Paving the Way for Mature Secondary Research: The Seven Types of Literature Review

Paul Ralph
Dalhousie University
Halifax, Canada
paulralph@dal.ca

Sebastian Baltes
University of Adelaide
Adelaide, Australia
sebastian.baltes@adelaide.edu.au

ABSTRACT

Confusion over different kinds of secondary research, and their divergent purposes, is undermining the effectiveness and usefulness of secondary studies in software engineering. This short paper therefore explains the differences between ad hoc reviews, case surveys, critical reviews, meta-analyses (aka systematic literature reviews), meta-synthesis (aka thematic analysis), rapid reviews and scoping reviews (aka systematic mapping studies). We envision that these guidelines will help researchers better select and describe their literature reviews, while helping reviewers select more appropriate evaluation criteria.

KEYWORDS

Case survey, critical review, meta-analysis, meta-synthesis, rapid review, systematic review, systematic literature review, scoping review

ACM Reference Format:
Paul Ralph and Sebastian Baltes. 2022. Paving the Way for Mature Secondary Research:
The Seven Types of Literature Review. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Scholarship and scholarly writing can be stratified into *primary*, *secondary*, and *tertiary* research.

Primary research involves making observations in the broadest sense of collecting data about objects that are not themselves studies. Computing code metrics, administering questionnaires, interviewing participants, counting bugs, collecting documents, downloading source code, and taking field notes while observing a retrospective meeting are all primary research.

Secondary research involves analyzing, synthesizing and critiquing primary studies. Ad hoc reviews, case surveys, critical reviews, meta-analysis, meta-synthesis and scoping reviews are all types of secondary research. Secondary research is central to evidence-based practice because (1) practitioners cannot read every study; (2) important decisions typically should be made based on the balance of evidence rather than a single study; (3) in many fields, individual studies are too small to produce accurate estimates of population parameters (e.g. the strength of the relationship between two variables). However, some secondary research degrades into predominately descriptive "papers about papers" with limited scope and usefulness.

Tertiary scholarship has two related meanings: (1) analyses of groups of secondary studies (meta-reviews); (2) summaries broad areas of scientific knowledge as found in textbooks, encyclopedia entries, etc.

This paper focuses on secondary research because confusion over different kinds of secondary research, and their divergent purposes, is undermining the effectiveness and usefulness of secondary studies in software engineering (SE). This observation motivates the following objective.

> ***Purpose:*** *The objective of this article is to compile, describe, compare, and contrast the different kinds of secondary research commonly published in software engineering venues, to pave the way for more mature secondary research.*

2 REVIEW TYPES

2.1 Ad Hoc Reviews

An *ad hoc* literature review is simply a discussion of some literature, as contained in most research papers as part of a background or related work section. Ad hoc reviews may develop theory [e.g. 27], or integrate a new theory into existing literature [cf. 31]. Ad hoc reviews can support a position paper, or tertiary scholarship.

Ad hoc reviews use purposive sampling [3]; that is, researchers purposefully select papers or studies that are useful, relevant, or support their arguments. Ad hoc reviews are often appropriate, for example, to support theory development, identify promising research topics, or prepare for a comprehensive exam. However, they suffer from important limitations: their unsystematic nature introduces sampling bias and defies replication. This may lead to cherry-picking of evidence supporting authors' arguments [10]. Therefore, *ad hoc reviews are inappropriate for supporting empirical statements* such as "x causes y", "most people/objects have property P", or "process P has structure S or follows rules R".

2.2 Systematic Reviews

The term *systematic review* has two meanings, as follows:

(1) a literature review that employs a systematic (hence the name), replicable process of selecting primary studies for inclusion (below referred to as *systematic review*), including case surveys, critical reviews, meta-analyses, meta-syntheses, and scoping reviews.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Table 1: The Seven Types of Literature Review

Type	Systematic	Purpose	Primary Studies	Analysis	Approach
ad hoc	no	discuss	any	any	discus purposively selected related work
case survey	yes	explain & predict	qualitative	quantitative	test causal hypothesis by aggregating case studies results
critical	yes	prescribe	any	any	defend a position and make recommendations by analyzing a sample of papers
meta-analysis	yes	explain & predict	quantitative	quantitative	estimate effect sizes by aggregating results of similar quantitative studies
meta-synthesis	yes	explain	qualitative	qualitative	synthesize the results of numerous studies using qualitative analysis
rapid	yes	explain & predict	quantitative	quantitative	a meta-analytic review that compromises rigor for speed
scoping	yes	describe	any	both	describe an area of research and maps studies into meaningful categories

(2) a systematic review that uses meta-analysis of quantitative studies (especially experiments) to assess the strength of the evidence for specific, usually causal, propositions (below referred to as *meta-analysis*).

