



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Mapping the landscape: Peer review in computing education research.
Author(s)	Petre, Marian; Sanders, Kate; McCartney, Robert; Ahmadzadeh, Marzieh; Connolly, Cornelia; Hamouda, Sally; Harrington, Brian; Lumbroso, Jérémie; Maguire, Joseph; Malmi, Lauri; McGill, Monica M.; Vahrenhold, Jan
Publication Date	2020-06-17
Publication Information	Petre, Marian, Sanders, Kate, McCartney, Robert, Ahmadzadeh, Marzieh, Connolly, Cornelia, Hamouda, Sally, Harrington, Brian, Lumbroso, Jérémie, Maguire, Joseph, Malmi, Lauri, McGill, Monica M., Vahrenhold, Jan. (2020). Mapping the landscape: Peer review in computing education research. Paper presented at the ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 20), Trondheim, Norway, June 17–18, 2020.
Publisher	Association for Computing Machinery (ACM)
Link to publisher's version	https://dl.acm.org/doi/proceedings/10.1145/3341525
Item record	http://hdl.handle.net/10379/16384
DOI	http://dx.doi.org/10.1145/3437800.3439207

Downloaded 2024-04-24T08:27:00Z

Some rights reserved. For more information, please see the item record link above.



Mapping the Landscape of Peer Review in Computing Education Research

Marian Petre*
The Open University
Milton Keynes, UK
m.petre@open.ac.uk

Kate Sanders*
Rhode Island College
Providence, RI, USA
ksanders@ric.edu

Robert McCartney*
University of Connecticut
Storrs, CT, USA
robert@enr.uconn.edu

Marzieh Ahmadzadeh
York University
Toronto, ON, Canada
marzieha@yorku.ca

Cornelia Connolly
National University of Ireland
Galway, Ireland
cornelia.connolly@nuigalway.ie

Sally Hamouda
Rhode Island College
Providence, RI, USA
shamouda@ric.edu

Brian Harrington
University of Toronto Scarborough
Toronto, ON, Canada
brian.harrington@utoronto.ca

J eremie Lumbroso
Princeton University
Princeton, NJ, USA
lumbroso@cs.princeton.edu

Joseph Maguire
University of Glasgow
Glasgow, Scotland, UK
joseph.maguire@glasgow.ac.uk

Lauri Malmi
Aalto University
Finland
lauri.malmi@aalto.fi

Monica M. McGill
Knox College
CSEdResearch.org
Galesburg, IL, USA
monica@cse dresearch.org

Jan Vahrenhold
Westfalische Wilhelms-Universitat
M nster
Germany
jan.vahrenhold@uni-muenster.de

ABSTRACT

Peer review is a mainstay of academic publication – indeed, it is the peer-review process that provides much of the publications’ credibility. As the number of computing education conferences and the number of submissions increase, the need for reviewers grows. This report does *not* attempt to set standards for reviewing; rather, as a first step toward meeting the need for well qualified reviewers, it presents an overview of the ways peer review is used in various venues, both inside computing education and, for comparison, in closely-related areas outside our field. It considers four key components of peer review in some depth: criteria, the review process, roles and responsibilities, and ethics and etiquette. To do so, it draws on relevant literature, guidance and forms associated with peer review, interviews with journal editors and conference chairs, and a limited survey of the computing education research community. In addition to providing an overview of practice, this report identifies a number of themes running through the discourse that have relevance for decision making about how best to conduct peer review for a given venue.

*Working group co-leader

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ITiCSE-WGR ’20, June 17–18, 2020, Trondheim, Norway

  2020 ACM.

ACM ISBN 978-1-4503-8293-9/20/06...\$15.00

<https://doi.org/10.1145/3437800.3439207>

CCS CONCEPTS

• **Social and professional topics** → **Computing education**.

KEYWORDS

peer review, computing education, peer review criteria, peer review process, peer review roles, peer review ethics

ACM Reference Format:

Marian Petre, Kate Sanders, Robert McCartney, Marzieh Ahmadzadeh, Cornelia Connolly, Sally Hamouda, Brian Harrington, J eremie Lumbroso, Joseph Maguire, Lauri Malmi, Monica M. McGill, and Jan Vahrenhold. 2020. Mapping the Landscape of Peer Review in Computing Education Research. In *2020 ITiCSE Working Group Reports (ITiCSE-WGR ’20)*, June 17–18, 2020, Trondheim, Norway. ACM, New York, NY, USA, 37 pages. <https://doi.org/10.1145/3437800.3439207>

1 INTRODUCTION

This report sketches a map of the way peer review is conducted in computing education research, indicating standards, practices, challenges, and issues – and indicating also, around the edges of the map, some points of interest in closely-related fields. Where possible, it provides insights about what influences decisions about the peer review process and criteria. As the number of computing education venues and submissions grows, there is a growing need for qualified reviewers. This report compiles expectations about what constitutes a good review, and an effective reviewer.

The members of this working group bring considerable experience to this investigation. They include current and former editors and conference chairs from *ACM Transactions on Computing Education*, *Computer Science Education*, the ACM International Computing

Education Research Conference (ICER), and Koli Calling, among others.

Nevertheless, our intention is to learn from others. Our goal is to map the landscape of peer reviewing for people in our field, and also for others who may be interested – to identify decisions that are made, to lay out a range of possible solutions and the trade-offs involved, and perhaps to dispel some assumptions that ‘this is the way things are always done.’

The report addresses four key components of peer review:

- (1) criteria: What constitutes a good paper? and What constitutes a good review?
- (2) the review process: What are the typical components, structure, and timeline of the peer-review process?
- (3) roles and responsibilities: What are the job titles involved in the peer review process, and what are the associated roles and responsibilities?
- (4) ethics and etiquette: What are the ethical issues associated with peer review, and what is the customary code of behavior?

These topics are addressed in turn (Sections 3-6), each drawing on related literature and on the data collected (described in Section 2): documents related to peer review; interviews with journal editors and conference chairs; and a survey of the computing education community. Each of the topic sections has its own coverage of literature and data, and each discusses issues and decisions that shape that topic – and peer review overall. The final section (7) summarises the themes that emerged across the topics and those that warrant further discussion and exploration by the community.

2 DATA COLLECTION

Our datasets offer different perspectives on the peer review process. They are discussed one by one in the following sections.

2.1 Documents

We examined documents related to peer review for 17 different venues. We limited our focus in three ways. First, we focused on full research papers, rather than other types of submission, such as workshop proposals, posters, etc. Second, we focused on conferences and journals, rather than magazines, because magazines have a very different review process, and most computing education research papers are published in conferences and journals. Third, we looked only at conferences that took place before the working group met for ITiCSE, in June 2020.

The large majority of the documents we collected were gathered from public websites. In addition, there were a few internal documents such as review forms and guidelines for reviewers that we used with explicit permission from the relevant editor or conference chair.

