

## Book Chapter

# Evidence-Based Software Engineering: A Checklist-Based Approach to Assess the Abstracts of Reviews Self-Identifying as Systematic Reviews

Alvine Boaye Belle\* and Yixi Zhao

Lassonde School of Engineering, York University, Canada

**\*Corresponding Author:** Alvine Boaye Belle, Lassonde School of Engineering, York University, 4700 Keele St, Toronto, ON M3J 1P3, Canada

Published **December 13, 2022**

This Book Chapter is a republication of an article published by Alvine Boaye Belle and Yixi Zhao at Applied Sciences in September 2022. (Boaye Belle, A.; Zhao, Y. Evidence-Based Software Engineering: A Checklist-Based Approach to Assess the Abstracts of Reviews Self-Identifying as Systematic Reviews. Appl. Sci. 2022, 12, 9017. <https://doi.org/10.3390/app12189017>)

**How to cite this book chapter:** Alvine Boaye Belle, Yixi Zhao. Evidence-Based Software Engineering: A Checklist-Based Approach to Assess the Abstracts of Reviews Self-Identifying as Systematic Reviews. In: Prime Archives in Applied Sciences. Hyderabad, India: Vide Leaf. 2022.

© The Author(s) 2022. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Author Contributions:** A.B.B., conceptualization, investigation, methodology, supervision, software, validation, visualization and writing-original draft; Y.Z., investigation, visualization, validation and writing-original draft. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the LURA (Lassonde undergraduate research Award) program. The grant number is 1. The APC (Article Processing Charges) were funded by the start-up grant of the lead author.

**Conflicts of Interest:** The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

A systematic review allows synthesizing the state of knowledge related to a clearly formulated research question as well as understanding the correlations between exposures and outcomes. A systematic review usually leverages explicit, reproducible, and systematic methods that allow reducing the potential bias that may arise when conducting a review. When properly conducted, a systematic review yields reliable findings from which conclusions and decisions can be made. Systematic reviews are increasingly popular and have several stakeholders to whom they allow making recommendations on how to act based on the review findings. They also help support future research prioritization. A systematic review usually has several components. The abstract is one of the most important parts of a review because it usually reflects the content of the review. It may be the only part of the review read by most readers when forming an opinion on a given topic. It may help more motivated readers decide whether the review is worth reading or not. But abstracts are sometimes poorly written and may, therefore, give a misleading and even harmful picture of the review's contents. To assess the extent to which a review's abstract is well constructed, we used a checklist-based approach to propose a measure that allows quantifying the systematicity of review abstracts i.e., the extent to which they exhibit good reporting

quality. Experiments conducted on 151 reviews published in the software engineering field showed that the abstracts of these reviews had suboptimal systematicity.

## Keywords

Systematic Reviews; Reporting Guideline Adherence; PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Statement; Systematicity; Evidence-Based Software Engineering

## Introduction

Systematic reviews are well established in the software engineering field [1]. A systematic review relies on explicit and systematic methods to collect and synthesize findings of studies that focus on a given research question [2]. It provides high-quality evidence-based syntheses for efficacy under real-world conditions and allows understanding the correlations between exposures and outcomes [3]. A systematic review also allows making recommendations on how to act based on the review findings and help support future research prioritization [4,5]. The core features of a methodologically sound systematic review include transparency, replicability, and reliance on clear eligibility criteria [6]. Systematic reviews are usually called *secondary studies*, and the documents they synthesize are usually called *primary studies* [1]. Systematic reviews are usually resource intensive, e.g., time-consuming, as they usually require that a team of systematic reviewers, i.e., the authors of the review, work over a long period of time (approximately 67.3 weeks) to complete the review [6].

Systematic reviews are crucial for various stakeholders because they allow them to make evidence-based decisions without being overwhelmed by a large volume of research [4,7–10]. These stakeholders include patients, healthcare providers, policy makers, educators, students, insurers, systematic groups/organizations, researchers, guideline developers, journal editors, and evidence users [4,9,10]. Systematic reviews are increasingly popular and are conducted in various areas such as

health care, social sciences, education, and computer science [2,8].

Systematic reviews that are clearly and completely reported allow the reader to understand how they were planned and conducted, to judge the extent to which the results are reliable, to judge the reproducibility and appropriateness of the methods, and to rely on the evidence to make informed recommendations for practice or policy [11,8]. However, systematic reviews are sometimes poorly conducted and reported, exhibit low methodological rigor and reporting quality, and may result in misleading evidence-based decision-making and in the waste of various (non)financial resources [12,13]. To tackle these issues, reporting guidelines have been developed to support various tasks such as review reporting and assessment.

The most widespread reporting guidelines usually consist of a statement paper describing how the guideline has been developed and presenting the guideline in the form of a checklist that often comprises up to thirty items that need to be reported when writing a systematic review [8]. When reporting a review, it is crucial to follow such guidelines to ensure that the review is transparent, complete, beneficial, trustworthy, and unbiased [14]. According to the EQUATOR Network registry, there are at least 529 registered reporting guidelines as of August 2022. These reporting guidelines include CONSORT, STRICTA, ARRIVE and PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [15,16]. In particular, PRISMA 2020 (the most recent version of PRISMA) was released in 2021 and comprises two checklists: the PRISMA 2020 checklist that consists of 27 items and the PRISMA 2020 for abstracts checklist that consists of 12 items. PRISMA's recommendations have been widely adopted by several areas including the health and medical area, social sciences, education, and computer science [2]. As of August 2020, PRISMA yielded citations in more than sixty thousand reports [4]. Besides, PRISMA has been endorsed by more than two hundred journals as well as systematic review organizations [4]. In this work, we focus on the PRISMA reporting guideline as it yields a more

concise and transparent reporting than other reporting guidelines [17] and is suitable for the systematic reviews reporting.

Several approaches have been proposed to assess the extent to which studies have adhered to reporting guidelines, such as CONSORT, COREQ, TRIPOD, STROBE, PRISMA, and STRICTA. Thabane et al. [10], Whiting et al. [18], Shea et al. [19,20], Pussegoda et al. [12], Jin et al. [21], Blanco et al. [22], Heus et al. [23], Walsh et al. [24], Fernández-Jané et al. [15], Logullo et al. [8], and Sargeant et al. [5] discussed such approaches. Some of these approaches focus on the assessment of the adherence of systematic reviews to reporting guidelines and/or on the assessment of the risk of bias in systematic reviews. Most of these approaches are either qualitative or quantitative. Qualitative approaches usually rely on ratings scales (e.g., low, medium) to assess the adherence to reporting guidelines. Quantitative approaches often rely on frequencies and/or percentages to assess that adherence. Most of the work that has assessed the methodological rigor and the reporting quality of systematic reviews has concluded that adherence to existing reporting guidelines is usually suboptimal [13,15,21,22]. However, none of that work is suitable for the assessment of existing reviews published in the software engineering field. This calls for the proposal of new approaches to assess existing reviews and provide to stakeholders of systematic reviews a quantitative indicator that will allow them to determine whether the review at hand is sufficiently systematic and that its findings are therefore credible enough to support decision-making.

As Page et al. [2] and Andrade [25] pointed out, the abstract is one of the most important components of a review as it usually synthesizes the content of a review. It may be the only part of the review that most readers will read when forming an opinion on a given topic. It may help readers decide whether the review content is worth reading or not. However, abstracts are sometimes poorly written and may therefore give a misleading and even harmful picture of the review's contents. To assess the reporting quality of review abstracts published in the software engineering field, we proceeded as follows: (1) we introduce the

Check-SE 2022 checklist for abstracts, an eleven-item question-based checklist that is derived from the PRISMA 2020 for abstracts checklist [26], and that embodies the best reporting guidelines and/or procedures the abstract of a review published in the SE (software engineering) field should comply with to exhibit good reporting quality; and (2) we introduce a measure that allows assessing compliance with that checklist. That measure quantifies the systematicity of review abstracts by assessing their reporting quality. Experiments conducted on published reviews claiming to be systematic shows that the systematicity of their abstracts is suboptimal. More efforts should therefore be made to increase the systematicity of reviews, starting from their abstracts. We further describe our work in the remainder of this paper.

