2018, revealed that only 3% ($p \leq 0.05$)[4] of these papers employed statistical tests to substantiate

[3] SAE-CSAR stands for Structure-based Approach for Evaluating Case Study and Action Research Reports.

[4] According to the Kruskal Wallis test for variables with more than two categories with non-Gausian distribution.

their findings. This statistic aligns with the arguments outlined in the existing literature.

Our systematic review significantly contributes to achieving the primary goal of this work while also addressing several secondary objectives. We identified three initiatives from the field of Health that served as the foundational basis for RODA. This process enabled us to pinpoint the information relevant to scientific publications within a methodological framework tailored to CS and specify the content that should be included within this structure.

Additionally, our systematic review played a crucial role in identifying key stakeholders in the publication process, such as Researchers, Consultants, Meta-analysts, Replicators, Reviewers, and Authors. These roles were proposed by Kitchenham et al. [Kitchenham 2008] and greatly informed the validation of our final proposal.

Throughout our research, we noted the absence of a uniform taxonomy for classifying types of studies within the realm of CS. It became evident that different authors employ varying nomenclatures based on their unique perspectives. In the systematic review, we identified "Objects of Study" and "Evaluative Methods," as initially proposed by Peffers et al. [Peffers et al. 2012]. Combining these elements, as articulated by Wazlawick [Wazlawick 2017], research can manifest in diverse presentation formats, including "Product," "Something Different," "Something Presumably Better," "Something Recognizably Better," and "Trial." These formats are further categorized into three primary types of research: "Formal," "Empirical," and "Exploratory." Moreover, an alternative source[5] identifies three research types: "Theoretical," "Engineering," and "Empirical." In our proposed guide, we have adopted a nomenclature based on the type of study object and the type of evaluative method.

Utilizing the insights garnered from our systematic review, we crafted a preliminary version of RODA and subjected it to two validation processes. The first process considered the perspectives of various research subjects, a methodology proposed by Kitchenham et al. [Kitchenham 2008]. This approach involved delineating the requirements and queries that a publication must address to cater to each reader profile associated with the research.

This validation method mirrors the approach undertaken by De Moura et al. [França 2016] and Runeson et al. [Runeson 2009] in assessing the comprehensiveness of their propositions. In our work, we followed a similar path, and the results are meticulously detailed in Table 1. Notably, our analysis reveals that, based on the gathered data, only the Researcher and Consultant profiles did not receive complete answers from the preliminary proposal. In both cases, the question "Is it a relevant paper?" (Question 2) remained a subject of subjective interpretation.

The essence of this question lies in its inherent subjectivity, as it hinges on the perspective of the individual reader. The determination of a paper's relevance is shaped by the reader's viewpoint and preferences. Different readers, driven by varying understandings and interests, may thus react diversely to the same work. It is worth noting that the application of a research guide during the preparation of a scientific paper serves to enhance its engagement, precision, and transparency, thereby augmenting the likelihood of readers finding it pertinent.

Our assessment of the preliminary protocol identified areas warranting improvement. Notably, with regard to the Researcher profile, one area pertains to the previously mentioned issue, while another relates to the extent of scientific contextualization within the research. In the Consultant profile, beyond the previously mentioned aspect, we identified five additional facets necessitating enhancement, particularly focusing on

[5] Scientific Methodology – https://www.youtube.com/c/jpsauveut Accessed 06/16/2021.

Perspective/Subject	Answered questions	Improvements
Researcher	16 (of 17)	2
Consultant	21 (of 22)	6
Meta-analyst	14	2
Replicator	9	1
Reviewer	7	-

Table 1: Proposal evaluation according to research profiles

the applicability of research results within real-world contexts (external validity). It is noteworthy that our preliminary protocol does not individually address these aspects; rather, it guides the inclusion of information about internal and external result validity and associated risks in a specific section.

Within the Meta-analyst profile, two questions demanded refinement, particularly concerning the suitability of analytical methods for the study model and the management of missing or dropout data. Lastly, the Replicator profile underscored the importance of bolstering training requirements for executing research. The final iteration of RODA encompasses all these adjustments.

For the second validation phase aimed at confirming the adequacy of our work's proposal, we conducted a survey among 29 professors in a Computer Science graduate program. This survey featured a two-part data collection instrument. The first part consisted of questions related to the suitability of our work's proposal and the accompanying support protocol. In the second part, we gathered individual data from each professor to better understand their roles as researchers.