Our list of venues, shown in Table 1, is a broad sample. While there are many other venues we could have considered (for example, CompEd - which was not included because it had occurred only once by the time the working group met), this list includes: both conferences and journals; venues with a variety of sponsors and held in a variety of geographical areas; some venues that are inside computing education research, some that have a broader scope but include computing education research, and – for comparison –

Table 1: The conferences and journals whose published guidance for peer reviewing we examined.

Type	Primary Topic	Venue
Conferences	Computing Education	ACE, ICER, ITiCSE, Koli, ICSE SEET, ISSEP, SIGCSE
	Engineering Education	LaTiCE
	Human-Computer Interaction	CHI
Journals	Software Engineering	ICSE
	Computing Education	CSE, TOCE
	Engineering Education	IEEE ToE, JEE
	Education/Educational Technology	IEEE TLT, JLS
	Software Engineering	IEEE TSE

some that are outside, but in a related field; and, within CER, some that focus on empirical research, some that include both empirical research and experience reports, some that focus on primary and secondary computing education, and some that focus on computing education at all levels, both inside and outside of formal education.

The venues we examined are:

- Seven computing education conferences: the Australasian Computing Education conference (ACE), the ACM International Computing Education Research conference (ICER), the conference on Innovation and Technology in Computer Science Education (ITiCSE), the Koli Calling International Conference on Computing Education Research (Koli), the Software Engineering Education and Training track of ICSE (ICSE SEET), the International Conference on Informatics in Schools: Situation, Evolution, Problems (ISSEP), and the ACM Technical Symposium on Computer Science Education (SIGCSE);
- Two computing education journals: *ACM Transactions on Computing Education* (TOCE) and *Computer Science Education* (CSE);
- Three conferences in overlapping or related fields: the International Conference on Learning and Teaching in Computing and Engineering (LaTiCE), the ACM CHI Conference on Human Factors in Computing Systems (CHI), and the International Conference on Software Engineering (ICSE); and
- Five journals in overlapping or related fields: *IEEE Transactions on Education* (IEEE ToE), *IEEE Transactions on Software Engineering* (IEEE TSE), *IEEE Transactions on Learning Technologies* (IEEE TLT), *The Journal of Engineering Education* (JEE), and *The Journal of the Learning Sciences* (JLS).

We also examined information related to peer review from the publishers or sponsoring organizations of these venues: ACM, IEEE, Elsevier, Taylor & Francis, and Wiley. Finally, we found additional relevant information in the websites published by CONSORT [44] and CORE (The Computing Research and Education Association of Australasia) [49].

All of the sub-groups working on the key topics (criteria, process, roles, ethics) included the core computing-education venues in their

analysis (ACE, ICER, ITiCSE, Koli, ICSE SEET, ISSEP, SIGCSE, CSE, TOCE). The review of documents for these venues was systematic, and the documents are representative of those venues (which are arguably representative of the domain). The coverage of other venues varied and is indicated in the relevant section. The analysis could certainly be expanded to other venues, for example for the purpose of further comparison with other domains.

2.2 Interviews

We conducted interviews with 11 informants, selected on the basis of broad experience with our core computing education research venues (as well as other venues) and different perspectives. Hence we spoke mainly to people with experience as journal editors and conference chairs, with a focus on a current role but the invitation to draw from the whole of their experience with academic review.

The interviews targeted information that is not in the documents: ethos and justification, decision processes, recruitment, and so on. The purpose of the interviews was to augment our map of the peer review landscape by eliciting insight into the rationale for choices about the review process; it was *not* a systematic survey of editors/chairs.

The primary venues discussed by the interviewees included:

- Journals: *ACM Transactions on Computing Education* (TOCE) and *Computer Science Education* (CSE);
- Conferences: the ACM International Computing Education Research conference (ICER), IEEE Frontiers in Education (FiE), Innovation and Technology in Computer Science Education (ITiCSE), Koli Calling International Conference on Computing Education Research (Koli), the ACM Technical Symposium on Computer Science Education (SIGCSE).

However, the experience of the informants covered many, many more venues, both within computing education research, and in other areas of computer science and software engineering.

The interviews were semi-structured, based on a flexible script, so that while all of the interviews covered the same over-arching questions, the individual conversations varied enormously, following the informant's lead. The interview script can be found in Appendix A. Ethics approval was obtained from the Open University's Human Research Ethics Committee (HREC/3598/Petre) and Princeton University's Institutional Review Board (IRB/12967/Lumbroso).

The interviews were transcribed automatically (using Trint transcription software¹) and then verified by working group members. The data were available to the whole working group, and different sub-groups conducted analysis. Some analysis was inductive, seeking broad themes; and some was targeted, seeking evidence pertinent to questions arising in the sub-groups.

The interviews were intended to elicit some of the rationale underpinning the peer review decisions of different computing education venues, and they are invoked in the discussion when they shed light on issues, challenges, or rationale. As such, they are indicative, rather than comprehensive, and further work with other groups (e.g., associate editors, meta-reviewers, reviewers) and comparison to other domains could be used to expand the work.

2.3 Survey

Because our interviews were with broadly experienced members of the community and addressed reviewing primarily from their perspective as editors and conference chairs, we also conducted a brief survey aimed at reviewers more broadly. The survey was designed to provide some simple, concrete information about the reviewer experience, in response to questions that arose in the interviews. It was also designed to be short - less than 5 minutes. The survey was disseminated to the SIGCSE community via the SIGCSE-members list (1472 recipients). The survey was intended to reveal perspectives the other sources may not have. A copy of the survey questions can be found in Appendix B. Ethics approval was obtained from Princeton University's Institutional Review Board (IRB/12967-02/Lumbroso).

We make no claim that the responses are representative. The return rate was low: 5.7 percent of surveys (84) were returned. The majority of those who responded were experienced reviewers:

- 95.2% of respondents had reviewed 10 or more CER papers.
- 76.2% had reviewed for 4 or more CER venues.
- 75.1% reviewed for 2 or more CER venues in the past year (47.7% overall reviewed for 3 or more CER venues in the past year).
- 65.6% of respondents review for other areas as well.
- 78.3% read, skimmed, or referred to review guidelines.

This indicates that there is an active review culture in the community, but it also suggests a particular constituency of respondents. Unfortunately, the working group did not have the time to analyze the data fully (for example, we have not looked for correlations between responses to different questions, or between responses and respondent profiles). As a result, we offer some preliminary results in the sections that follow, but emphasize the need for a broader, more systematic survey.

3 WHAT IS A GOOD REVIEW?

The research questions addressed in this section are:

What are the criteria applied in the peer-review process: What constitutes a good review, and what constitutes a good research paper?

The high-level criteria for any review include helping the editor or program chair to decide whether to accept a submission and helping the author(s) to improve their work. The literature and the published documents suggest many ways to interpret and apply these criteria, ranging from general guidelines to very specific details.