## **Background and Related Work**

In this section, we describe some concepts that are useful to understand the remainder of this paper.

### **Overview of Systematic Reviews**

Moher et al. [9,27] as well as Page et al. [2] defined a systematic review as a “review that uses explicit, systematic methods to collate and synthesize findings of studies that address a clearly formulated question”. Thus, a systematic review focuses on a clearly formulated research question and relies on systematic, reproducible, and explicit methods to support the identification, selection, and critical appraisal of relevant research meeting some eligibility criteria as well as the collection and analysis of data retrieved from that research. A systematic review also relies on explicit, reproducible, and systematic methods that allow reducing the potential bias that may arise when identifying, selecting, synthesizing, and summarizing studies. As such, a systematic review allows synthesizing the state of knowledge of a given topic [10]. A systematic review that focuses on evidence coming from the analysis of the literature deemed relevant for a given research question/topic area is referred to as systematic literature review [26]. A systematic mapping study is a form of systematic review that does not aim at synthesizing the

literature, but rather focuses on categorizing primary studies against a given framework/model, to identify the relevant research that has been undertaken in an area and to identify potential gaps in that research area [1].

Page et al. [4,11], Thabane et al. [10] and Moher et al. [9] stated that a systematic review can have several stakeholders (end-users) for which it yields different types of knowledge. Each of these stakeholders has specific interests in the systematic review and needs to know the actions they must perform based on the findings of a given review of a set of interventions. As Page et al. [2], Logullo et al. [8], Shea et al. [20], and Nair et al. [28] pointed out, systematic reviews are crucial to such stakeholders for several reasons including the following ones:

- By synthesizing the state of knowledge in a given field, systematic reviews allow identifying future research priorities, challenges, and directions without overwhelming the reader with a high volume of research.
- Systematic reviews serve as a basis of major practice and policy decisions and can address research questions for which individual studies have no answer.
- Unlike ad hoc search, systematic reviews allow minimizing subjectivity and bias, and therefore provide a higher degree of confidence regarding their coverage of the relevant studies included in the review and the credibility of their findings.
- Systematic reviews allow evaluating the trustworthiness as well as the applicability of the results they synthesize, and allow replicating or updating reviews.

To plan, conduct and report a systematic review, systematic reviewers can rely on a reporting guideline. A reporting guideline is “*a checklist, flow diagram, or explicit text to guide authors in reporting a specific type of research, developed using explicit methodology*” [29]. When systematic reviewers rely on up-to-date reporting guidelines, it helps them report a review in a transparent, complete, and accurate way, and therefore ensures that the review they are conducting is valuable to its users [2]. Besides, completing systematic reviews can be a very daunting

task that a reporting guideline can help alleviate by providing rigorous methodological guidance [30].

Based on the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network registry (<https://www.equator-network.org/library/reporting-guidelines-underdevelopment/> (accessed on 14 August 2022)), there are currently more than 529 reporting guidelines. Some of these guidelines, e.g., PRISMA, can be used to conduct and write systematic reviews. The PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guideline yields a more concise and transparent reporting than other guidelines [17].

The PRISMA reporting guideline is the result of a consensus of experts in systematic reviews [5]. As Page et al. [2,4,11] further explained, the PRISMA statement is a minimalistic set of evidence-based items that systematic reviewers can use to report systematic reviews as well as meta-analyses. The PRISMA guideline provides recommendations that help address poor writing of systematic reviews by helping systematic reviewers to transparently, completely and accurately report the rationale for conducting the review, what they did to complete the review, and their findings.

Page et al. [2,4,11] stated that PRISMA was initially released in 2009 as a 27-item checklist, and it was then referred to as PRISMA 2009. Since then, PRISMA's recommendations have been widely adopted by various areas such as the health and medical area, social science, education, and computer science. Besides, PRISMA has been endorsed by more than two hundred journals as well as systematic review organizations. As of August 2020, PRISMA has yielded citations in more than sixty thousand reports. Page et al. [2,4,11] and Veroniki et al. [31] further explained that PRISMA has evolved over time and its updated guidelines embody recent methodological and technological advances, review processes related terminology, and publishing landscape opportunities.

The PRISMA 2020 statement is the most recent release of PRISMA, and it replaces PRISMA 2009 that should no longer



be used. The PRISMA 2020 refines the structure as well as the presentation of the PRISMA 2009 checklist and includes new reporting guidance that embodies the recent progress about systematic review methodology and terminology. The PRISMA 2020 comprises a 27-item checklist focusing on reviews reporting as well as a 12-item checklist focusing on the reporting of reviews' abstracts. To properly report a systematic review, systematic reviewers should rely on both checklists. The PRISMA 2020 statement 27-item checklist provides some recommendations on how to report the following sections of a systematic review: the title (1 item), abstract (1 item), introduction (2 items), methods (11 items), results (7 items), discussion (1 item), and other information (4 items). Table A1 in the Appendix A shows the PRISMA 2020 12-item checklist for abstracts. Table A2 in the Appendix A shows an excerpt adapted from the PRISMA 2020 27-item checklist. The corresponding 12 abstract items are associated to the item 2, i.e., the abstract item, of the PRISMA 2020 27-item checklist. Page et al. [2,4] provided a description of both checklists together with their original phrasing.

The literature also describes several reporting procedures that can be used to conduct and write systematic (literature) reviews. They are notably discussed in Kitchenham [32], Petticrew and Roberts [33], Kitchenham et al. [26,34], Pickering and Byrne [35], Wohlin [36], and Okoli [30].

However, despite the availability of reporting guidelines/procedures to support clear, reproducible as well as transparent reporting, systematic reviewers are not required to adhere to them [21]. This may result in systematic reviews that lack methodological rigor, yield low-credible findings, and may mislead decision-makers [12,13]. Still, the software engineering research community has not proposed measures to assess the reporting quality of existing reviews to find out which ones are the most appropriate to support decision-making in the software engineering field.

## Related Work

Several approaches have assessed the extent to which studies have adhered to reporting guidelines. These include [5,8,10,12,13,15,18–24,37,38]. Some of these approaches focused on the assessment of the adherence of systematic reviews to reporting guidelines and/or on the assessment of the risk of bias in systematic reviews. Assessing the adherence of reviews to reporting guidelines is a means to assess the systematicity of these reviews. Most of the existing approaches are either qualitative or quantitative. Qualitative approaches usually rely on ratings scales (e.g., low, medium) to assess the adherence to reporting guidelines, while quantitative approaches usually rely on frequencies and/or percentages to assess that adherence. Several of these approaches, e.g., Thabane et al. [10], Fernández-Jané et al. [15], and Oliveira et al. [13], concluded that the adherence to reporting guidelines is suboptimal, usually vary from a checklist item to another one, and that more efforts should be made to increase that adherence and therefore yield systematic reviews of superior quality. However, none of these approaches focused on reviews published in the software engineering field.

Whiting et al. [18] implemented ROBIS (Risk of Bias in Systematic reviews), a tool that allows evaluating the level of bias in systematic reviews. The ROBIS tool mainly targets systematic reviews published in the context of health care settings, e.g., interventions, diagnosis, and prognosis. The ROBIS tool relies on three phases to assess a systematic review. Phase 1 consists in assessing the relevance of the systematic review for the investigated research question. Phase 2 consists in identifying concerns that may arise from the review process. Phase 2 focuses on biases that may be introduced into a systematic review during its completion, i.e., when studying eligibility criteria, identifying, and selecting studies, collecting data, and performing study appraisal or when synthesizing data and reporting findings. Phase 3 consists in judging the level, e.g., low, high, unclear, of the risk of bias by notably assessing the overall risk of bias in the interpretation of review results. Notably, ROBIS relies on the answers to a set of signaling

questions to complete phases 2 and 3, which can be one of the following: “yes”, “probably yes”, “probably no”, “no”, and “no information”. The answer “yes” means the concerns are low. Once the evaluation is complete, ROBIS outputs a risk rating (level).