This double meaning is due to the history of systematic reviews. The concept of a meta-analytic systematic review emerged from health and medicine in the late 20^{th} century [25], when practitioners could not keep current with the acceleration of research results. When Chalmers [6] founded the Cochrane Library in 1993, the medical community coalesced around using meta-analysis to inform evidence-based practice. Now, *systematic review* is often conflated with *meta-analysis*.

However, other kinds of systematic reviews have also been around for decades. Scoping reviews were proposed no later than 2005 [1]. Meta-synthesis goes back at least to 1996 [14]. Case surveys were proposed as far back as 1974 [18].

Systematic reviews begin by searching databases for primary studies that meet pre-established criteria. Most types of systematic reviews seek to identify *all* the primary studies that meet the selection criteria by applying various techniques to mitigate sampling bias and publication bias including: "backward and forward snowballing searches; checking profiles of prolific authors in the area; searching both formal databases (e.g. ACM Digital Library) and indexes (e.g. Google Scholar); searching for relevant dissertations; searching pre-print servers (e.g. arXiv); soliciting unpublished manuscripts through appropriate listservs or social media; contacting known authors in the area" [28].

The way that the primary studies are analyzed determines the *type* of systematic review.

2.2.1 *Meta-analysis.*

In an archetypal systematic review, researchers identify a group of randomized controlled experiments with the same hypotheses, and the same independent and dependent variables.

Suppose, for example, we have 10 experiments in which SE undergrads were randomized into a control group (who complete tasks individually) and a treatment group (who complete tasks in pairs); that is, individual vs. pair programming. In all 10 experiments, the dependent variable was the number of tasks completed successfully, and the hypothesis was that the treatment group would complete more tasks successfully. Each primary study reports the results of an independent samples t-test including the mean and standard deviation for each group, the t-statistic, the p-value, Cohen's D (effect size) and the 95% confidence interval for D.

Our aim, then, is to combine the results of these ten experiments to estimate the effect size of pair programming on effectiveness. Suppose that four of the studies found a negative effect (hypothesis not supported), three found no significant effect and three found a positive effect (hypothesis supported). We **DO NOT** proceed by vote counting; that is, assuming that the effect is negative because four studies found a negative effect vs. three finding a positive effect. Vote counting is not valid because primary studies can have wildly different sizes. What if the studies that found positive effects were much larger and more rigorous; while the studies that found negative effects were small and confounded? What if, when the three studies without significant results are aggregated, together their results are significant?

Instead of vote-counting, we apply meta-analysis [9]; that is, we use a statistical model to aggregate the results from the primary studies into a global estimate of effect size. Meta-analysis is often possible with the summary data reported in papers, without the original datasets.

Meta-analysis is not limited to randomized controlled experiments. Results from different sorts of (quantitative) methods can aggregated through meta-analysis as long as the studies have the same independent and dependent variables. Sometimes, instead of exactly the same variables, studies have overlapping sets of variables. The more complicated the overlaps, the more complicated the meta-analytic model. A comprehensive tutorial on statistical procedures for meta-analysis is beyond the scope of this paper, but readily available [cf. 4]

Meta-analytic reviews essentially have the same research question as the studies being reviewed. The purpose of the meta-analysis is to reach a more reliable, robust conclusion by aggregating all available data, implying two important criteria for meta-analysis: (1) researchers are supposed to go to great lengths to find **ALL** relevant studies; and (2) research *must* evaluate the quality of each primary study and either exclude low quality studies or include study quality as a covariate in the meta-analytic model.

In summary, when scientists equate systematic reviews with evidence-based practice, they usually mean meta-analytic reviews. Meta analysis aggregates quantitative studies that investigate the same or overlapping hypotheses. *They do not simply describe* existing research. Meta-analysis is rare in software engineering. While there are several good examples [e.g. 12, 26, 29], quality meta-analysis is dwarfed by superficial scoping reviews [7].