For clarity, we divide all these criteria into two types: *review criteria* and *paper criteria*. *Review criteria* are the factors by which the review itself is assessed. *Paper criteria* are used to evaluate the paper. A good review must take the paper criteria into account, but it must also have additional qualities, such as being submitted on time, being constructive and polite, etc.

3.1 Analysis

3.1.1 Document analysis. In order to determine the criteria by which venues evaluate papers and reviews, we first attempted to

¹<https://trint.com>

collect and synthesize all published criteria from four types of documents: guidance for authors; guidance for reviewers; guidance for associate editors or meta-reviewers as to what to look for in a review; and review forms. The goal of this exercise was to develop a representative set of meta-criteria that could be compared across venues. We performed this analysis on computing education conferences and journals detailed in Table 1, namely: ACE, ICER, ITiCSE, Koli, ICSE-SEET, ISSEP, SIGCSE, CSE, and TOCE.

After collecting all relevant documents, we extracted any sentence relating to the expectations for or evaluations of either a paper or review. While statements about criteria for reviews were almost exclusively extracted from the guidance for reviewers, statements referring to criteria for papers were found across all document types.

The extracted statements were grouped into similar thematic categories in an iterative process, with discussion and debate among several group members. Eventually, the group was able to categorize all statements into meta-criteria both for papers and for reviews.

3.1.2 Literature analysis. Next, other researchers, who had not been involved in the previous analysis, examined prior work related to peer-review criteria. We examined the closely-related literature from computing education and computer science, and, in addition, some of the STEM and education literature. This analysis was both deductive and inductive: it started with the meta-criteria found in the documents and looked for support in the literature, but was open to additional meta-criteria. Two new criteria emerged from the literature, “decisive” and “timely”. After discussion and re-examination of the documents, we concluded that there was support in the documents for these two criteria as well, and they were added to the list.

As Table 2 shows, all meta-criteria – both review criteria and paper criteria – were found in the literature, all were found in more than one source, and some were found in almost all sources. Some of the sources provide a comprehensive discussion of both paper and review criteria [23, 27, 52, 109, 125]. Others focus on paper criteria, with little or no attention given to review criteria [29, 30, 107]. One examines only review criteria, specifically the proportion of reviews that are polite and/or constructive [51]. Finally, some focus on a more specific question, such as whether reviewers consider researchers’ ethics [30] or the differing views of methodology held by researchers and reviewers [127].

The meta-criteria and related issues are discussed further in the remainder of this section.

3.2 Review criteria

The review criteria we found are listed with brief definitions in Table 3. Some observations about these criteria follow:

3.2.1 Decisive. At a high level, all the core CER venues require reviewers to make a recommendation, but the details vary. Sometimes the reviewers’ recommendations are shared with the authors, for example at ACE [3]; sometimes they are not, as at SIGCSE [123]. ITiCSE appears to have considered SIGCSE’s policy and rejected it: the language about *not* sharing the recommendations with authors is visible on its website, but crossed out [91].

In addition, at the concrete level, reviewers and meta-reviewers may not actually be required to make a decisive accept/reject recommendation. For example, TOCE provides an “Other” option for reviewers [17]. CSE allows its reviewers a range of recommendations from 1 (“Definitely Reject”) to 5 (“Definitely Accept”) [42], and ICER gives its reviewers a range from -2 (strongly disagree) to 2 (strongly agree) [72]; in all cases, the number in the middle of the range indicates “no decision”. ICER 2020’s meta-reviewers had three choices: “accept”, “reject”, and “discuss at the program committee meeting” [72]. By contrast, the ICER 2018 review form did require an accept/reject recommendation: it used the options “Strong accept”, “Accept”, “Weak accept”, “Weak reject”, “Reject”, and “Strong reject” [68].

3.2.2 Coverage. A basic level of *Coverage* is often enforced by editorial management software that requires papers to be scored with regard to key paper criteria. Such software does not, however, enforce either any justification of the scores or any connection between the individual criteria scores and the overall recommendation.

There is a range of opinion as to *Coverage*. Some say that a review should cover all of the key aspects of a paper; others disagree. Two different reasons are given for disagreeing: first, that if a paper has fatal flaws, at some point it is acceptable to stop listing its flaws; and second, that there can be good reasons for a reviewer not to address part of a paper. A reviewer may be requested to address only a single area in which he or she has expertise (for example, the use of statistics). Alternatively, a reviewer may have expertise in most aspects of the paper, but not all. As one author noted (with regard to a medical domain) [27, p. 48]:

[I]t is unlikely that any single reviewer will be expert in all of the protocols encountered in a given paper. The reviewer should comment only on those aspects of the work with which he/she has familiarity ...

In either case, if reviewers have a good reason to omit one or more aspects of the paper entirely, then software that enforces *Coverage* (e.g., requiring the reviewer to enter a numerical score for each aspect of the paper) can be problematic.

3.2.3 Justified. To be *Justified*, a review must not only address all of the key issues, but provide sufficient evidence for its conclusions. The line between insufficient and sufficient evidence can be difficult to draw, however. How many examples (positive or negative) are enough to justify a score for a particular criterion or the overall recommendation? Some justification is definitely expected, but it is not enforced; software generally provides the opportunity to write as much or as little as the reviewer chooses.

3.2.4 Helpful to Authors. For the benefit of authors, it can be desirable for a review to include more feedback than the minimum needed to justify a recommendation. For example, as noted in Table 6, the ACE reviewer instructions observe that reviewers are not required to point out “every flaw” in a paper’s writing, but “some reviewers choose to do this, and many authors appear to appreciate the additional service” [3]. One of our interviewees also discussed this issue:

Table 2: Our meta-criteria, organized by the sources that explicitly address them. The names of the criteria are fairly self-explanatory; for definitions, see Tables 3 and 4.

Citation	Review Meta-Criteria							Paper Meta-Criteria									
	Timely	Decisive	Coverage	Justified	Helpful to Authors	Civil	Actionable	Reviewer Ethics	Scope	Important Question	Related Work	Methodology	Sufficient Detail	Novelty/Contribution	Real-World Impact	Presentation	Author Ethics
Aleksic et al. [21]						1	1	1				1					
Azer et al. [23]	1	1	1	1		1	1	1		1	1	1	1		1	1	
Ben-Ari [25]									1		1	1		1		1	
Benos et al. [27]	1	1	1	1		1	1	1		1			1		1	1	
Birman and Schneider [28]												1	1	1			
Bordage [29]									1	1	1	1	1	1	1	1	1
Bornmann et al. [30]			1						1	1	1	1	1	1	1	1	1
Chubin and Hackett [38]											1	1		1		1	
Dobele [51]						1	1										
Duchesne and Jannin [52]	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
Jefferson [95]		1							1	1	1	1	1	1	1	1	1
McGill et al. [106]												1					
Meyer et al. [107]													1				
Parberry [109]	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	
Smith [125]	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1
Street and Ward [127]					1		1					1					

Interviewer: Do you think that [listing grammatical problems with a paper] is a reasonable expectation for reviewers?