Shea et al. [19] developed AMSTAR (A MeaSurement Tool to Assess systematic Reviews) to assess the methodological quality of systematic reviews of randomized controlled trials (RCTs). The main users of AMSTAR are health professionals, systematic reviewers, and policy makers. The AMSTAR tool implements a checklist approach and therefore relies on eleven checklist items, i.e., questions to assess systematic reviews. The answers to these questions fall in several response categories: “yes”, “no”, “cannot answer”, and “not applicable”. However, when used to assess a systematic review, AMSTAR does not output an overall score.

Shea et al. [20] developed AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews) to assess the quality of systematic reviews that include (non-)randomized studies of healthcare interventions. The AMSTAR 2 tool is an improvement of AMSTAR; it assesses all the important steps that systematic reviewers need to follow when conducting a review. The AMSTAR 2 tool is meant to allow decision-makers to distinguish high-quality reviews from low-quality ones and therefore to avoid accepting right away the findings of a single systematic review without second-guessing them. Unlike AMSTAR, AMSTAR 2 can evaluate the risk of bias in non-randomized studies included in a review. The AMSTAR 2 tool consists of sixteen questions, i.e., checklist items, whose responses can be one of the following: “yes”, “partial yes”, and “no”. The AMSTAR 2 tool outputs a rating of the overall confidence in the review results that indicates the methodological quality of the analyzed systematic review and can be one of the following: high, moderate, low, and critically low. High means that the systematic review yields either zero or one non-critical weakness.

Fernández-Jané et al. [15] relied on the STRICTA (The STAndards for Reporting Interventions in Controlled Trials of Acupuncture) reporting guidelines to assess the quality of the reporting in systematic reviews focusing on acupuncture interventions. In their work, Fernández-Jané et al. [15] focused on sixteen STRICTA checklist subitems and used them to assess twenty-eight acupuncture trials for chronic obstructive pulmonary disease (COPD). They concluded that, only four of the sixteen investigated STRICTA checklist subitems were appropriately reported in more than 70% of the trials. They also concluded that seven of the sixteen investigated STRICTA checklist subitems were correctly reported in less than 30% of the trials. They stated that the main limitation in their work was that it used STRICTA as a rating scale. Using reporting guidelines as tools to assess the quality of reporting is usually discouraged because reporting guidelines do not give guidance on how to judge/rate each of their checklist items and on when to consider each of these items as fully reported [15,5].

Sargeant et al. [5] conducted a study to evaluate the completeness of reporting in the animal health literature. They analyzed 91 systematic reviews to retrieve information on whether the reviews had reported each individual PRISMA item. Their analysis focused on three review types: systematic reviews only, systematic reviews with meta-analyses, and meta-analyses only. They leveraged the retrieved information to compute, for each review type, the percentages (proportion) of reviews for which information to address each PRISMA item was reported. These percentages allow them to conclude that the adherence to the recommended PRISMA reporting items is suboptimal and varies from an item to another. Notably, Sargeant et al. [5] performed the information retrieval step before the release of the PRISMA 2020 statement and may have missed some information related to the revised items of the PRISMA 2020 checklist.

## **An Approach to Support the Systematicity Assessment**

In this section, we introduce the notion of systematicity in the context of reviews, and propose a checklist-based approach to assess that systematicity.

### **What Is Systematicity?**

As stated above, according to Moher et al. [9,27] and Page et al. [2], a systematic review is a “review that uses explicit, systematic methods to collate and synthesize findings of studies that address a clearly formulated question”. A systematic review, therefore, focuses on a clearly formulated research question and relies on systematic, reproducible, and explicit methods to support the identification, selection, and critical appraisal of relevant research meeting some eligibility criteria as well as the collection and analysis of data retrieved from that research. A systematic review also relies on explicit, reproducible, and systematic methods that allow reducing the potential bias that may arise when identifying, selecting, synthesizing, and summarizing studies.

In that context, we define the systematicity of a review as the extent to which it is systematic, i.e., the extent to which a review exhibits good reporting quality. Based on the literature, we can conclude that the systematicity property is usually akin to several sub-properties, such as the completeness of the reporting, the transparency of the reporting, the trustworthiness in the review findings, the appropriateness of the methods used to conduct the review, the unbiasedness of the reporting, and the reproducibility of the study [2,6,8,14,29,39]. Given a review that was planned, conducted, and reported using a specific reporting guideline, we can also perceive the systematicity as the extent to which systematic reviewers who authored that review have adhered with the target reporting guideline. In this regard, Blanco et al. [22] defined adherence as the “Action(s) taken by authors to ensure that a research report is compliant with the items recommended by the appropriate/relevant reporting

guideline. These can take place before or after the first version of the manuscript is published.”

Based on the literature, e.g., Thabane et al. [10], Fernández-Jané et al. [15], Logullo et al. [8], and Oliveira et al. [13], Belur et al. [6], focusing on reporting quality and methodological rigor, we can conclude that the notion of systematicity is usually subjective, variable from one review to another and suboptimal given that each review does not usually abide by the same standards of quality and methodological rigor. The rationale is that some systematic reviewers usually prefer to report the items of reporting guidelines that seem the most important to them and to not report the remaining ones [5]. Besides, as Belur et al. [6] pointed out, conducting a systematic review usually implies making subjective decisions at various points of the review process. These decisions that are usually made by multiple systematic reviewers may hinder the reproducibility of the review if that review does not report with sufficient details how coding and screening decisions were made and how were resolved the potential disagreements that occurred between systematic reviewers involved in the review process. Relying on quantitative assessment to assess the systematicity of reviews will, therefore, foster the objective and consistent comparison of adherence to reporting guidelines across studies [23]. Besides, quantitative assessment will help decision makers to distinguish high-quality reviews from low-quality ones and therefore to avoid accepting right away the findings of a single systematic review without second-guessing their credibility [20].

Similar to the literature, e.g., Shea et al. [19,20], and Whiting et al. [18], we follow in this paper a checklist-based approach and, therefore, rely on a set of eleven items, i.e., on a set of eleven questions, to quantitatively assess the systematicity of the reviews that self-identify as “systematic reviews”. We further describe these questions below.

## A Checklist-Based Approach to Quantitatively Assess Review Abstracts

In this paper, we chose to focus on the systematicity of abstracts because the abstract is one of the most important components of a systematic review given it usually synthesizes the content of the review. In this regard, Page et al. [4] and Kitchenham et al. [26] stated that after reading the information a review abstract provides, readers should be able to decide if they want to access the full content of the review or not. Furthermore, as Andrade [25] pointed out, the abstract is usually the only part of an article, e.g., a review, that a referee reads when choosing to accept the invitation made by an editor to review a manuscript. Besides, the abstract may be the only part of a review that most readers will read when forming an opinion on a research topic.

As Page et al. [4] and Kitchenham et al. [26] pointed out, an abstract should, therefore, report the main information regarding the key objective(s) or question(s) that the review deals with. It should also specify the methods, results, and implications of the review findings. Besides, some readers, may only have access to a review abstract, e.g., in the case of journals that require a payment/subscription to access their articles. Before requesting the full content of a review, it is therefore critical for such readers to find in the abstract sufficient information about the main objective(s) or question(s) that the review addresses, methods, findings, and implications of the findings for future practice, policy, and future research. Besides, the keywords that are part of a review abstract are usually used to index the systematic review in bibliographic databases. It is, therefore, recommended to properly write abstracts so that they report keywords that accurately specify the review question, e.g., population, interventions, outcomes. However, as Andrade [25] pointed out, abstracts are sometimes poorly written, may lack crucial information, and may convey a biased and misleading overview of the review they summarize.