To administer this instrument, we utilized an electronic questionnaire hosted on a restricted-access website. Professors were invited to participate via email, where they received a brief explanation of the study and a link to access the questionnaire. Upon clicking the link, participants were directed to an introductory page containing more information about the research, along with a consent form. After providing their email addresses and an encryption key to ensure the confidentiality of their responses, participants gained access to the first part of the instrument.

In the section concerning the adequacy of the proposed protocol, we individually presented the guidelines from the guide, along with a description of their respective purposes as outlined in the preliminary version. After this presentation, respondents were asked to evaluate the protocol's adequacy through a closed-ended question, offering response options such as "Adapted," "Needs Improvement," "Inadequate," and "I do not have an opinion." Subsequently, we provided an open-ended question to allow respondents to explain their reasons for suggesting improvements, deeming it inadequate, or indicating their inability to form an opinion. To conclude this section of the questionnaire, an open-ended question was included to solicit comments, suggestions, or criticisms regarding the paper's proposal.

The second part of the instrument focused on characterizing the faculty members whose data were collected. This section included questions related to the respondent's research experience, years of training, experience in writing and publishing papers, and familiarity with research guides. These data were collected to establish potential relations with the feedback received in the first part of the instrument.

Out of 26 professors affiliated with the Computer Science and Computer Engineering

Answer Option	**n**	**%**
Suited	528	84.5
Require improvements	82	13.1
Inadequate	11	1.8
Can not comment	4	0.6
Total	625	100

Table 2: Overall Adequacy Assessment

undergraduate courses, and Computer Science Graduate Program at the university studied, 25 professors (96% of the total) submitted valid instruments. These valid instruments were defined as those containing responses in at least one section. It is worth noting that all participants accessed the instrument's form. However, one professor merely viewed the form, while two others did not complete the respondent's characterization section, providing responses solely in the proposal adequacy section.

In the analysis, the majority of responses were considered, with 84.5% ($p \leq 0.05$[6]) of the faculty rating the proposal as adequate, as detailed in Table 2. This high level of adequacy is encouraging, given that responses categorized as other than "Adequate" provided constructive feedback that significantly contributed to the refinement of the final protocol.

To ensure transparency and reproducibility, we have made available the report of the instrument's data analysis, conducted in RStudio, which includes the summarized data and results presented here. You can access this report, along with additional information about the study, at the following link: https://github.com/roda-protocol/roda.

Armed with a solid theoretical foundation, we have meticulously designed RODA (Table 3). RODA comprises a comprehensive set of 26 guidelines, thoughtfully organized to align seamlessly with the logical structure of a typical scientific publication. These guidelines encompass the key sections of a research paper, including the Cover Page or First Page, Introduction, Methodology, Results, Discussion, and Conclusion. Each section is accompanied by specific guidelines that outline the necessary information to be included.

The adaptability of RODA's structure, and the flexibility to incorporate relevant information, are closely tied to the specific publication model of the journal where the research report is intended for submission. Moreover, there are no restrictions preventing the integration of supplementary information related to one guideline within a different section, deviating from RODA's original proposal. In the context of this guide, the term "section" primarily denotes the category of information that aligns with the corresponding recommendations. We anticipate a commitment to comprehensive information, with authors placing it within the reporting context they find most fitting.

Considering the potential level of detail within some guidelines, which may be constrained by journal space limitations or page count restrictions, authors have the opportunity to enhance their research by providing additional information in a publicly accessible repository, duly cited in their report. This supplementary information not only enriches the study but also bolsters its reliability and transparency. To this end, we have

[6] According to the Kruskal Wallis test for variables with more than two categories with non-Gausian distribution.

made supplementary information about our work publicly available.[7]