Chair 3: Definitely not ... I have never believed that ... it is the function of a reviewer to be an editor. I believe it is the function of a reviewer to tell the authors and the chairs what's good about a paper and what's not good about the paper, but not to fix it. And so when I do that, I knowingly do it as an added extra.

As with the *Justified* criterion, there is a line to be drawn here. Parberry argues that: "The referee may voluntarily give the author the benefit of his or her expertise in proofreading, debugging and improving proofs, or technical writing," but cautions not to go too far: "However, the referee should beware of those who abuse the system by using the referee to perform time consuming tasks that are the responsibility of the author" [109, p. 97]. Similarly, Smith [125, p. 99] emphasizes, "You are expected to spend some time in error-detection and correction, but you are not expected to do the author's research."

In sum, reviewers must determine how much feedback to supply. This may vary depending on the type of flaws and the potential quality of the paper, as well as on the reviewer.

3.2.5 Reviewer Ethics. Parberry gives a high-level definition for this criterion [109, p.102]:

The referee has great power over the author. A series of bad referees' reports can seriously damage the career of a scientist, or at the least severely damage his or her self-esteem to the point where productivity is reduced. With power comes the ethical responsibility to avoid its abuse.

This general principle covers many specific issues, which are discussed in more detail in Section 6.

3.2.6 Meta-criteria for meta-reviews. It would seem reasonable to apply all of these criteria to meta-reviews as well, but there is very little discussion of meta-reviews in either the literature or the documents.

3.3 Paper criteria

The paper meta-criteria we found are listed with brief definitions in Table 4. They are all supported by both the documents and the literature. One dimension of variation is the level of detail: some sources offer a general guiding principle, such as "Are the methods selected by the authors able to answer the research questions?" [23], while others go into great detail on specific issues [See, e.g., 52]. In addition, paper criteria are sometimes instantiated for a particular domain. For example, one paper expressed the criterion Research Contribution/Added Value/Novelty as "Does the idea or

Table 3: Review Meta-Criteria. All have support in both the venues’ published documents and the literature.

Meta-Criterion	Explanation
Timely	Is the review turned in on time?
Decisive	Does the review explicitly make an accept/reject recommendation?
Coverage	Does the review address all elements of the paper explicitly (at some level)?
Justified	Does the review clearly explain and justify its judgments and recommendation?
Helpful to Authors	Does the review provide more help to the authors than required to justify its judgments and recommendations?
Civil	Is the tone of the review positive and appropriate to the ethos of the community?
Actionable	Does the review provide constructive feedback? Is it sufficient to allow the author to assess criticisms and repair issues raised?
Reviewer Ethics	Is the review fair and objective?

work presented have potential to contribute to *CS and/or CS education*” [emphasis added] [25, p. 4].

Methodology is covered by nearly all the papers, and several provide detailed lists of criteria. Notably, the detailed lists all concerned quantitative methods. Only one paper [23], in the medical-education domain, mentioned qualitative methods even in passing. This is currently being addressed in the computing education research community; for example, at the time this working group convened, TOCE had its own working group developing detailed guidance to include various quantitative and qualitative methods.

Effective Presentation typically addresses writing issues such as syntax, grammar, usage, organization, and clarity. As with other criteria, the implementation of this criterion varies from giving a broad general principle to very detailed specifications. One thing that is relatively rare, however, is any mention of the non-textual aspects of the paper. Azer et al. gives a nice version of this part of the criterion [23, p. 699]:

Do the figures and tables illustrate key features of the study? Are the figures and graphs clear and easy to follow with appropriate legends where needed?

3.3.1 Variation Within Venues. We expected that criteria might vary from one venue to another, but that they would be consistent

Table 4: Paper Meta-Criteria. All have support in both the venues’ published documents and the literature.

Meta-Criterion	Explanation
Scope	Do the topics covered in the paper fall within the scope of the venue?
Important Question	Is the problem or study addressed in the paper likely to be of interest or benefit to the community?
Situation in Prior Work	Does the paper clearly cite and build upon relevant existing literature? Is the coverage of current knowledge satisfactory, and is the paper situated within this context?
Rigor/Sound Methodology	Was the study designed and implemented appropriately? Does the analysis of the results adequately support the claims made?
Sufficient Detail	Does the paper provide enough detail to properly implement, replicate, or assess the research or practice?
Research Contribution/ Added Value/ Novelty	Do the findings or ideas presented enhance the understanding or practice of the community?
Real-World Impact/Application	Are the findings or ideas in the paper likely to affect the practices of the community?
Effective Presentation	Is the paper well written, organized, and structured? Do the tables and figures (if any) enhance the text, and are they readable, well designed, and integrated with the text?
Author Ethics	Includes issues such as plagiarism, ethical treatment of animals and human participants, honestly acknowledging limitations of the work, etc. See Section 6 for further discussion.

within the documents for a particular venue. Surprisingly, we found that there are often variations even within venues.

As an example, Table 5 gives representative articulations of paper criteria addressed to three different audiences: authors, associate editors, and reviewers. The examples were drawn from TOCE, which was judged to have the most complete and consistent coverage of criteria across its published documents.

Even here there are some inconsistencies. For example, while the articulation of “Situation in Previous Research” is relatively standard across the documents, the articulation of “Real World Application/Impact” points the reviewer’s attention to geographic

specificity in the review form, despite it not being mentioned in the author or associate editor instructions.

Most of these variations are relatively minor. However, differences in emphasis and an overall lack of consistency could potentially lead to reviewers or associate editors not having specific instructions to address paper criteria that had been emphasized to authors, or authors not being informed of criteria that will be evaluated. One solution to this problem might be to have a single page describing review and paper criteria for all audiences.

3.3.2 Variation Between Venues. Almost all venues offered at least some criteria (implicitly or explicitly), but the amount of detail provided to authors and reviewers in formal documents varied widely between venues. Some author instructions gave very clear and specific instructions; others only listed technicalities such as the number of pages and formatting guidelines. Some review forms broke the review down into many small categories with specific guidance for each category; some offered a simple prompt with an open text form for reviewer comments. Some venues focused very heavily on specific meta-criteria, and didn't mention others at all. Table 6 provides example text taken from reviewer guidelines for several core venues.

3.4 Interviews - insights on criteria

We reviewed the interviews for any statements relating to "criteria"; we then focused in particular on insights into the reasoning about how criteria are identified, adopted, interpreted, reviewed, and so on. This was a preliminary analysis, identifying the most salient themes in the very rich interviews. Those themes are summarized here with selected, representative quotations.

3.4.1 For whom are reviews written? One of the challenges of effective reviewing is the multiple audiences: authors, editors/chairs, associate editors/meta-reviewers, other reviewers. The different audiences may prioritize different presentations of the assessment. For example, decision-makers faced with a large number of submissions may design their venue's review form to gather scores for key criteria, facilitating compilation and summary for large numbers of reviews – in addition to detailed commentary, to justify the assessment and assist authors in improving the paper.