Similar to the literature, e.g., Shea et al. [19,20], Whiting et al. [18], to assess the systematicity of the selected review abstracts, we analyzed them using a checklist-based approach. The latter

relies on a set of eleven questions. We have inferred these questions from a thorough analysis of the literature focusing on existing reporting guidelines and procedures. In our work, we focused on:

- (a) The analysis of reporting guidelines and procedures, including the PRISMA reporting guideline that was discussed in Page et al. [2,4,11], Moher et al. [9] and Kitchenham et al. [26]. That analysis also focused on the reporting procedures discussed by Kitchenham [32], Petticrew and Roberts [33], Kitchenham [34], Pickering and Byrne [35], Wohlin [36], and Okoli [30].
- (b) The analysis of references explaining how to properly write abstracts. These include: [25], the references describing the PRISMA 2020 for abstracts checklist (e.g., [2,4]) and the reference describing revisions that Kitchenham et al. [26] made to that abstract's checklist in the context of the software engineering field.

From our analysis, we decided to mainly focus on PRISMA because it is a well-established reporting guideline that federates several existing reporting procedures used in several research areas, including computing. The PRISMA statement embeds several good reporting practices used in several disciplines including software engineering. The PRISMA 2020 checklist focuses on the adequate reporting of the whole systematic review, while the PRISMA 2020 for abstracts checklist focuses on the adequate reporting of the abstract of a review. Kitchenham et al. [26] further explained how to apply PRISMA 2020 in the software engineering field. In particular, Kitchenham et al. [26] provided supplementary explanations and examples of the SEGRESS (Software Engineering Guidelines for REporting Secondary Studies) items. These explanations of the SEGRESS items mainly focus on quantitative and qualitative systematic reviews. These guidelines derive from the PRISMA 2020 statement (and some of its extensions), the guidelines used to carry out qualitative reviews, and the guidelines used to carry out realist syntheses. Kitchenham et al. [26] revised some of the existing PRISMA 2020 explanations to make them suitable for software engineering research.



We also decided to focus on PRISMA 2020 because it has been endorsed by more than 200 journals, and it embeds more recent reporting practices and, therefore, reflects the advances in methods used to plan and conduct systematic reviews [2,4,11,31].

Our analysis allowed us to propose the Check-SE 2022 checklist for abstracts, a modified question-based version of the PRISMA 2020 checklist for abstracts. The Check-SE 2022 checklist for abstracts contains eleven items and is derived from the PRISMA 2020 for abstracts checklist and from Kitchenham et al. [26], and that applies to the reporting of reviews' abstracts published in the SE (software engineering) field. The Check-SE 2022 checklist for abstracts embodies the best reporting procedures and/or guidelines that the reviews' abstracts published in the SE (software engineering) field should comply with to exhibit good reporting quality.

Table A1 in Appendix A reports the PRISMA 2020 for abstracts checklist. Our first ten questions only consider the first ten items in the PRISMA 2020 for abstracts checklist. Thus, our questions do not take into account the 11th and the 12th item of the PRISMA 2020 for abstracts checklist, as, according to Kitchenham et al. [26], these two items do not usually apply to the software engineering field. Besides, in the software engineering community, researchers usually perform quality assessment instead of risk of bias evaluations [26] as the fifth item of the PRISMA 2020 for abstracts checklist recommends. Our fifth question does not, therefore, focus on risk of bias as the PRISMA 2020 for abstracts checklists recommends, but on the quality assessment. Besides, we have noticed that some software engineering journals usually recommend having well-formed structured abstracts by using keywords that embody the review structure, e.g., context, background, methods, results, conclusion. We, therefore, consider that recommendation as an additional reporting guideline item, and our 11th question addresses that item by attempting to determine whether existing reviews comprise structured abstracts. This is consistent with Andrade [25] who stated that journals usually require abstracts to have a formal structure consisting of the following sections:

background, methods, results, and conclusions. Some journals may recommend using additional sections, e.g., objectives, limitations, when reporting abstracts.

To assess the systematicity of a review abstract, we extracted data from the abstract to populate the form that Table 1 depicts. This allowed us to answer the eleven questions that our checklist-based approach comprises. The first row of Table 1 specifies the review reference information, e.g., names of the review authors, the publication year, and the title of the review. The second column of Table 1 shows the questions we created to assess the systematicity of reviews abstracts. The third column of Table 1 indicates the sentence(s)/keyword(s) of a reference abstract’s that allows answering a given question. The third column of Table 1 indicates the confidence, i.e., certainty, in the answers to each question, using a 3-point scale that is widely adopted in the literature (e.g., [20,1]).

**Table 1:** Question-Driven Data Extraction Form to Assess the Systematicity of Review Abstracts.

<b>Full Review Reference (APA Style)</b>			
<b>ID</b>	<b>Question</b>	<b>Review Abstract’s Sentence(s)/Keyword(s) Allowing to Answer the Question</b>	<b>Confidence of the Answer</b>
1.	Did the systematic reviewers, i.e., the authors of the review, use in the abstract the expressions “systematic review”, “systematic literature review”, “PRISMA-driven review” or a similar expression to identify the report as a systematic review?		
2.	Did the reviewers explicitly state in the abstract the main objective(s) or question(s) that the review addresses?		
3.	Did the reviewers adequately report in the abstract the inclusion and		

	exclusion criteria used in the review?		
4.	Did the reviewers adequately report in the abstract the information sources, e.g., databases, registers, used to identify studies and the date when each was last searched?		
5.	Did the reviewers mention in the abstract the methods used to perform quality assessment?		
6.	Did the reviewers specify in the abstract the methods used to present and synthesizes results?		
7.	Did the reviewers specify in the abstract the total number of included studies and participants and summarize relevant characteristics of studies?		
8.	Did the reviewers present in the abstract the results for the main outcomes, preferably indicating the number of included studies and participants for each?		
9.	Did the reviewers provide in the abstract a brief summary of the limitations of the evidence included in the review?		
10.	Did the reviewers provide in the abstract a general interpretation of the results and important implications?		
11.	Is the abstract well-structured, i.e., does it include keywords, e.g., “context”, “background”, “methods”, “results”, “conclusion”, reflecting its structure and followed		

	by a sentence(s) summarizing what has been done in the corresponding section of the review?		
<b>OConfidence (Overall confidence over 100%)</b>			

That confidence can take one of the following values: *Certain* (i.e., 1), *Partially certain* (i.e., 0.5), or *Not certain* (i.e., 0). That confidence therefore quantifies the extent to which the abstract review provides information that allows answering the so-called question.

According to Kitchenham et al. [26], the gray cells in Table 1 correspond to items in the PRISMA 2020 for abstracts checklist that are crucial in the Software Engineering field, and that every review published in this field should report. The other items (unshaded cells) seem less crucial and may not be reported, especially if the lack of space prevents systematic reviewers from reporting them. The rationale is that abstracts in software engineering journals are usually short, between 100 and 250 words. Software engineering journals therefore have more stringent limits regarding the length of abstracts than medical journals [26] for which systematic reviewers are offered more space to report all the items of the PRISMA 2020 for abstracts checklist.