2.2.2 *Meta-synthesis.*

Meta-synthesis—aka thematic analysis, narrative synthesis, meta-ethnography, and interpretive synthesis—refers to a family of methods of aggregating qualitative studies [14]. Meta-synthesis is approximately the constructivist analogue of meta-analysis. After identifying the primary studies, the researcher applies hermeneutical and dialectical analyses to understand each primary study, translate them into each other, and construct an account of the body of research; for example, a theory of the central phenomenon that unites the primary studies.

Meta-synthesis requires expertise in qualitative methods and some understanding of the underlying philosophical assumptions. If one does not know what "hermeneutical and dialectical analyses" means, one should not attempt meta-synthesis. Meta-synthesis is **NOT** organizing papers into categories (as we see in scoping reviews). Meta-synthesis is synthesizing a credible, nuanced account of a phenomenon, grounded in a body of interconnected qualitative research.

In principle, meta-synthesis can be applied to both qualitative and quantitative work. In practice, such combinations are philosophically strained.

2.2.3 Case Survey (aka Case Meta-analysis). In a case survey, the primary studies are (typically qualitative) case studies in the broadest sense of a case study (i.e., a scholarly account of some events). Experience reports and grey literature may or may not be included, depending on the study's purposes. However, unlike meta-synthesis, a case survey transforms qualitative accounts into a quantitative dataset that supports null-hypothesis testing. Case studies share the philosophy of meta-analysis (positivist), not meta-synthesis (constructivism). Case surveys typically begin with a priori hypotheses and an a priori coding scheme. The researcher reads each case and extracts data into the coding scheme, often using simple dichotomous variables like 'did the team have retrospective meetings? [yes/no]' or 'does the case mention coordination problems [yes/no]?' This resulting dataset is often too sparse for regression modeling, so researchers use simple bivariate correlations to test hypotheses [5]. Case surveys were proposed by the Rand Corporation as a "way to aggregate existing research" [18], quickly picked up by Yin [32], and later elaborated in management [5, 17]. Now case meta-analysis is widely used in management and information systems [15]. While SE-specfic case survey guidelines are available [e.g. 19, 22], SE case surveys remain rare. However, case surveys have been used to investigate strategic pivots at software start-ups [2] and how organizations select component sourcing options [23]. We see great potential for case surveys in SE research because case studies are so common.

2.2.4 Critical Reviews. In a *critical review*, researchers analyze a sample of primary studies to support an argument or critique.[1] For example, Stol et al. [31]'s critical review of the use of grounded theory in software engineering *criticizes* method slurring; that is, claiming to have used a research methodology that was not actually used to create the illusion of legitimacy. Similarly, Baltes and Ralph [3] used a critical review to describe and criticize how SE researchers overstate sample representativeness and often conflate random sampling with representative sampling. Indeed, critical reviews in software engineering often investigate methodological topics, including how ethnography is reported [33] or how qualitative research is synthesized [13].

Critical reviews differ from case surveys and meta-analysis in two important ways. First, meta-analytic reviews aggregate evidence regarding causal relationships to generate evidence-based recommendations, whereas a critical review critically evaluates issues. Critical reviews are not for supporting evidence-based practice, or summarizing the evidence for a theory. Critical reviews are part of the meta-scientific discourse; i.e., the conversations a scientific community has internally about how it conducts research.

Second, most systematic reviews aims to collect *all* relevant primary studies on a specific topic [16]. Including all relevant studies for critical reviews is often impossible. A critical review of adherence to the *Introduction, Method, Results and Discussion framework* (IMRaD) framework [30], for example, could include all SE papers, ever. However, this is unnecessary. A random sample of papers from a selection of top journals and conferences is sufficient to assess compliance with IMRaD. For meta-analysis, we worry about publication bias. Critical reviews do not assess causal claims so worrying "what if significant results were published but non-significant results were not?" is irrelevant.

Analysis performed within a critical review can be quantitative, qualitative or both. However, critical reviews typically adopt a *critical* stance; that is, they go beyond mere description and offer specific critiques of the work being reviewed.

2.2.5 Scoping Reviews (aka Systematic Mapping Studies). What is often called a "systematic mapping study" in software engineering [24] is usually called a *scoping review* elsewhere (e.g. in health, medicine, and psychology). The purpose of a scoping review is to understand the status of research on a particular topic, typically by mapping primary studies into categories (hence, "mapping study").