Guideline		Description
Section - Title page / First page		
O1	Title	It is desirable that the title, besides the description of the technique developed, contains the type of study object and the evaluation method.
O2	Abstract	The abstract must be the miniature of the work. It is desirable to include the scientific background, objectives, methods, the main result and conclusion of the work.
O3	Authorship	Present the authors' names, their affiliations, and at least a contact for further information.
O4	Acknowledgments or Contractual Agreements	Citing important persons or institutions, or bound contractual terms with funding agencies or third parties.
Section - Introduction		
O5	Scientific context	Justify and place the work within the scientific context of what has already been done.
O6	Objectives	Cite the specific objectives of the paper, including possible research questions and hypotheses
Section - Methodology		
O7	Protocol	Cite the protocol/guideline used in the research.
O8	Research Context	Contextualize the research object within the environment in which it was developed, including key dates and places.
O9	Type	Show and justify the type of research object and evaluative method.
O10	Sample Size	Show the total number of samples to be taken and the number of repetitions of the experiment.
O11	Data Source	Show the source of the data to be collected.
O12	Items/Variables of Interest	Describe all the variables and metrics that will be collected.
O13	Criteria	Describe the inclusion criteria of data in the collection.
O14	Data Collect	Describe the entire process used with the data sources described in the study design to capture the data.
O15	Publicity or Access	Explain the publicity of the research object and data.
O16	Data Analysis	Describe the statistical methods used.
O17	Results Summary	Summarize the obtained data in the analysis.
O18	Data Synthesis	Describe the main findings of the data summary.
Section - Results		
O19	Risks	Describe the possible biases and risks to the results validity.
O20	Selected	Highlight the total data and groupings.
O21	Results Individualization	Evidence the results in an individual way.
O22	Additional Analyses	Report additional analyses.
Section - Discussion		
O23	Evidence Summarizing	Summarizing the major results.
O24	Limitations	Describe the limitations of the research.
Concluding Section		
O25	Conclusions	Describe the interpretations of the results.
O26	Continuity	List possible continuations, developments, complementing of the research and future works.

Table 3: RODA Quick Guide.

5 Case Study: a Guided Reading

This section documents a case study in which a guided reading process of scientific papers was conducted. The experiment involved guiding students from Bachelor's programs in Computer Science and Computer Engineering, as well as graduate students in CS, master and doctoral degrees, through the process of reading scientific publications.

A total of 38 students participated in the activity during the 2022/1 and 2022/2 semesters. The classes comprised a total of 29[8] undergraduate students and 9[9] graduate

[7] https://github.com/roda-protocol/roda

[8] 2022/1, Special Topics in Computing III.

[9] 2022/1, Research Seminars in Computing II, and 2022/2, Special Topics in Computing I.

students at the master's or doctoral level. None of those participants is a native English speaker, and the undergraduate students have little to no research experience.

The RODA protocol was used as a guide for reading and studying the publications. Each student analyzed three papers during the process: two were freely chosen from the literature, and the third was selected from among the choices made by their peers. As a result, 76 papers were analyzed, with half of them undergoing two evaluations. The analysis of the papers was conducted in six stages, one for each group of recommendations in RODA. Each student recorded their impressions of how well these papers adhered to RODA's guidelines at the end of each stage.[10] The results of these analyses are presented in Section 5.1.

The papers could have been chosen from journals or conferences, irrespective of their subarea. However, the journal or conference needed to hold a "Qualis" restricted classification and provide results in the form of quantitative analysis, representing a performance evaluation.[11] Additionally, the selection was limited to full papers, allowing authors adequate space for discussing their findings.

In applying the method, the different sets of criteria in the RODA protocol were presented in class, and then survey forms[12] was applied for students to record their perceptions of the papers. As there are six recommendation groups in RODA, there were six stages of guided reading, each lasting two to three weeks. The form included open and closed-ended questions, all of a qualitative nature, to record students' impressions regarding the adherence of the papers to the RODA protocol. The number of questions in each group (1 to 6) was 18, 11, 26, 9, 7, and 5 questions, respectively, with some guidelines having more than one question in the form. Thus, more than one question per form guidance. Among the questions presented, for each guidance, at least one had the following format "Does the paper satisfactorily meet orientation X?", with closed options allowing the student to inform (i) Yes, definitely, (ii) Probably yes, (iii) or (iv) I don't know how to answer. In the forms, the language used was Portuguese.

The development process for each stage consisted of three steps for each set of RODA guidelines. In the first step, conducted in the classroom, the professor presented and discussed the group of guidelines that was the focus of the stage. After a two- or three-week period of reading and analyzing the papers, students completed the corresponding form to record their impressions. During a round-table discussion, the professor prompted students to identify both positive and negative aspects observed in the analyzed papers. The professor used the responses in the forms as a foundation for facilitating the discussion. The set of responses was qualitatively analyzed.

5.1 Results

At the end of the two semesters, a total of 114 responses should have been recorded, considering the number of students who completed the course and the number of papers

[10] Forms, as well as the anonymized responses are available at http://www.github.com/roda-protocol/estudo-de-caso.

[11] The Qualis is a tool used by the Coordination for the Improvement of Higher Education Personnel (CAPES) in Brazil to assess the quality of academic journals and events (conferences and symposia). It classifies these journals and events into strata, using external metrics such as SCOPUS and Web of Science, as well as other criteria. Only journals and events on which Brazilian researchers have published within a specific time frame are considered. There are eight strata, four of which are referred to as restricted, indicating the journals or events with the highest scientific impact. The current journal and event classification are available at https://ppgcc.github.io/discentesPPGCC/pt-BR/qualis/.