In Table 1, we refer to the overall confidence in the systematicity of the abstract of a review as *OConfidence*. We compute the *OConfidence* value of a review as a percentage whose maximal value is 100% and whose minimal value is 0%. For a given review abstract  $Abs_i$ , we therefore compute the *OConfidence* as the sum of the confidence in the answer to each individual question multiplied by one hundred and divided by the total number of questions in our checklist. Equation (1) formalizes that reasoning. In that equation,  $|Questions|$  represents the number eleven, i.e., the total number questions reported in Table 1. Recall from above that for a given review abstract  $Abs_i$ , the confidence in the answer  $A_k$  provided for a given question  $Q_k$ , can be one of the following: 1 (certain), 0.5

(partially certain) or 0 (not certain). Here,  $k$  designates the identifier of a question and can vary from one to eleven.

$$OConfidence (Abs_i) = \frac{100 * \sum_{k=1}^{|Questions|} confidence(A_k, Q_k)}{|Questions|} \quad (1)$$

Similar to the literature, e.g., Weaver et al. [40], Boaye Belle et al. [41], we also used a  $n$ -level scale to interpret the values of the overall confidence based on their respective ranges. That scale varies from Invalid (i.e., 1) to Valid (i.e., 10), with  $n = 10$  in their case. We adopted that scale in a more quantitative manner, considering that the confidence and relevance are expressed in percentage and vary from 0% to 100% instead of 1 to 10. Table 2 therefore specifies the various interpretations of the *OConfidence* measure based on its value range. Table 2 maps a confidence specified in percentage to a confidence expressed using a level. The latter allows indicating to the review stakeholders the level of systematicity that a given review yields. For instance, a review’s abstract whose *OConfidence* is equal to 29% (i.e., belonging to the 20–40 range) has a Low-level confidence i.e., exhibits a low systematicity.

**Table 2:** Interpretation of the Overall Confidence, from Boaye Belle et al. [41].

<i>OConfidence</i> Range (Interval)	<i>OConfidence</i> Interpretation
[0, 1[	Invalid
[1, 5[	Extremely Low
[5, 20[	Very Low
[20, 40[	Low
[40, 60[	Medium
[60, 80[	Medium High
[80, 95[	High
[95, 99[	Very High
[99, 99.9[	Extremely High
[99.9, 100]	Valid

## Empirical Evaluation

In the remainder of this section, we present the experiments we have carried out to validate our approach.

### Dataset

We have run our experiments on a dataset consisting of reviews that self-identified as systematic reviews. To obtain these references, we searched the following well established databases that are popular in the software engineering field: Google Scholar, IEEE Xplore and ACM Digital Library. We used the following query and specifying a year range varying from 2012 and 2022:

1. **IEEE** (using the Command Search, and searching in metadata): (((*"All Metadata": "systematic review"*) OR (*"All Metadata": "PRISMA review"*)) AND (*"All Metadata": "software engineering"*))
2. **Google scholar** (using the ***Publish or Perish*** tool, and searching in the titles): (*"systematic review"* OR *"PRISMA review"*) AND (*"software engineering"*)
3. **ACM** (using the Advanced Search option and searching in the titles): (*"systematic review"* OR *"PRISMA review"*) AND (*"software engineering"*).

Searching the three databases using that query identified 267 reviews. We imported these references into Endnote (a reference manager) by creating an EndNote library of them. We then eliminated the duplicates in the so-obtained EndNote reference list. After that, we analyzed the content of the corresponding reviews and decided to include in our dataset the reviews that met the following inclusion criteria:

- It was written in English
- It was a conference paper, a journal paper, or a workshop paper
- It was peer-reviewed
- It was published between 2012 and 2022
- It self-identified as a systematic review

- It discussed software engineering concepts. If the reference deals with interdisciplinary concepts i.e., focuses on several fields, then software engineering should be part of these fields.

We did not include in our dataset the reviews that met the following exclusion criteria:

- It was the reference of a magazine, a chapter of the book, a book, conference Protocol, or a tutorial
- It was the reference of a paper that does not present results
- It was the reference of a paper that cannot be downloaded and/or is not available
- It was not related to software engineering
- It was the reference of a very short paper (i.e., a paper that has less than four pages).

Applying these inclusion and exclusion criteria allowed us to include 151 reviews in our dataset. Eight of these reviews have complied with PRISMA; 143 of them complied with other reporting procedures/methods such as the ones discussed in [32,34,42,43].

## Research Questions

Our experiments focused on the dataset described above, and aimed at answering the following research questions:

**(RQ1):** To what extent are the abstracts of reviews (that self-identify as systematic reviews) systematic? Here we rely on the answers to the eleven questions specified above to assess the systematicity of the reviews that are included in our dataset.

**(RQ2):** To what extent does the publication year influence the systematicity of the analyzed review abstracts?

**(RQ3):** Does the use of PRISMA increase the systematicity of the review i.e., their reporting quality?

**(RQ4):** What are the items that were appropriately reported in the review abstracts i.e., what are the questions to which most of the abstracts of the analyzed reviews were able to provide answers? Investigating this question will help us determine the minimum set of core abstract items to which systematic literature reviews commonly comply with in the software engineering field.

**(RQ5):** Does the systematicity of the review abstracts influence the number of citations of these reviews? The rationale is that the more a review abstract is well-reported, the more the user may feel like the review findings are credible, may be motivated to read the full content of the review to eventually exploit its results, and may decide to cite the review as a relevant related work for instance.

**(RQ6):** How do systematic reviewers usually structure reviews?

When discussing the adherence of reviews to reporting guidelines, Logullo et al. [8] stressed that some researchers estimate that a 100% adherence to the checklist items of a reporting guideline would mean that a review constitutes “a quality paper”. However, in the currently published research, it is virtually impossible to find a review that yields a 100% adherence with the target reporting guideline. Hence, that “top quality” may not be reachable and reviews adhering to, say, 80% of the items should be considered as “well reported”. A threshold equal to 80% seems to be an optimal value and involves the coverage of most items in the reporting guideline checklist. Based on that reasoning, we therefore considered that if the overall confidence in the systematicity of an analyzed review abstract is at least 80% (i.e., High), then that review abstract exhibits an optimal systematicity. Otherwise, that abstract exhibits a suboptimal systematicity.

## Data Collection and Analysis Process

To answer the six research questions specified above, we analyzed the reviews included in our dataset and, for each analyzed review abstract, we populated the question-driven data



extraction form that Table 1 reports. Similar to Fernandez-Jané et al. [15], we tested our question-based data extraction form with pilot data coming from three systematic reviews. This allows ensuring the usability of the data extraction form and helped solve potential disagreements on its understanding. We also included data of the pilot test in the final analysis. Two researchers then extracted data from the abstracts of the reviews and assessed them. More specifically, one researcher iteratively extracted all the data from the reviews included in the dataset by populating the question-driven data extraction forms (see Table 1) and specifying the individual confidence values for each answer to a given question. A more experienced researcher then validated the extracted data by double-checking the content of the question-driven data extraction forms. After the validation of the extracted data, one researcher then used Equation (1) to compute the overall confidence in the systematicity of each review's abstract.

To further answer RQ1, RQ2, RQ3, and RQ4, we created Python scripts that allowed aggregating the overall confidence values obtained for all the abstracts of the reviews included in the dataset. This allowed us to create the charts used to answer these four research questions.

To answer RQ5, we used Google Scholar to retrieve the number of citations of each review as of July 2022. We then computed the Spearman correlation between the systematicity of the review abstracts and their number of Google citations.

To answer RQ6, we relied on the answers to questions 11 that we collected from the analyzed review abstracts (when applicable). Recall from above that question 11 focused on identifying the various keywords that the systematic reviewers often use to structure their review abstracts. This allowed us to generate a word cloud depicting the most frequent keywords used to structure review abstracts.

## Results Analysis

- (1) (RQ1): Investigation of the systematicity of the review abstracts

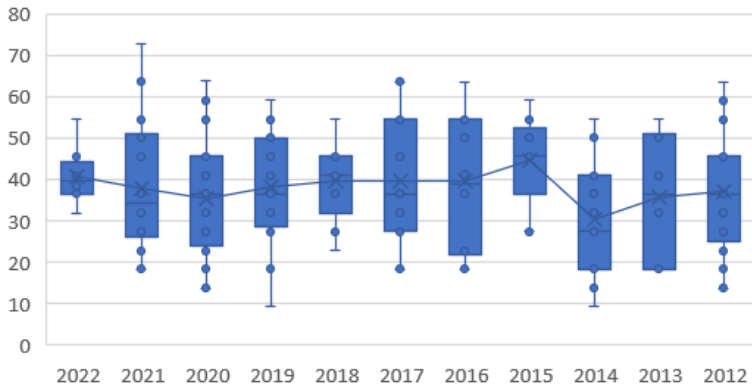
Table 3 shows the distribution of the confidence values that each review abstract yielded. All the confidence values, therefore, vary between 5 and 80 (80 excluded). Sixty-two of the review abstracts yield a confidence varying between 40 and 60 (60 excluded): these, therefore, have a Medium level confidence. Fifty-nine of the review abstracts yield a confidence varying between 20 and 40 (40 excluded): these therefore have a Low-level confidence. Only seven review abstracts have a confidence varying between 60 and 80 (80 excluded): they therefore have a Medium High confidence. None of the review abstracts has a confidence that is at least 80% (i.e., High). This allows us to conclude that all the analyzed review abstracts exhibit a suboptimal systematicity and hence, a suboptimal adherence to Check-SE 2022.