Scoping reviews are often primarily descriptive. They count the number of studies on a topic. They often organize studies by research method, subtopic, authors, geographical location, publication venue, etc. They often conclude that more research is needed in this or that subtopic. For example, Mohanani et al. [20] mapped primary studies according to which cognitive bias (e.g. confirmation bias) they investigated and in which area of software development (e.g. design, management) they investigated it, then called for more research on *debiasing* (preventing or mitigating cognitive biases).

The problem with scoping reviews is that they typically fail to fulfill any of the core purposes of systematic reviews:

Meta-analysis and *case surveys* synthesize the results of many studies to answer specific empirical (often causal) questions about the world. Scoping reviews include a similar search but typically do not provide sufficient quantitative synthesis to answer important empirical questions. Therefore, scoping reviews do not inform evidence-based practice like meta-analyses and case surveys.

Meta-synthesis involves deep, theory-oriented reinterpretation of related qualitative studies. While scoping reviews often include qualitative analysis (e.g. mapping or categorization), that analysis is often too superficial to generate novel, useful theories.

Critical reviews use a sample of papers to demonstrate an important pattern for a scientific community's internal discourse (e.g. widespread construct validity problems). While scoping reviews often give recommendations regarding future research, they focus on an empirical topic (e.g. cognitive biases in SE); not a meta-scientific topic (e.g. construct validity); therefore, they are not configured, from the outset, to deliver useful meta-scientific critique.

In summary, scoping reviews begin like other kinds of systematic reviews, but stop short of *synthesizing* the data into aggregate

[1] The term *critical review* has been used differently elsewere, but we focus on the meaning in SE.

empirical results, theory, or meta-scientific critique. This is by definition: if a scoping review applies a meta-analytic model to aggregate primary study results, or applies hermeneutics and dialectics to synthesize qualitative accounts, or develops an evidence-based critique of a scientific practice, it is no longer a scoping review; it is a meta-analysis, a case survey, a meta-synthesis or a critical review. Some authors therefore recommend a scoping review "as a precursor to a systematic review" [21].

2.2.6 Rapid Reviews. A *rapid review* is a meta-analysis that makes methodological compromises to reduce completion time [8]. Ganann et al. found many such compromises including restricting the literature search, truncating results, omitting techniques for overcoming publication bias (e.g. reference snowballing); streamlining screening and data extraction, and skipping quality assessment [8].

Rapid reviews are justified *if and only if* evidence is needed to support imminent decisions, and waiting for a comprehensive meta-analytic review would be harmful. These conditions occur in health and medicine, for example, when an unprecedented viral pandemic strikes. These conditions do not occur frequently in SE. The term *rapid review* should not be used to legitimize doing a bad job of a systematic review where there is no pressing need for immediate results. Moreover, SE-related topics rarely have so many primary studies that it would take more than a year to complete a comprehensive review.

3 DISCUSSION

Secondary research that synthesizes the results of a body of work into robust, reliable, findings and practical ramifications is crucial for evidence-based practice. Secondary research is the funnel that concentrates the overwhelming flow of empirical results into something practitioners can digest and use. However, most secondary research in SE contains no synthesis of findings [7], because most SE secondary studies are scoping reviews. Scoping reviews struggle to get from a broad description of what is happening in a field to specific recommendations for practitioners. Whereas the purpose of meta-analysis is to answer a specific empirical question by combining the results of all of the studies that tested corresponding hypotheses, scoping reviews just put research in buckets. Meanwhile, SE researchers perceive inadequate number and quality of primary studies as a major barrier to good secondary research [11].

3.1 Recommendations

Our main recommendation is: **software engineering research needs more meta-analysis, more case surveys, more meta-synthesis, and fewer scoping and rapid reviews.**

If there is a sufficient body of evidence to facilitate a meta-analysis, case survey, or meta-synthesis, by all means, do one. Top venues should welcome competent meta-analyses, case surveys, and meta-syntheses as these approaches are central to evidence-based practice. In contrast, predominately descriptive scoping reviews are essentially works-in-progress. They belong in workshops, poster sessions and short paper tracks. If there is not enough evidence for a meta-analysis, case survey, or meta-synthesis, then a scoping review will not be able to make a substantive contribution to the scholarly discourse. If there is enough evidence, the scoping review is like a pilot study for a more ambitious systematic review.

Rapid reviews similarly have little place in SE. The conditions that justify a rapid review rarely occur in our field. Top venues should expect rapid reviews to justify methodological compromises and explain why results are needed immediately. The term "rapid review" should not be used to legitimize a bad meta-analysis.