[12] Forms were administered using the Google Forms tool.

they were supposed to analyze. However, the actual totals differed from the expected count, decreasing over time. This was due to students dropping out of the course and failing to submit their assignments, as shown in Table 4. In this table, you can also see the number of protocol guidelines per group and the number of questions in the form for each group.

Group / Section	Guideline	Questions	Answers
Group 1 / Title page	3	18	122
Group 2 / Introduction	2	11	115
Group 3 / Methodology	12	26	112
Group 4 / Results	4	9	108
Group 5 / Discussion	2	7	107
Group 6 / Conclusion	2	5	107

Table 4: Number of responses to the survey forms per group/section

The following results are presented individually for each protocol guideline. It is important to note that recording impressions regarding compliance with each guideline should involve responding to more than one question, as the conceptual density of the protocol guidelines requires multiple responses for comprehensive coverage. Once again, it is emphasized that the results summarized here are based on a qualitative analysis of the students' records during the readings of their respective papers.

Guideline 1: Overall, students report that the paper's title indicates the subject of study and its potential innovation. However, they observe that, in most cases, the title does not characterize the evaluative method used.

Guideline 2: In the analyses conducted, abstracts of the papers generally clearly presented both the scientific context and the work's objectives. However, in a significant number of cases, it was not possible to identify the development method applied in the research. Similarly, many analyses indicated that reading the abstract did not provide a perspective on the achieved results and the authors' conclusions.

Guideline 3: Author identification is clear in the vast majority of works, but it was reported that in two works, the authors were not clearly presented. Information about institutional affiliation and contact information was present in most papers, although significant cases were noted where this information was missing.

Guideline 4: Acknowledgments to research funding agencies were quite common in the analyzed papers, and a considerable number of papers mentioned individuals important to the research who did not participate in the papers themselves.

Guideline 5: Among the questions related to this guideline, students were asked to construct a sentence characterizing the scientific context of the chosen paper. Some students could only identify the paper's research area, while others could produce a synthesis paragraph. The majority of reports indicate that it was relatively easy to create the requested sentence, but some students reported that the paper did not provide the necessary information. There were also reports about papers that did not cite related works or papers that cited works seemingly unrelated to the research.

Guideline 6: More than 50% of the responses indicate that students had no difficulty identifying and understanding the research hypotheses in the analyzed papers. In the remaining responses, the majority reported difficulty in finding this information, and in some cases, the complete absence of it.

Guideline 7: This result reflects the current maturity of research in CS regarding the use of research protocols, as few papers mentioned any protocols. When mentioned, they referred to protocols developed for other areas than CS or specific to Software Engineering.

Guideline 8: The characterization of the research context was considered adequate in most papers. However, there were reports that details for interpretation were omitted or only the hardware used was described. Some reports indicated that temporal issues that could be relevant to the research were not considered. Additionally, in a significant number of reports, the location of the research object's development as reported in the publications could have some influence on the results but was not considered according to the reports. It is important to note that location also refers to the description of the conditions of the equipment used.

Guideline 9: Reports indicated that there were no difficulties in identifying the research object and the evaluative method in the papers. As expected, due to paper selection guidelines, most papers referred to the study of an algorithm and its experimental evaluation.

Guideline 10: Reports indicate that in the papers, the sample size is usually clearly presented, and students found the presentation satisfactory. However, the same does not apply to the number of repetitions, which was absent in half of the analyzed papers.

Guideline 11 and 12: The records of the analysis of these guidelines indicated that the source of the data and the metrics used throughout the paper were adequately characterized, whether it was from the research object itself or from another source.

Guideline 13 and 14: Most records indicated that the inclusion criteria for the collected data were not presented, although the majority of records indicated that the papers adequately described the parameters used in data collection.

Guideline 15: Most papers do not indicate the publication of data or research objects, meaning they do not make the sample data available for third-party use. A few works still indicate that the data used is confidential, but the mere omission of data availability or not indicates that this practice is not common in the field of Computing.

Guideline 16: Impressions regarding compliance with this guideline indicate that less than half of the analyzed papers clearly report the statistical methods used in the results analysis. Nevertheless, it was found that a significant number of papers present statistical analysis of results without specifying the method(s) used.

Guideline 17 and 18: Almost all papers presented their results using some visual resource (tables or graphs) in an objective manner. However, in about half of the papers, the presentation involved some data manipulation or transformation (e.g., logarithmic base, shifting of X and/or Y axes relative to the origin). Impressions also indicated that more than half of the papers did not provide any reference model for analysis and did not perform any comparison with related works.