Chair 1: I'm pretty sure that what Program Chairs and Program Committees do, is they look ... only at the numbers. And everything that's a clear accept you accept; everything that's a clear reject, you reject; and you only ever read the reviews or the meta-reviews of the papers that are right on the borderline. So ... the purpose of writing the reviews was to provide thoughtful feedback to the author, from the community.

How editors and chairs read reviews may depend on the volume of submissions, how much time is allocated for decisions, the perceived role of meta-reviewers, and agreement among reviewers.

Chair 2: [B]ecause ICER had been growing, ... there's something needs to be done, because it's simply too heavy, the old process that the chairs read all the papers. ... And that was major, major work.

The interviews emphasized that reviews should be "written for authors," and that a review that provides thorough commentary for authors also provides sufficient rationale for chairs or editors in cases where there isn't clear consensus among reviewers.

3.4.2 Review of paper criteria. The interviews made clear that there is periodic reflection on criteria in all the venues with which the informants had experience, but that discussions about criteria on committees and editorial boards take place at different intervals for different venues, and are prompted in different ways, e.g.: by change-over of personnel, issues arising in reviewing or in the decision-making based on reviews, changes in norms in the broader community (as reflected in the practices of other venues).

Chair 3: So these criteria are becoming clearer year by year, based on discussions and misunderstandings in preceding years.

That discussion is considered to be important in broadening perspectives and developing a shared understanding of criteria:

Chair 4: So most program committee meetings are huge philosophical debates around what is the difference between an experience report and a research paper. Can experience reports ever be research? What's the line between those two? Does there need to be a line between those two? What is a sufficient amount of evidence or a sufficient amount of novel discovery to publish something? When is a replication novel and when is a replication not novel? So that's where those debates happen... I think that's valuable.

3.4.3 Differences of interpretation. Several informants commented that differences in the interpretation of criteria, when made explicit, are an important part of the discourse.

Chair 4: Academics don't agree on what constitutes good research... Even within a community, there's disagreement about that. And that disagreement is part of the progress of academic communities, as we discuss and debate and we reconsider and we toss out old ideas and start new ones. And so criteria are where that debate happens.

Conversely, seeking convergence of interpretation can limit or suppress variation.

Chair 1: I think that we need to remember that conferences do much more than just publish papers. And I think that the increasing focus on ... what one considers to be very high-quality scholarship can converge on what everybody agrees is a norm for good scholarship. And that has a potential to exclude diversity of thought and therefore contribute to an overall paucity of the research field as a whole over time.

3.4.4 Different criteria for different types of paper. All of the informants referred to the diversity of epistemology and of methodology in computer science education research, and to the challenge of reviewing different types of paper in terms of relevant criteria.

Chair 4: There are many, many diverse ways of knowing that academia uses.

Table 5: Paper Criteria Within a Single Venue. Sample quotations from *ACM Transactions on Computing Education* documents, illustrating that even within an exemplary venue site, there remain variations.

Criteria	Author Instructions [16]	Associate Editor Instructions [14]	Reviewer Instructions [17]
Scope	Appeal to a broad audience interested in computing education	Articles have a broad appeal to computing instructors and curriculum designers	
Important Question	Address significant problem(s) of lasting value	Address a question or issue of significance	How significant is the research question or problem that the paper addresses?
Situation in Prior Work	Build upon (and cite) relevant references	Linked to previous relevant research	How well grounded is the paper in relevant literature and bodies of knowledge?
Rigor / Sound Methodology	Use appropriate methodology, both for the teaching intervention and its evaluation	Methods of investigation are appropriate to the problem studied	How well does the paper use evidence to address its research questions? To what degree are the paper's chosen research methods appropriate to addressing the paper's research questions?
Sufficient Detail	Sufficient detail to replicate and evaluate	Provide sufficient details for these educational practitioners to replicate the approaches taken	Does the paper provide sufficient detail to evaluate the work?
Research Contribution / Novelty / Added Value	At least 30% new material		Please summarize the contributions of this paper to the computing education research literature.
Real World Impact / Application	Content that can be directly applied by classroom instructors or curriculum designers	Establish a clear connection to student learning	How specific are the paper's results to a geographic region?
Effective Presentation	Clearly and carefully written adhere to accepted standards of style, usage, and composition		How well is the paper written and organized? How well does the paper adhere to accepted standards of style, usage, and composition?
Author Ethics	No paper submitted to TOCE may be under review elsewhere. All papers are subject to the ACM policy on plagiarism. [A]uthors are expected to take reasonable measures to conceal their identities in their papers ...	The reporting of the research is honest and careful	

Chair 5: Some previous versions of the review criteria talked about things like, was there clear empirical evidence for something? But that's not what all papers do. Some papers make arguments like philosophers

do. There is no empirical evidence in an argument. There has to be room to evaluate an argument. But we can't have the criteria presume empirical evidence is the foundation of everything.

Table 6: Review Criteria Across Venues: Sample Quotations From Reviewer Instructions. (The criteria may also be addressed in other parts of the website; here we look only at reviewer guideline pages and review forms.) Citations included in the column header apply to all entries in the column.

Criteria	TOCE [17]	SIGCSE [123]	ITiCSE [91]	ICER	ACE [3]
Coverage			Reviews just a few lines long are generally not helpful to the authors, the APCs, or the program chairs.		[N]ote positive and negative aspects for each of the 7 criteria.
Justified		The best reviews clearly justify the reviewer’s choice The least valuable review gives a low score with no written comments	[E]nsure that your review gives reasons, explanations, examples of the points you make		Be sure that your overall recommendation tallies with your impression of the paper
Helpful to Authors		Please point out typographic and grammatical errors; if there are too many to list, please state so in your review.	The reviewers are not required to be editors, pointing out every flaw that requires correction		The reviewers are not required to be editors, pointing out every flaw that requires correction. However, some reviewers choose to do this, and many authors appear to appreciate the additional service.
Civil	Please maintain a polite and constructive tone.... We all know how hurtful a needlessly negative review can be, and how helpful a constructive one can be.	In addition to telling the authors what you didn’t like about their paper, be sure to include what you did like.		Even a very negative review should be respectful to the author(s). Avoid comments about the author(s) themselves; focus on the document ... [70]	
Actionable	The more specific you can be with regard to what the authors need to change ... and how ... the more helpful your review will be to the authors	Even if your opinion is that the paper is poorly written or poorly thought-out, you can still provide constructive criticisms to help the authors	Authors should be able to see the text parts of the reviews as a constructive critique of their submission.	Edit your review, making it as constructive and clear as possible. [70]	Wherever possible, it will include suggestions for improvement.
Reviewer Ethics		If you recognize the authors, it is your responsibility to give a fair and unbiased review ...	[R]eviewers will be required to note papers with which they have a conflict of interest	[Reviewers] must register ... [any] conflict in EasyChair. [72]. If ... you suspect ... [plagiarism], do the following ... [70].	