The case for critical reviews is more complicated. SE does not need more critical reviews per se; SE needs a more vivid, robust meta-scientific discourse. Meta-science in SE has been hamstrung by the insistence, among many reviewers and top venues, that every paper reports an empirical study, regardless of whether the questions at hand are empirical. As long as this unwise position remains, we need critical reviews because they provide a mechanism to inject meta-science into our publications. We would also like to remind reviewers not to misapply the "find ALL relevant studies" or "exclude low quality studies" criteria to critical reviews.

Finally, authors should identify the kind of "systematic review" they have done. Avoid using "systematic review" as a synonym for meta-analysis or, worse, scoping review. In the broadest sense, any review that uses a systematic process of identifying primary studies is a systematic review.

3.2 Limitations

This is not an empirical paper. This paper discusses a methodological issue. It is meta-science: part of our community's internal discourse about what we are doing, why, and how to do it. One major reason SE has so many methodological challenges is that we have effectively abandoned our meta-scientific discourse by expecting every paper to report an empirical study. This paper is credible because it meets the expectations laid out in the Empirical Standard for Meta-science: it is useful because most SE researchers read or perform literature reviews; makes clear recommendations (Section 3.1); presents coherent arguments; "goes beyond summarizing methodological guidance from existing works... [and] provides insight specifically for software engineering" [28]. It is limited by excluding tertiary studies and providing relatively brief descriptions of each type of review to respect the venue's page limit.

4 CONCLUSION

Systematic review is often conflated with *meta-analysis*. We recommend using *systematic review* as the category of secondary research in which primary studies are selected according to a systematic, replicable process; that is, the opposite of an *ad hoc* review. Meta-analysis, case survey, meta-synthesis, critical review, scoping review and rapid review are all types of systematic review.

SE needs more meta-analysis, meta-synthesis and case surveys. These methods synthesize bodies of related work into actionable recommendations for practitioners and are therefore crucial for evidence-based practice. Scoping reviews, in contrast, do not include enough synthesis to support evidence-based practice. A scoping review is like a prelude to a more ambitious project. If published at all, scoping reviews belong with other work-in-progress in posters, workshops or short papers. Rapid reviews meanwhile, are a valuable tool but SE phenomena rarely manifest the conditions under which rapid reviews are justified. Finally, critical reviews have different best practices and reviewers should stop misapplying meta-analysis criteria to critical reviews.

We hope this paper helps researchers distinguish between different types of literature reviews—paving the way for more mature secondary research in software engineering.