Guideline 19: Based on the analysis of this guideline, students generally found that the papers provide robustness of results with good argumentation and acceptable confidence levels, although in some cases, the argumentation was considered superficial. Almost all students were unable to identify any situation or nuance of the subject that was not explored by the authors. This may be due to the fact that most students involved in the experiment are still in undergraduate programs.

Guideline 20: The complete set of collected data was observed in the majority of papers, and the results obtained by the dataset were statistically supported for just over half of the papers. Impressions also indicated that a significant portion of the analyzed papers intends to extend this analysis in the future.

Guideline 21: Also, in most papers, results were presented individually and visually using graphics and/or tables. The specifics of each result were equally highlighted and discussed.

Guideline 22: A significant number of students reported identifying references in the text to further analysis of the data recorded in the research results. In some cases, this extension was carried out within the paper itself, while in others, it was indicated as future work.

Guideline 23: In presenting the results, the objectives were clearly restated for the majority of the papers, and the results that allowed them to be achieved were presented. Just over half of the impressions recorded highlighted that, in the discussion, statistical analysis is mentioned, providing adequate confidence levels.

Guideline 25: Almost all impressions indicate that the papers highlight the main contributions of the work and position the obtained results, in one way or another, in relation to the state of the art.

Guideline 24 and 26: More than half of the papers present stages for continuing research or opportunities for future work, relating these opportunities to the limitations of the work performed.

5.2 Discussion

After evaluating the results in general, it was found that the students were able to extract from the text what was asked of them as long as they followed the RODA protocol guidelines. Additionally, it can be observed that the selected papers met the quality expectations set by the selection rules. However, even well-written papers have some open points according to the protocol guidelines, which may indicate that the authors overlook the importance of providing a complete account of their experience. The closed questions, which allow students to indicate the degree of adherence of the paper to each protocol guideline on a four-level scale, showed that the papers address the protocol's demands but without sufficient detail.

As can be gleaned from the record of impressions, the aspects considered negative or not entirely positive are specific. Problems were even found in the identification of authors, underscoring the importance of using research protocols. However, the references to these contents in undergraduate courses are not specific enough to draw attention to this point.

Regarding the results, the issues of publicizing the results, and even the object of study itself, were not openly and publicly disclosed in the vast majority of the analyzed works, hindering the basic principles of reproducibility and repeatability that science should have. Almost all papers report having achieved their objectives, and there is no case where a paper reports a negative experience or the complete failure of the developed research object.

It is worth noting that there are no recommendations for teaching Scientific Methodology and Statistics in postgraduate courses; there is only guidance to use "appropriate research methods" for the objectives of the work. This possibly reflects the fact that these contents have already been recommended in undergraduate courses, further reinforcing the argument for prior knowledge.

At the end of the analysis of the case study results, it is concluded that the protocol is fully applicable. The problems pointed out by the results are precisely part of the gap that the protocol's purpose seeks to address, reinforcing the argument for its full utilization.

6 Conclusions and Future Studies

In this paper, we introduce RODA, a research protocol comprising 26 guidelines. These guidelines are thoughtfully organized into six logical structural sections commonly found in scientific publications. RODA's inception was influenced by the experiences of other research domains that have successfully employed established research protocols such as PRISMA, STROBE, and CONSORT. These protocols served as a wellspring of inspiration for the development of our proposal. To validate the presented protocol, we conducted a comprehensive evaluation within an experimental research context, encompassing a diverse spectrum of investigations conducted in the field of CS.

Our primary objective is to elevate the quality of research by enhancing researchers' training, both in terms of improving the presentation of their research findings and documenting their software artifacts. The proposal underwent rigorous validation processes using established methodologies and received approval from the faculties of the Engineering and CS programs at studied university.

The main reason for this work comes from the difficulties that come with proving scientific results in the field of CS. We have found these challenges through detailed case studies, which showed a lack of statistical tests being used to confirm research findings. Nevertheless, it's important to mention that recent years have seen a change in how we view research results. They are no longer just presentations; now they are expected to be strong, dependable, and transparent from start to finish. This aligns well with the fundamental principles of scientific inquiry, which focus on innovation, reproducibility, and the sharing of knowledge.

Furthermore, these aspects become even more significant when considering the intended software lifecycle in an academic setting. The quality of a research paper not only helps share the obtained results but also allows the researcher or research group to become essential contributors to the exchange of experiences and the transfer of technology. This broadens the understanding of the work being done, especially in the context of requirement engineering, where clarity, reliability, and transparency are vital in shaping software development practices and encouraging the sharing of collaborative knowledge.