Editor 1: [I]f you do a literature review, you have to go beyond what is just summative...you have to add value. You have to identify gaps in the literature. You have to... come up with theories. And few papers have actually done it.

Many venues explicitly identify multiple categories of submissions, and some conferences run different tracks for different types of paper. The challenge then is to identify review criteria appropriate to the work.

Chair 2: If you ... solicit different types of papers like research papers, systems papers, theoretical papers, discussion papers, then if you have the same review criteria for these tracks, it doesn't work.

Chair 1: So there's pure research, research-to-practice, and practice papers. And so you've written a decent practice paper that describes your practice, has some empirical data, and draws some interesting insights from that, then the likelihood is it'll get accepted as something that's of value to other practitioners in the community.

3.4.5 The challenge of "weighting" criteria. Although the interviews included specific statements about criteria, and informants often identified key criteria (e.g., "Soundness is the number one criterion", "So the overarching criterion is it has to be relevant to the teaching and learning of computing, education"), those interviewed indicated that the relative weighting of criteria is a matter of judgment.

Chair 3: I don't think I've ever been aware of reviewers who say, all right, we will give this criterion three out of 10, and we will give this criterion eight out of 10, and we'll add them all up and come to a numerical score, and then we'll cut there. I think they always look at everything and make a holistic decision on papers on which they are undecided.

And the balance among the criteria may also be considered in the broader context of what the venue is trying to achieve. Judgments may take into consideration factors other than the reviewer feedback, such as the likely impact of papers, space considerations, or the diversity of topics covered.

Chair 3: So, yes, I have observed papers that have been accepted even though they were lower down on average review score, than rejected papers. And they've been accepted because we so seldom get papers on this topic area. So it is worth including because of its... novelty. Because it's different from all these papers on how we've attempted to improve the pass rate in introductory programming.

Chair 1: I think that's just a matter of bravery in the Program Committee, to be able to stand up to the decision and say, OK, it was ... a good paper, but you didn't get in any way, because we didn't have the space, and we've balanced the program based on diversity of topics and other issues, and you just didn't make it in this year.

3.4.6 Effective Presentation. Several informants discussed the challenge of the interpretation of an "appropriate standard of English" for English-language venues, given the number of submissions by non-native English speakers.

Chair 3: [G]iven that we are talking typically about venues that are published in English, and that a significant portion of our community, if not the majority ... is non-native English speaking, there's an interesting question about ... where the right level of tolerance sits between 'that's accented English' and 'that's now dysfunctional in some way'.

3.4.7 Review criteria. There was general agreement across the interviews (resonating with the review meta-criteria in Table 3) about what constitutes a good review:

Chair 1: We emphasize to the reviewers that we expect them to provide concise, usable, constructive comment.

Similarly, there was broad agreement that reviews should be written for both the authors, and for the associate editors or meta-reviewers.

Chair 3: So the reviewers and the meta-reviewer together put a case for the chairs to help the chairs make a decision of accept or reject. The other audience, equally important, is the paper's authors. And a good review will be written for both of those distinct and disparate audiences.

3.5 Discussion - criteria

3.5.1 The influence of venue identity/purpose on paper criteria. The themes that have emerged from each of the analyses draw attention to the fluid and evolving understanding of paper criteria and how they should be applied in the context of a given venue. Although a surprising consistency emerged at the "meta-criteria" level (Table 4), the interpretation of individual criteria, the perceived interplay between the criteria, and the prioritization of criteria all vary, not just between venues (for examples, see Table 6), but for a given venue over time as the community develops. The reflective discussions among chairs and editors-in-chief, and in program committees and editorial boards, shape the guidance, but need to be communicated clearly to authors and reviewers.

The interviews highlighted the relationship between the "identity" of the venue – that is, how it characterizes its purpose and its values – and the review criteria. For example, a venue that identifies itself as "archival" may prioritize rigor, whereas one that identifies itself as "community-building" may prioritize the potential of a paper to prompt discussion.

Editor 2: [A] lot of times I say... this is promising, but submit it to a conference. And then... get some feedback there. I think this is a good conference paper. This is an archival quality journal, I keep saying. So, one of the things I'd like to put into our authors page more is... what is expected in archival quality journal versus conference paper.

Chair 1: I think a conference is a good opportunity to get a good idea with a... theoretical underpinning,

out into the community for discussion. And also to... establish new research trajectories... a lot of, maybe not-super-well thought-through, conference papers have sparked off whole new areas of inquiry. And it doesn't mean that the the ground paper was the most fantastic piece of scholarship or the most amazingly well-thought-through empirical study design or whatever it was, but it served as... a spark to ignite an interest in a completely new under-explored area. And I think more of that is needed, because the computing education research community has had difficulty diversifying itself from the focus on introductory programming.

To be clear, the distinction is not simply "journal vs conference," but is based on the values and priorities of the venue. And those values and priorities influence the criteria. For example:

Chair 2: So I, if I compare Koli Calling and ICER, then I think Koli has a clearly richer set, a... more diverse set of submissions. ... And the review process itself... has supported this. So let's take an example that, in ICER there have been frequently some challenges with the reviewers; so, interpreting... what does it mean that the paper has to have a strong theoretical basis? That is one of the general criteria for ICER, and it's also frequently in the review criteria in some form. And in Koli, this hasn't been so much an issue, because we had the different tracks. OK... all of them don't need to have theoretical basis. It's a bonus... But it's not kind of the key criteria.

The informants valued the diversity of venues, recognising that each could play a particular role in the discourse.

Chair 2: So I think it's good that it has its own flavor. And I wouldn't like to make Koli a second ICER. And neither is it a good idea that ICER would change into to a second Koli.

3.5.2 Checklists vs. deterministic rubrics. Duchesne and Jannin [52] provide a lengthy checklist based on a qualitative analysis of existing lists, guidelines, and instructions from journals in their field of medicine. They stress, however, that the checklist is *not* intended as a mechanical solution [52, p.1783]:

[W]e suggest a tool to improve reviewer's [sic] productivity... Thought of as a guide, it is not meant to be a categorical tool to arrive at a deterministic assessment of the quality of a manuscript, but rather, as an aide-memoire to help reviewers in their task.

Chubin and Hackett, on the other hand, argue [38, p.46]:

[A]s peer review measures scientific performance, it should adhere to technical standards of good measurement; validity and reliability. A valid measure is one that measures the quality it is claimed to measure, not something else.. a reliable measure is one that yields the same values on repeated measures of the same object, with little random variation.

4 THE PEER REVIEW PROCESS

The research question addressed in this section is:

What are the typical components, structure, and timeline of the peer-review process?

The review process can be considered as consisting of a number of sub-processes that are largely sequential in nature, although some sub-processes can happen simultaneously. The working group analyzed guidance documents from core CS education research venues to map and compare the various editorial cycles. It used information from the interviews to clarify practices and expose rationale associated with the process decisions.