REFERENCES

[1] Hilary Arksey and Lisa O'Malley. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology* 8, 1 (2005), 19–32.
[2] Sohaib Shahid Bajwa, Xiaofeng Wang, Anh Nguyen Duc, and Pekka Abrahamsson. 2017. "Failures" to be celebrated: an analysis of major pivots of software startups. *Empirical Software Engineering* 22, 5 (2017), 2373–2408.
[3] Sebastian Baltes and Paul Ralph. 2022. Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering* 27 (2022), article number 94.
[4] Michael Borenstein, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2021. *Introduction to meta-analysis.* John Wiley & Sons.
[5] RJ Bullock and Mark E Tubbs. 1987. The case meta-analysis method for OD. *Research in organizational change and development* 1 (1987), 171–228.
[6] Iain Chalmers. 1993. The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Annals of the New York Academy of Sciences* 703, 1 (1993), 156–165.
[7] Daniela S Cruzes and Tore Dybå. 2011. Research synthesis in software engineering: A tertiary study. *Information and Software Technology* 53, 5 (2011), 440–455.
[8] Rebecca Ganann, Donna Ciliska, and Helen Thomas. 2010. Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science* 5, 1 (2010), 1–10.
[9] Gene V Glass. 1976. Primary, secondary, and meta-analysis of research. *Educational researcher* 5, 10 (1976), 3–8.
[10] David Gough, Sandy Oliver, and James Thomas. 2017. *An introduction to systematic reviews.* Sage.
[11] Liliana Guzmán, Constanza Lampasona, Carolyn Seaman, and Dieter Rombach. 2014. Survey on research synthesis in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering.* 1–10.
[12] Jo E Hannay, Tore Dybå, Erik Arisholm, and Dag IK Sjøberg. 2009. The effectiveness of pair programming: A meta-analysis. *Information and software technology* 51, 7 (2009), 1110–1122.
[13] Xin Huang, He Zhang, Xin Zhou, Muhammad Ali Babar, and Song Yang. 2018. Synthesizing qualitative research in software engineering: A critical review. In *Proceedings of the 40th International Conference on Software Engineering.* 1207–1218.
[14] Louise A Jensen and Marion N Allen. 1996. Meta-synthesis of qualitative findings. *Qualitative health research* 6, 4 (1996), 553–560.
[15] Marlen C Jurisch, Petra Wolf, and Helmut Krcmar. 2013. Using the case survey method for synthesizing case study evidence in information systems research. In *Proceedings of the 19th Americas Conference on Information Systems (AMCIS).*
[16] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering.* Technical Report. Keele University and University of Durham.
[17] Rikard Larsson. 1993. Case survey methodology: Quantitative analysis of patterns across case studies. *Academy of Management Journal* 36, 6 (1993), 1515–1546.
[18] William A Lucas. 1974. *The case survey method: Aggregating case experience.* Number 1515. Rand.
[19] Jorge Melegati and Xiaofeng Wang. 2020. Case Survey Studies in Software Engineering Research. In *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM).* 1–12.
[20] Rahul Mohanani, Iflaah Salman, Burak Turhan, Pilar Rodríguez, and Paul Ralph. 2018. Cognitive biases in software engineering: a systematic mapping study. *IEEE Transactions on Software Engineering* 46, 12 (2018), 1318–1339.
[21] Zachary Munn, Micah DJ Peters, Cindy Stern, Catalin Tufanaru, Alexa McArthur, and Edoardo Aromataris. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology* 18, 1 (2018), 1–7.
[22] Kai Petersen. 2020. Guidelines for Case Survey Research in Software Engineering. In *Contemporary Empirical Methods in Software Engineering.* Springer, 63–92.
[23] Kai Petersen, Deepika Badampudi, Syed Muhammad Ali Shah, Krzysztof Wnuk, Tony Gorschek, Efi Papatheocharous, Jakob Axelsson, Severine Sentilles, Ivica Crnkovic, and Antonio Cicchetti. 2017. Choosing component origins for software intensive systems: In-house, COTS, OSS or outsourcing?—A case survey. *IEEE Transactions on Software Engineering* 44, 3 (2017), 237–261.
[24] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. Systematic Mapping Studies in Software Engineering.. In *EASE*, Vol. 8. 68–77.
[25] Edward Purssell and Niall McCrae. 2020. A Brief History of the Systematic Review. *How to Perform a Systematic Literature Review* (2020), 5–17.
[26] Yahya Rafique and Vojislav B Mišić. 2012. The effects of test-driven development on external quality and productivity: A meta-analysis. *IEEE Transactions on Software Engineering* 39, 6 (2012), 835–856.
[27] Paul Ralph. 2018. Toward methodological guidelines for process theories and taxonomies in software engineering. *IEEE Transactions on Software Engineering* 45, 7 (2018), 712–735.
[28] Paul Ralph, Sebastian Baltes, Domenico Bianculli, Yvonne Dittrich, Michael Felderer, Robert Feldt, Antonio Filieri, Carlo Alberto Furia, Daniel Graziotin, Pinjia He, et al. 2020. ACM SIGSOFT Empirical Standards. *arXiv: 2010.03525 [cs.SE]* (2020).
[29] Martin Shepperd, David Bowes, and Tracy Hall. 2014. Researcher bias: The use of machine learning in software defect prediction. *IEEE Transactions on Software Engineering* 40, 6 (2014), 603–616.
[30] Luciana B Sollaci and Mauricio G Pereira. 2004. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the medical library association* 92, 3 (2004), 364.
[31] Klaas-Jan Stol, Paul Ralph, and Brian Fitzgerald. 2016. Grounded theory in software engineering research: a critical review and guidelines. In *Proceedings of the 38th International Conference on Software Engineering.* 120–131.
[32] Robert K Yin and Karen A Heald. 1975. Using the case survey method to analyze policy studies. *Administrative science quarterly* (1975), 371–381.
[33] He Zhang, Xin Huang, Xin Zhou, Huang Huang, and Muhammad Ali Babar. 2019. Ethnographic research in software engineering: a critical review and checklist. In *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering.* 659–670.