**Table 3:** Confidence of the Review Abstracts.

<i>O</i> Confidence Range (Interval)	<i>O</i> Confidence Interpretation	Number of Reviews in the Range
[0, 1]	Invalid	0
[1, 5]	Extremely Low	0
[5, 20]	Very Low	23
[20, 40]	Low	59
[40, 60]	Medium	62
[60, 80]	Medium High	7
[80, 95]	High	0
[95, 99]	Very High	0
[99, 99.9]	Extremely High	0
[99.9, 100]	Valid	0

- (2) (RQ2): Influence of the Publication Year on the Systematicity of the Reviews' Abstracts

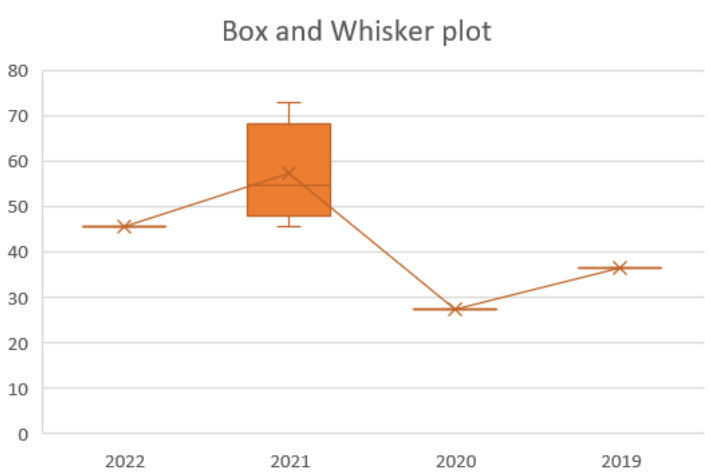
Figure 1 shows the variations of the *O*Confidence values that the analyzed review abstracts yield. It, therefore, indicates that the confidence of these reviews varies from one year to another, with highest variations scored in years 2016 and 2017.



**Figure 1:** Variation of the *O*Confidence throughout the years.

(3) (RQ3): Influence of the Well-Established Prisma Reporting Guideline on the Systematicity of the Review Abstracts

Figure 2 shows the variations of the *O*Confidence values that the analyzed PRISMA-driven review abstracts yield. Based on Figure 2, we can conclude that the PRISMA-driven review abstracts, therefore, yield relatively higher *O*Confidence values than all the analyzed review abstracts (see Figure 1). Based on the timeline that Figure 2 reports, we can also conclude that the software engineering community has recently started to use the well-established PRISMA reporting guideline, the least recent analyzed PRISMA-driven review dating back 2019.



**Figure 2:** Variation of the *OConfidence* throughout the years for the PRISMA-driven reviews.

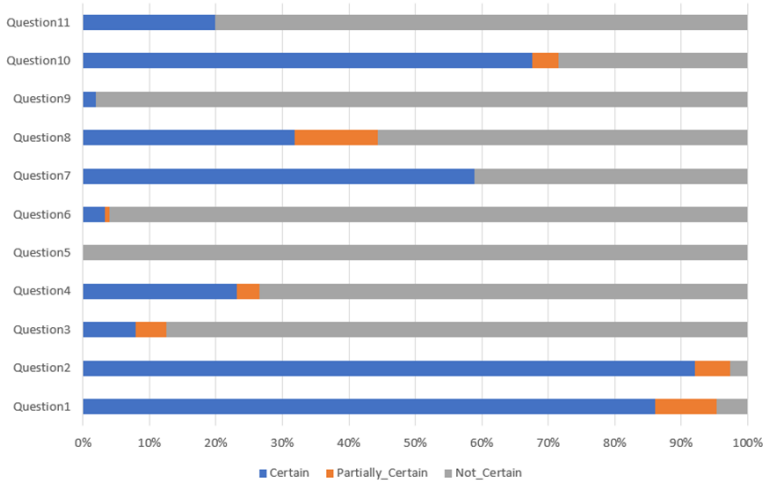
(4) (RQ4): Minimal set of items that review abstracts usually report

Figure 3 shows a summary of the answers of the eleven questions that we extracted from the analyzed review abstracts. That Figure allows us to conclude that questions 1, 2, 7, and 10 are those that most of the analyzed abstracts have appropriately answered. Hence, when writing the review abstracts, most of systematic reviewers ensured that:

- The abstract identifies the review as systematic (question 1)
- The abstract stated the main objective(s) or question(s) that the review addressed (question 2)
- The abstract specified the total number of included studies and participants and summarize relevant characteristics of studies (question 7)
- The abstract provided a general interpretation of the results as well as important implications (question 10).

As stated in Section III.B, Kitchenham et al. [26] recommended reporting at least items 2, 4, 6, 8, and 10 of the PRISMA 2020 for abstracts checklist as they are the most critical. But Figure 3 shows that existing reviews do not comply with that

requirement, as they mostly focused on answering questions 1, 2, 7, and 10 that correspond to items 1, 2, 7, and 10 in the PRISMA 2020 for abstracts checklist.



**Figure 3:** Summary of the items answered by reviews.

- (5) (RQ5): Influence of the systematicity of review’ abstracts on the number of citations of these reviews

We computed the correlation between the *OConfidence* values of the analyzed review abstracts and the number of Google citations (as of mid-July 2022). The corresponding value of the Spearman coefficient

(<https://www.socscistatistics.com/tests/spearman/default2.aspx> (accessed on 17 July 2022))  $r_s$  is 0.01235, with a  $p$ -value of 0.97685. By normal standards, the association between the two variables would not be considered statistically significant. We therefore cannot conclude whether there is a relationship between the systematicity of the review abstracts and the number of times these reviews get cited. Hence, there is no evidence that the systematicity of reviews has an impact on the influence, e.g., number of citations, of the review on the research community. Other factors may come into play and are worth investigating as future work. Such factors may include the reputation of the systematic reviewers in the research

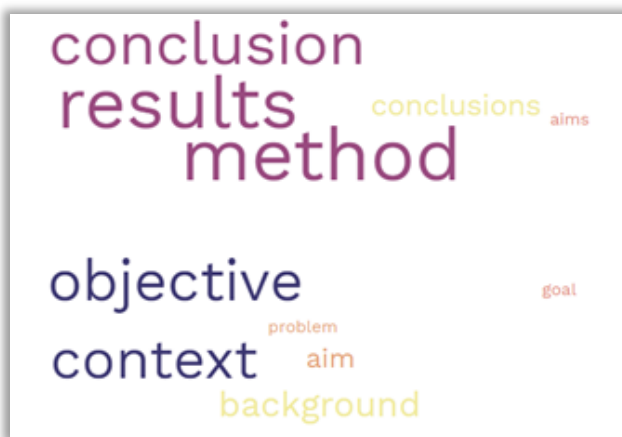
community, the importance, e.g., impact factor, conference ranking, conference acceptance rate, of the venue in which the review is published.

(6) (RQ6): On the well-formedness of the review abstracts

Thirty of the analyzed reviews had structured their abstracts by using specific keywords that mirrored the structure of the reviews they summarize. Based on the answers to question 11 that we collected from these thirty reviews, we identified the various keywords that the systematic reviewers usually used to structure their abstracts. This allowed us to generate the word cloud shown in Figure 4, which was generated using the online tool called Free Cloud Generator

(<https://www.freewordcloudgenerator.com/> (accessed on 19 July 2022)).