4.1 Review process structure

The structure of the review process is often inherited, for example from the previous incarnation of the conference, from a prior editorial board, or from publisher guidance, although some venues reassess the process annually. Re-assessment of the review process is often triggered by other factors: change of personnel (chairs/editors), increase in volume of submissions, external influences.

A fairly typical (simplified) journal review process is: submission, editor assessment, first round of peer review, decision – potentially iterating, following revision and re-submission, hence giving some authors the opportunity to strengthen their submissions in light of reviewer commentary. This review process may be amplified, for example, by Associate Editors, who invite reviewers, manage the review process, and make a recommendation to the Editor in Chief.

The review processes for conferences and journals have a notable difference: most computing-education research conferences have no rebuttal or revision sub-process [25]. Conferences do, however, permit revision following acceptance as part of the preparation of the camera-ready version.

Among the computing-education research conferences, notable differences in review processes are the presence or absence of three items: abstract submission preceding full paper submission, discussion, and meta-reviewing. These are discussed further below. Appendix C illustrates fairly typical conference and journal review processes, showing some of the key differences.

4.2 Timeline

Figure 1 illustrates components that are common to the peer-review processes for many of the computing science education venues (these are discussed in Section 4.3). Figure 2 refers to those components and visualizes the timelines for several computer science venues.

Halpern and Parkes argue that journals will need to provide much faster review cycles to reduce the significance of conferences and to increase their importance in computing science research [62]. Interestingly, the timeline for the first round of peer review for TOCE is comparable to the review period for the ICER and ITiCSE conferences. The target review time for TOCE was 70 days in 2016; the average actual review time was 47 days across all decision types [64]. The time is calculated from the start of the monthly review cycle, not from the date the author initially submitted the article. If the analysis is restricted to submissions that were dispatched for peer-review, the average actual review time was 71 days. If restricted further to those that were initial submissions



Figure 1: Common components in the conference peer-review process for many computing science education research venues. Variations (discussed in Section 4.3) are marked with an asterisk (*).

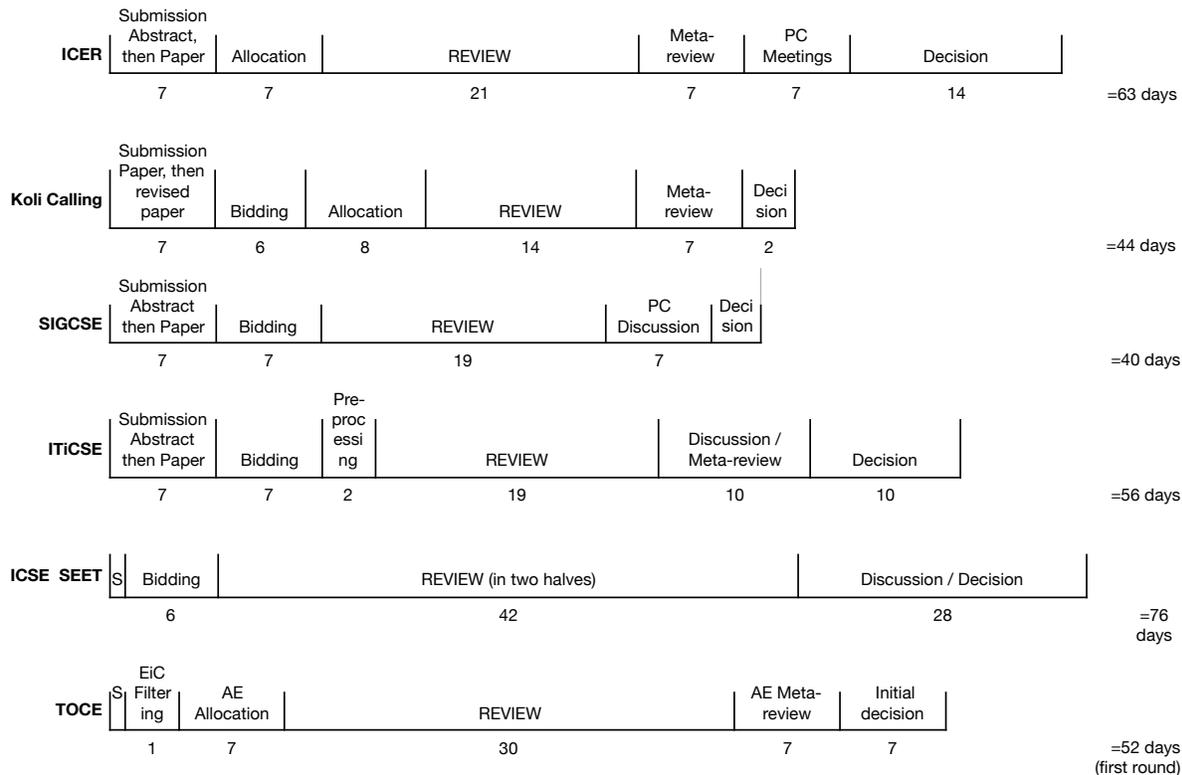


Figure 2: Best-estimate timelines for the review process. The figure summarizes the recent (2019-2020) review process timelines for ICER, Koli Calling, SIGCSE, ITiCSE, ICSE SEET, and TOCE (the order is arbitrary), based on published guidance, input from editors and chairs, and reviewer experience. Although they are as accurate as possible, they should be treated as rough indications, presented here to give an impression of the variations in process.

(and exclude resubmissions), the average actual review time was 90 days.

In contrast, Vardi argues that not only does computing science research have an over-reliance on conferences, but that those conferences are slow, compared to other disciplines, in disseminating research [132].

4.3 Common components

The components that are common to peer-review processes are discussed in turn below.

4.3.1 Abstract submission. Some conferences (e.g., ICER, ITiCSE, ISSEP, SIGCSE) ask authors to submit abstracts in advance of the full paper submission. This gives authors slightly more time to prepare their submissions, while allowing Program Committees to initiate the bidding and reviewer assignment processes in advance

of full paper submission. Where abstract submission is used, it is normally mandatory; those who do not submit an abstract may not submit a paper. Koli Calling offers an interesting variation, an “extended submission deadline” for revisions of papers for which a prior draft was submitted on time.

4.3.2 Paper submission. Papers are typically submitted via an online editorial management system, such as EasyChair, Manuscript-Central, or Taylor & Francis’s Editorial Manager. These systems assist the editors or program chairs in handling the different elements of the review process, provide access for all concerned (authors, reviewers, meta-reviewers/associate editors, chairs/editors, administrators), and maintain “history” across different conference years and journal issues. They may allow elements of automation; for example, they may include tools for plagiarism detection, allocation of papers to reviewers based on bidding, recommendation

of reviewers based on publication and review history, and checks against submission constraints such as word counts.