This transformative shift has already transpired in other academic domains, where the adoption of research guides has proven instrumental. Thus, there is a compelling need to focus attention on the issues highlighted herein. The findings of this work should prompt contemplation regarding the integration of statistical analysis, validation procedures, and result reproducibility methodologies into the foundational graduate disciplines of CS. The RODA Protocol emerges as a powerful tool in this regard, offering standards that have the potential to elevate the field of CS knowledge and enhance the quality of research outcomes.

Hence, the contribution of this work is still in its early stages in the Computer Science Department at our own university. This first version of RODA marks the start of a big change in how professors see and do their scientific research. Suggesting that everyone should use it to improve the CS research group's work is not just a good idea but really important. We also think that using RODA can make a big difference in how good and careful research is done in this area. It can help make the development of software requirements more organized and structured. This change can lead to software systems that work better and solve big problems in requirement engineering.

Within this perspective, some unfolding of this work is envisioned. The first is the improvement of the protocol itself. The verification of its adequacy to the various research types needs to be better studied. It is understood that RODA is suitable for experimental research, but concerning theoretical research, it needs improvement. To adapt the guidelines not only to the proof of the results through statistical analysis, but also to the theoretical inductive proof.

As part of our strategy to enhance research and publications among faculty and students, we propose an additional initiative, offering seminars for the dissemination of RODA's guidelines. These seminars would also provide comprehensive support and guidance to research participants throughout the entire research process, from the initial conception phase to the final report's elaboration. This initiative would be particularly beneficial within the context of our own university CS courses. The field of meta-science as applied to CS offers a vast and promising landscape for study. It presents numerous distinct avenues of exploration, making it a compelling area for future research initiatives.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

References

[Adler et al. 2015] Adler, S., Schmitt, S., Wolter, K. and Kyas, M. (2015) A survey of experimental evaluation in indoor localization research. *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-10.

[Arshed and Danson 2015] Arshed, N., Danson, M. (2015). The Literature Review. In R. MacIntosh, K. D. O'Gorman (Eds.), *Research Methods for Business and Management: A Guide to Writing Your Dissertation*. Goodfellow Publishers, pp. 31-49.

[Brings et al. 2018] Brings, J., Daun, M., Brinckmann, S., Keller, K. and Weyer, T. (2018) Approaches, Success Factors, and Barriers for Technology Transfer in Software Engineering—Results of a Systematic Literature Review. *Journal of Software: Evolution and Process*. 30(11): e1981.

[Budgen et al. 2009] Budgen, D. and Others (2009) Preliminary Reporting Guidelines for Experience Papers. *International Conference on Evaluation and Assessment in Software Engineering (EASE)*. 13 pp. 1-10.

[Cico et al. 2021] Cico, Orges; Jaccheri, Letizia; Nguyen-Duc, Anh; Zhang, He (2021) Exploring the intersection between software industry and Software Engineering education - A systematic mapping of Software Engineering Trends. *Journal of Systems and Software*, 172, pp. 110736.

[Ebrahim and Clarke 2007] Ebrahim, S. and Clarke, M. (2007) *STROBE: New standards for reporting observational epidemiology, a chance to improve*. Oxford University Press.

[Ferrari et al. 2021] Ferrari, A., Zhao, L. and Alhoshan, W. (2021) NLP for Requirements Engineering: Tasks, Techniques, Tools, and Technologies. *Proceedings of the 43rd International Conference on Software Engineering: Companion Proceedings*. pp. 322-323.

[Fong 2009] Fong, P. (2009) Reading a Computer Science Research Paper. *ACM SIGCSE Bulletin*. 41(2): 138-140.

[França 2016] França, B. and Travassos, G. (2016) Experimentation with dynamic simulation models in software engineering: planning and reporting guidelines. *Empirical Software Engineering*. 21(3): 1302-1345.

[Franch et al. 2022] Franch, X., Mendez, D., Vogelsang, A., Heldal, R., Knauss, E., Oriol, M., Travassos, G., Carver, J. and Zimmermann, T. (2022) How do Practitioners Perceive the Relevance of Requirements Engineering Research? *IEEE Transactions on Software Engineering*. 48(6): 1947-1964.

[Garousi et al. 2019] Garousi, V. and Others (2019) Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*. 2019(106): 101-121.

[Ghezzi 2020] Ghezzi, Carlo (2020) What Is Research and Why We Do It. *Being a Researcher*, pp. 1-23, Springer.