Based on that figure, we can conclude that the most common structure of abstracts comprised the following sections: Context, Objectives, Method, Results, and Conclusion(s). Some of abstracts also include an additional Background section to improve the well-formedness of review abstracts. These various sections are expected to comprise sufficient information to allow readers to conclude that the review is transparent, reproducible, trustworthy, unbiased, and complete enough.



**Figure 4:** Word cloud summarizing the structure of abstracts.

## Discussion

Our experiments allow us to conclude that all the analyzed review abstracts exhibit a sub-optimal systematicity. This is probably because most of the reporting procedures followed by non-PRISMA driven reviews did not necessarily put the emphasis on the reporting quality that review abstracts should exhibit. To increase that systematicity and encourage systematic reviewers to report higher quality abstracts, existing reporting procedures should be revised to encourage reviewers to create more informative and structured abstracts that properly mirror the content of the reviews. That systematicity could also be improved by using well established reporting guidelines such as PRISMA that have been specified by a panel of experts. The reviews included in our dataset indicate that PRISMA has just recently started to be used by systematic reviewers of the software engineering community. Furthermore, our experiments indicate that using PRISMA may yield a relatively higher systematicity when it comes to review abstracts among others.

Besides, providing more examples illustrating how to write good reviews and more particularly good review abstracts in computing-related disciplines, e.g., in artificial intelligence, internet of things, software engineering, could lead to the reporting of higher quality reviews. This is consistent with Kitchenham et al. [26] who proposed some examples illustrating the use of the PRISMA reporting guidelines to write good reviews in the software engineering field and have notably put the emphasis on how to properly write abstracts. This is also consistent with Okoli [30] who provided exemplars of high-quality literature reviews in the Information Systems field.

## Conclusions

To assess the extent to which a review abstract exhibits good reporting quality, we have introduced the notion of *systematicity*. The latter embodies a review's reporting quality. We have then proposed the Check-SE 2022 checklist for abstracts, an eleven-item question-based checklist for abstracts that is essentially derived from the PRISMA 2020 for abstracts checklist and

Kitchenham et al. [26]. The Check-SE 2022 checklist for abstracts embodies the best reporting guidelines and/or procedures that the abstracts of reviews published in the SE (software engineering) field should comply with to exhibit good reporting quality. We have also proposed a measure that allows assessing compliance with that checklist. That quantitative measure allows assessing the systematicity of review abstracts. Our experiments on several reviews published in the software engineering field and self-identifying as systematic reviews have shown that the abstracts of these reviews exhibited a suboptimal systematicity.

As future work, we are planning to expand our approach to assess the systematicity of the whole content of each review included in our dataset. This will help us assess the extent to which existing reviews published in the software engineering field have complied with well-established reporting guidelines such as PRISMA 2020. To support that assessment, we will propose additional questions that allow checking the adherence to the target reporting guidelines. Similar to Boaye Belle et al. [44], in future work, we will also implement an NLP (Natural language Processing)-based extraction tool that will automatically analyze systematic reviews to find answers to the proposed questions.

As future work, we will also focus on the assessment of the reviews published in the software engineering field. In this regard, we will compare the systematicity of reviews self-identifying as systematic reviews to the systematicity of reviews that do not self-identify as systematic. This will help us assess the extent to which the use of reporting guidelines/procedures to write systematic reviews increases the quality of reviews published in the software engineering field.

As we have stated above, the notion of systematicity is usually akin to several sub-properties, such as the completeness of the reporting, the transparency of the reporting, the trustworthiness in the review findings, the appropriateness of the methods used to conduct the review, the unbiasedness of the reporting, and the reproducibility of the study [2,6,8,14,29,39]. In future work, we



will carry out a systematic review (systematic mapping study) focusing on the systematicity property to further explore these sub-properties. The resulting taxonomy will help us refine our checklist-based approach to create questions that capture the various systematicity sub-properties. This may result in a more advanced confidence measure that quantifies more accurately the systematicity of reviews.

## References

1. Budgen D, Brereton P, Drummond S, Williams N. Reporting systematic reviews: Some lessons from a tertiary study. *Inf. Softw. Technol.* 2018; 95: 62–74.
2. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ.* 2021; 372: n71.
3. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J. Clin. Nurs.* 2003; 12: 77–84.
4. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ.* 2021; 372: n160.
5. Sargeant JM, Reynolds K, Winder CB, O'Connor AM. Completeness of reporting of systematic reviews in the animal health literature: A meta-research study. *Prev. Vet. Med.* 2021; 195: 105472.
6. Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociol. Methods Res.* 2021; 50: 837–865.
7. Abdul Qadir A, Mujadidi Z, Boaye Belle A. Incidium competition 2022: A preliminary systematic review centered on guidelines for reporting systematic reviews. *Appear. STEM Fellowsh. J.* 2022; 10.
8. Logullo P, MacCarthy A, Kirtley S, Collins GS. Reporting guideline checklists are not quality evaluation forms: They are guidance for writing. *Health Sci. Rep.* 2020; 3: e165.

9. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 2015; 4: 1–9.
10. Thabane L, Samaan Z, Mbuagbaw L, Kosa S, Debono VB, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J. Multidiscip. Healthc.* 2013; 6: 169.
11. Page MJ, E McKenzie J, Bossuyt PM, Boutron I, Hoffmann TC, et al. Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *J. Clin. Epidemiol.* 2021; 134: 103–112.
12. Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, et al. Systematic review adherence to methodological or reporting quality. *Syst. Rev.* 2017; 6: 1–14.
13. Oliveira NL, Botton CE, De Nardi AT, Umpierre D. Methodological quality and reporting standards in systematic reviews with meta-analysis of physical activity studies: A report from the Strengthening the Evidence in Exercise Sciences Initiative (SEES Initiative). *Syst. Rev.* 2021; 10: 1–13.
14. Bougioukas KI, Liakos A, Tsapas A, Ntzani E, Haidich AB. Preferred reporting items for overviews of systematic reviews including harms checklist: A pilot tool to be used for balanced reporting of benefits and harms. *J. Clin. Epidemiol.* 2018; 93: 9–24.
15. Fernández-Jané C, Solà-Madurell M, Yu M, Liang C, Fei Y, et al. Completeness of reporting acupuncture interventions for chronic obstructive pulmonary disease: Review of adherence to the STRICTA statement. *F1000Research.* 2020; 9: 226.
16. Wiehn E, Ricci C, Alvarez-Perea A, Perkin MR, Jones CJ, et al. Task Force ‘Adherence to reporting guidelines in articles published in EAACI Journals: A systematic review’ of the EAACI Working Group on Epidemiology. Adherence to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist in articles published in EAACI Journals: A bibliographic study. *Allergy.* 2021; 76: 3581–3588.

17. Ferdinansyah A, Purwandari B. Challenges in combining agile development and CMMI: A systematic literature review. In Proceedings of the 2021 10th International Conference on Software and Computer Applications, Kuala Lumpur, Malaysia. 2021; 63–69.
18. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J. Clin. Epidemiol.* 2016; 69: 225–234.
19. Shea BJ, Grimshaw JM, Wells AG, Boers M, Andersson N, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Med. Res. Methodol.* 2007; 7: 1–7.
20. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017; 358: j4008.
21. Jin Y, Sanger N, Shams I, Luo C, Shahid H, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years?—A systematic review of reviews: An update. *J. Multidiscip. Healthc.* 2018; 11: 495.
22. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, et al. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open.* 2019; 9: e026589.
23. Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, et al. Uniformity in measuring adherence to reporting guidelines: The example of TRIPOD for assessing completeness of reporting of prediction model studies. *BMJ Open.* 2019; 9: e025611.
24. Walsh S, Jones M, Bressington D, McKenna L, Brown E, et al. Adherence to COREQ reporting guidelines for qualitative research: A scientometric study in nursing social science. *Int. J. Qual. Methods.* 2020; 19: 1609406920982145.
25. Andrade C. How to write a good abstract for a scientific paper or conference presentation. *Indian J. Psychiatry.* 2011; 53: 172.