[Hassani 2017] Hassani, Hossein (2017) Research methods in Computer Science: The challenges and issues. *arXiv preprint* arXiv:1703.04080, New York. https://doi.org/10.48550/arXiv.1703.04080

[Hermann and Fehr 2022] Hermann, S. and Fehr, J. (2022) Documenting research software in engineering science. *Scientific Reports*. 12(6567): 2045-2322.

[Institute 1992] Institute of Medicine (1992) *Responsible Science: Ensuring the Integrity of the Research Process*. The National Academies Press.

[Jedlitschka et al. 2014] Jedlitschka, A. and Others (2014) Reporting experiments to satisfy professionals' information needs. *Empirical Sw. Eng.*. 2014(19): 1921-1955.

[Kelly 2008] Kelly, D., Sanders, R. and Others (2008) Assessing the quality of scientific software. *First International Workshop on Software Engineering for Computational Science and Engineering*.

[Kitchenham 2007] Kitchenham, B. and Charters, S. *Guidelines for performing systematic literature reviews in software engineering*. Citeseer.

[Kitchenham 2008] Kitchenham and Others (2008) Evaluating guidelines for reporting empirical software engineering studies. *Empirical Software Engineering*. 2018(13): 97-121

[Kitchenham 2022] Kitchenham, B., Madeyski, L. and Budgen, D. (2022) SEGRESS: Software Engineering Guidelines for REporting Secondary Studies. *IEEE Transactions on Software Engineering*.

[Li et al. 2015] Li, Y., Guzman, E., Tsiamoura, K., Schneider, F. and Bruegge, B. (2015) Automated Requirements Extraction for Scientific Software. *Procedia Computer Science*. 2015(51): 582-591, International Conference on Computational Science.

[Makowska 2018] Makowska, Aleksandra. Research Articles as a Means of Communicating Science: Polish and Global Conventions (2018) In *University Writing in Central and Eastern Europe: Tradition, Transition, and Innovation*. Ed. Chitez, Mădălina, Doroholschi, Claudia Ioana, Kruse, Otto Salski, Łukasz and Tucan, Dumitru. Cham: Springer.

[Malta et al. 2010] Malta, M., Cardoso, L., Bastos, F., Magnanini, M. and Silva, C. (2010) Iniciativa STROBE: subsídios para a comunicação de estudos observacionais. *Revista De Saúde Pública*. 2010(44): 559-565.

[Moher et al. 2009] Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*. 2009(151), pp. 264-269.

[Moher et al. 2012] Moher, D., Hopewell, S., Schulz, K., Montori, V., Gøtzsche, P., Devereaux, P., Elbourne, D., Egger, M. and Altman, D. (2012) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*. 2012(10): 28-55.

[Mourão et al. 2024] Mourão, Erica; Trevisan, Daniela; Viterbo, José; Pantoja, Carlos Eduardo (2024) A Systematic Mapping Study on the Use and Development of Research Software. *Information Technology and Systems*. Springer Nature Switzerland, Cham, pp. 129–138. ISBN: 978-3-031-54256-5.

[Munafo et al. 2017] Munafò, M., Nosek, B., Bishop, D., Button, K., Chambers, C., Du Sert, N., Simonsohn, U., Wagenmakers, E., Ware, J. and Ioannidis, J. (2017) A manifesto for reproducible science. *Nature Human Behaviour*. **1**, 0021.

[Nguyen-Hoan 2010] Nguyen-Hoan, L., Flint, S. and Sankaranarayana, R. (2010) A Survey of Scientific Software Development. *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*.

[Nistala 2022] Nistala, P., Rajbhoj, A., Kulkarni, V., Soni, S., Nori, K. and Reddy, R. (2022) Towards digitalization of requirements: generating context-sensitive user stories from diverse specifications. *Automated Software Engineering*, 29(1):26.

[Nistala et al. 2020] Nistala, P., Rajbhoj, A., Kulkarni, V. and Nori, K. (2020) Digital Re-Imagination of Software and Systems Processes for Quality Engineering: ISPIN Approach. *Proceedings of the International Conference on Software and System Processes*. pp. 21-30.

[Osorio et al 2020] Osorio, Alessander; Cavalheiro, Gerson G. H. (2020) *A Guidelines Proposal For Computer Science Research*. PPGC/UFPel. Master Thesis. In Portuguese.

[Osorio et al. 2020] Osorio, A., Dias, M. and Cavalheiro, G. (2020) Tangible Assets to Improve Research Quality: A Meta Analysis Case Study. *High Performance Computing Systems*. pp. 117-132

[Osorio et al 2024] Osorio, Alessander; Ferreira, Paulo; Cavalheiro, Gerson G.H. (2024) Computer Metascience: A Study on Research Taxonomy. *PPG /UFPel*.

[Pawlik and Petre 2012] Pawlik, A., Segal, J. and Petre, M. (2012) Documentation practices in scientific software development. *2012 5th International Workshop on Co-operative and Human Aspects of Software Engineering (CHASE)*. pp. 113-119

[Peffers et al. 2012] Peffers, K. and Others (2012) Design science research evaluation. *International Conference on Design Science Research in Information Systems*. pp. 398-410.

[Peng et al. 2022] Peng, Z., Liu, Y., Zhou, H., Xu, Z. and Ma, X. (2022) CReBot: Exploring interactive question prompts for critical paper reading. *International Journal of Human-Computer Studies*. **167** pp. 102898.

[Rajbhoj et al. 2023] Rajbhoj, A., Nistala, P., Kulkarni, V., Soni, S. and Pathan, A. (2023) DocToModel: Automated Authoring of Models from Diverse Requirements Specification Documents. *2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. pp. 199-210.

[Runeson 2009] Runeson, P. and Höst, M. (2009) Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*. 14, 131.

[Sarawagi 2008] Sarawagi, S. Information Extraction. *Found. Trends Databases*. 1, 261-377 (2008), https://doi.org/10.1561/1900000003

[Schmidberger and Brugge 2012] Schmidberger, M. and Brugge, B. (2012) Need of Software Engineering Methods for High Performance Computing Applications. *Proceedings of The 2012 11th International Symposium on Parallel and Distributed Computing*. pp. 40-46, https://doi.org/10.1109/ISPDC.2012.14

[Souza 2009] Souza, R. (2009) O que é um estudo clínico randomizado? *Medicina (Ribeirão Preto. Online)*. 42, 3-8. In Portuguese.

[Sun 2020] Sun, T. ())2006) Active versus passive reading: how to read scientific papers?. *National Science Review*. **7**, 1422-1427 (2020,6), https://doi.org/10.1093/nsr/nwaa130

[Tedre 2014] Tedre, M. and Moisseinen, N. (2014) Experiments in computing: A survey. *The Scientific World Journal* 2014(549398): 1–11.

[Tichy 1995] Tichy, W., Lukowicz, P., Prechelt, L. and Heinz, E. (1995) Experimental evaluation in Computer Science: A quantitative study. *Journal of Systems and Software*. 28, 9-18.

[Vrhovec et al 2020] Vrhovec S, Fujs D, Jelovčan L, Mihelič A (2020) Evaluating Case Study and Action Research Reports: Real-world Research in Cybersecurity. *Journal of Universal Computer Science* 26(7): 827-853.

[Wainer et al. 2007] Wainer, J. et al. (2007) Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação. *Atualização Em Informática*. 1 pp. 221-262. In Portuguese.

[Wainer et al. 2009] Wainer, J., Barsottini, C., Lacerda, D. and Marco, L. (2009) Empirical evaluation in Computer Science research published by ACM. *Information and Software Technology*. 51, 1081-1085.

[Wazlawick 2017] Wazlawick, R. (2017) *Metodologia de pesquisa para Ciência da Computação*. Elsevier Brasil. In Portuguese.

[Wyrich 2023] Wyrich, M. and Wagner, S. (2023) Teaching Computer Science Students to Communicate Scientific Findings More Effectively. In *International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, pp. 107-114.

[Yamamoto 2021] Yamamoto, S., Suzuki, R., Kataoka, H. and Morishima, S. (2021) Comprehending Research Article in Minutes: A User Study of Reading Computer Generated Summary for Young Researchers. *Human Interface and The Management of Information. Information Presentation and Visualization – Thematic Area, HIMI 2021, held as part of the 23rd HCI International Conference, HCII 2021, Proceedings*. pp. 101-112.

[Yamashita 2015] Yamashita, A. (2015) Integration of SE Research and Industry: Reflections, Theories and Illustrative Example. *Proceedings of the Second International Workshop on Software Engineering Research and Industrial Practice*. pp. 11-17.

[Yuan et al. 2023] Yuan, K., Lin, H., Cao, S., Peng, Z., Guo, Q. and Ma, X. (2023) CriTrainer: An Adaptive Training Tool for Critical Paper Reading. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/3586183.3606816

[Zielczynski 2007] Zielczynski, P. *Requirements Management Using IBM Rational RequisitePro*. IBM Press.