26. Kitchenham BA, Madeyski L, Budgen D. Segress: Software Engineering Guidelines for REporting Secondary Studies. *IEEE Trans. Softw. Eng.* 2022.
27. Moher D, Liberati A, Tetzlaff J, Altman DG, Group, Prisma. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* 2009; 6: e1000097.
28. Nair S, de la Vara JL, Sabetzadeh M, Briand L. Classification, structuring, and assessment of evidence for safety—a systematic literature review. In *Proceedings of the 2013 IEEE Sixth International Conference on Software Testing, Verification and Validation*. Luxembourg. 2013; 94–103.
29. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med.* 2010; 7: e1000217.
30. Okoli C. A guide to conducting a standalone systematic literature review. *Commun. Assoc. Inf. Syst.* 2015; 37: 43.
31. Veroniki AA, Tsokani S, Zevgiti S, Pagkalidou I, Kontouli K, et al. Do reporting guidelines have an impact? Empirical assessment of changes in reporting before and after the PRISMA extension statement for network meta-analysis. *Syst. Rev.* 2021; 10: 1–12.
32. Kitchenham B. Procedures for performing systematic reviews. Keele UK Keele Univ. 2004; 33: 1–26.
33. Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A practical Guide*. Hoboken: John Wiley & Sons. 2008.
34. Kitchenham B, Pretorius R, Budgen D, Brereton OP, Turner M, et al. Systematic literature reviews in software engineering—A tertiary study. *Inf. Softw. Technol.* 2010; 52: 792–805.
35. Pickering C, Byrne J. The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *High. Educ. Res. Dev.* 2014; 33: 534–548.
36. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on*

- Evaluation and Assessment in Software Engineering. London, UK. 2014; 1–10.
37. Yu Y, Shi Q, Zheng P, Gao L, Li H, et al. Assessment of the quality of systematic reviews on COVID-19: A comparative study of previous coronavirus outbreaks. *J. Med. Virol.* 2020; 92: 883–890.
  38. Gates M, Gates A, Duarte G, Cary M, Becker M, et al. Quality and risk of bias appraisals of systematic reviews are inconsistent across reviewers and centers. *J. Clin. Epidemiol.* 2020; 125: 9–15.
  39. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst. Rev.* 2021; 10: 1–19.
  40. Weaver R, Fenn J, Kelly T. A pragmatic approach to reasoning about the assurance of safety arguments. In *Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software*, Canberra, Australia. 2003; 33: 57–67.
  41. Boaye Belle A, Lethbridge TC, Kpodjedo S, Adesina OO, Garzón MA. A novel approach to measure confidence and uncertainty in assurance cases. In *Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. Jeju, Korea. 2019; 24-33.
  42. Kitchenham B, Charters S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Version 2.3, Technical Report EBSE 2007-01, Keele University Newcastle, UK. Durham: Durham University. 2007.
  43. Petersen K, Feldt R, Mujtaba S, Mattsson M. Systematic mapping studies in software engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, University of Bari, Italy. 2008; 1–10.
  44. Boaye Belle A, El Boussaidi G, Kpodjedo S. Combining lexical and structural information to reconstruct software layers. *Inf. Softw. Technol.* 2016; 74: 1–16.

## Appendix A: The Two PRISMA 2020 Checklists

Tables A1 and A2 respectively report the two PRISMA 2020 checklists.

**Table A1:** PRISMA 2020 12-item checklist for abstracts adapted from [2] and [4].

Abstract Section	Topic	Item Number	Abstract Checklist Item
<b>Title</b>	Title	1	In the abstract, the systematic reviewers should identify the report as a systematic review
<b>Background</b>	Objectives	2	In the abstract, systematic reviewers should explicitly state the objective(s)/question(s) the review addresses.
<b>Methods</b>	Eligibility criteria	5	In the abstract, systematic reviewers should mention the inclusion and exclusion criteria they used in the review
	Sources of information	4	In the abstract, systematic reviewers should indicate all the sources (e.g., databases, registers) they searched or consulted to identify studies and the most recent date of the search/consultation of these sources.
	Risk of bias	5	In the abstract, systematic reviewers should report the different methods they used when assessing risk of bias in the studies they included in the review
	Synthesis of results	6	In the abstract, systematic reviewers should describe the methods they used to report and synthesize their findings
<b>Results</b>	Included studies	7	In the abstract, systematic reviewers should indicate

			the total number of studies they included in the review, the participants and provide a summary of the relevant studies' characteristics
	Synthesis of results	8	In the abstract, systematic reviewers should specify the results for main outcomes, ideally specifying the number of included studies and participants for each.
<b>Discussion</b>	Limitations of evidence	9	In the abstract, systematic reviewers should make a brief summary of the limitations (e.g., study risk of bias, inconsistency) of the evidence they included in the review
	Interpretation	10	In the abstract, systematic reviewers should make an interpretation of their findings as well as significant implications
<b>Other</b>	Funding	11	In the abstract, systematic reviewers should specify the main sources of financial support they received to complete the review
	Registration	12	In the abstract, systematic reviewers should specify the review's register name and registration number

**Table A2:** An excerpt of the PRISMA 2020 27-item checklist adapted from [2] and [4].

<b>Review Section</b>	<b>Topic</b>	<b>Item Number</b>	<b>Checklist Item</b>
<b>Title</b>	Title	1	Systematic reviewers should identify the report as a systematic review
<b>Abstract</b>	Abstract	2	This item is further described through the PRISMA 2020 for Abstracts 12-items checklist (see Table 2).

<b>Introduction</b>	Rationale	3	Systematic reviewers should state the rationale for the review in the context of existing knowledge
	Objectives	4	Systematic reviewers should explicitly state the objective(s)/question(s) the review addresses.
<b>Methods</b>	Eligibility criteria	5	Systematic reviewers should mention the inclusion and exclusion criteria they used in the review and explain how they grouped studies for the syntheses.
	Sources of information	6	Systematic reviewers should indicate all the sources (e.g., databases, registers, websites, organizations, reference lists) they searched or consulted to identify studies and the most recent date of the search/consultation of these sources.
	Strategy used for the search	7	Systematic reviewers should report the full search strategies used for all databases, registers, and websites, and specify the different filters and limits used during the search.
	...	...	...
<b>Results</b>	Study selection	16a	Systematic reviewers should report the results of the search and selection process by preferably relying on a flow diagram (The flow diagram is a template that illustrates the flow of records retrieved and analyzed in the different stages of the search and selection process (Page et al. [4]))
		16b	Systematic reviewers should specify the studies seem to meet the inclusion criteria, but that they decided to exclude, and provide a



			rationale for their inclusion
	Characteristics of the study	17	Systematic reviewers should specify each included study and describe its characteristics.
	Risk of bias in studies	18	Systematic reviewers should report assessments of risk of bias for each study they included
	...	...	...
<b>Discussion</b>	...	...	...
<b>Other information</b>	Protocol and registration	24a	Systematic reviewers should specify the review's registration information (e.g., register name and registration number) or state that they did not register the review
		24b	Systematic reviewers should specify the location of the review protocol, or state that they did not prepare a protocol
		24c	Systematic reviewers should indicate any amendments they made to the information they provided at registration/in the protocol.
	Support	25	Systematic reviewers should specify sources of financial or non-financial they received to support the completion of the review, and specify role of the funders/sponsors in the completion of the review
	Competing interests	26	Systematic reviewers should state any relevant competing interests regarding the review
	Data, code, and other materials availability	27	Systematic reviewers should specify whether some additional information (e.g., template data collection forms, data extracted from included studies, data used for all analyses, analytic code) is publicly available and specify its location