Each venue has its own submission requirements, such as format, page or word limits, anonymization, and supporting material. In addition, each typically requires adherence to ethical standards (as discussed in Section 6), for which explicit confirmation may be required. Conformance with the submission requirements is one of the standard filters in the review process; papers that do not conform may be rejected without review.

Different paper types. Few of the core venues distinguish different types of research papers (although many distinguish between research and various practice, tools, and curriculum papers, for which they provide different submission instructions). The exceptions are CSE, which explicitly invites both empirical and review papers, ISSEP, which invites both research and theory papers, and Koli Calling, which invites both full and short papers.

Anonymization. Most of the core venues use double-blind reviewing, on the assumption that this will reduce partiality and unconscious bias. Hundhausen reported that TOCE adopted a double-blind review process to improve the rigor of the venue and to improve parity of esteem with other computing education venues [64]. Double-blind review requires the authors to provide an anonymized version of their submission, i.e., one that removes any information that might identify the author(s) or institution(s) from the paper, and words self-citations carefully. However, anonymization is not reliable in a relatively small research community, and reviewers familiar with the discourse can often identify authors from the content. They are expected either to provide an “unbiased” review, or to declare a conflict of interest so that the paper can be re-assigned.

4.3.3 Editorial screening/desk rejection. Initial editorial screening occurs after the submission deadline. Prior to start of the core peer-review process, the individuals shepherding the peer-review process, either conference Program Chairs or journal Editors, review submissions for suitability, and reject submissions that are out-of-scope, are of poor quality, or do not conform to the submission requirements. A desk-rejected submission does not enter the core peer-review process.

The peer-review process is fundamental to the academic standing of a venue, and the reviewers (typically volunteers) are a valued resource:

Editor 2: I view the editorial board as a precious resource, a precious commodity, and I don't want to waste their time reviewing papers that aren't going to be accepted or aren't of a level of quality.

It is important that the process not be undermined nor the resource wasted. The process can be undermined when authors do not comply with instructions to facilitate the review process, such as anonymizing submissions, or understanding the scope of the venue. The resource can be considered wasted, if the peer-review process is being used to evaluate a submission outside the scope of the venue.

Both conferences and journals use desk-rejects for submissions that do not comply with submission instructions (e.g., formatting, page limits, anonymization), are outside the scope of the venue,

or obviously lack academic integrity (e.g., submissions that fail plagiarism checks). ICER documentation, for example, makes it clear that submissions will be desk-rejected and authors notified when authors do not follow submission instructions [72].

Venues may also desk-reject submissions that are not of sufficient quality (e.g., insufficient “value added” when previously-published material is included or fundamental methodological flaws). Screening is usually conducted by Program Chairs or Editors, but reviewers may flag papers of concern. For example, ICSE SEET [64]:

If you see ... issues that make you concerned, write the program chairs and we will discuss and make a decision about desk rejection.

It is not unreasonable for a journal to have a ~50% desk-reject rate.

4.3.4 Declaration of conflicts. The ACM Conflict of Interest Policy provides a definition of conflict of interest (COI) [10]:

A COI occurs when a person's objective judgment is — or is perceived by a reasonable observer to be — compromised by an existing relationship, affiliation, or connection to a person whose work they must evaluate.

This is discussed further in Section 6.

All of those involved in the peer review process are expected to declare any conflicts of interest, in order to avoid partiality or unconscious bias. Some venues (e.g., ICER, SIGCSE, ITiCSE) make this explicit in their reviewer guidance, and professional organisations such as ACM and IEEE typically have published policies. For example:

All reviewers, meta-reviewers, and program chairs who have a conflict of interest with a paper must register that conflict in EasyChair, after which they are excluded from all future evaluation, discussion, and decisions of that paper [72].

The declaration of conflicts of interest is typically integrated into the (online) bidding process (Section 4.3.5); in addition, there are both mechanisms within editorial management systems, and direct communication with the editors or chairs.

4.3.5 Matching reviewers to submissions - keywords and bidding. The optimal matching of submissions to reviewers with sufficient knowledge and experience to review them is central to any effective review process. Historically, program chairs would spend time matching submissions with the expertise of the program committee they had assembled or inherited. However, as participation in many computing science education venues grows, it becomes less feasible for program chairs to individually match submissions with reviewers. Consequently, different solutions have been proposed and adopted to make the process of matching submissions as effective as possible.

Keywords: Kumar et al. [102] suggest keywords could be better utilized by authors and program chairs to improve matching submissions. However, Kumar et al. concede that keywords may need to be optimized to encompass not only diverse disciplinary areas, such as graphics and cybersecurity, but also research focus, such as qualitative and quantitative methods. For example, in 2020

ICER asked authors to identify what skills and knowledge reviewers should have in order to review their paper. SIGCSE also asks reviewers to update their reviewer profiles to “include 3-5 topics that you are most qualified to review” and “please ensure that you select at least one of the new Methods Topics” [124].

ITiCSE has historically used such a keyword-style approach [86]. The contact author is expected to assign one or more topic keywords to support the matching process. The conference also constrains matching to reduce conflicts, for example by ensuring reviewers from the same institution and/or the same country, or the same state in the case of the United States, as the author, are not matched.

Bidding. In 2017, the ITiCSE matching process evolved significantly with the inclusion of a bidding process for reviewers [88]. The ITiCSE bidding process expects reviewers to bid for a number of submissions, far more than they would be actually expected to review. The expectation is that reviewers will be allocated submissions that they are qualified to evaluate effectively.

Reviewers are also expected to register any conflicts of interest during the bidding process, so as to avoid undermining the review process.

Editorial management applications such as EasyChair can match reviewers to papers automatically, based on profile and bidding data.

Chair 3: I have for many years noticed that the chairs are sometimes likely to say, no, we won't have that person reviewing that paper. We want somebody who's got a really solid basis in statistics, for example. And so they will make the decision that way. And in recent years, they have been taking further steps to try to help the reviewers self-select in a better way. So, for example, [in 2020] they asked authors to say what skills and knowledge should reviewers have in order to review this paper? And they asked reviewers to look at those responses while bidding for the papers. So more and more effort is being put into trying to get the reviewers to nominate the papers that they are best suited to review, so that the automatic process can do its job.

Here, the conference chair acknowledges that bidding is not perfect. The individuals shepherding the process appreciate the limitations, and intervene when necessary to achieve the best match they can.

Cabanac and Preuss [32] suggest that various factors may affect the bidding process, both positively and negatively. They argue that individuals are more likely to select items at the top of a list, even when randomized. Moreover, an individual's knowledge and experience may influence their choice. These issues were also noted in one of our interviews:

Chair 3: EasyChair presents the papers in a different order ... So, if reviewers start at the top of the list and work down for a while and think 'that's enough, I don't want to do any more', we don't have the first 50 papers getting lots of bids and the last 50 getting none at all, because they are quasi-randomized. So there are certainly some reviewers who look at every one of them and who often bid 'yes' or 'maybe' or

'conflict' for every paper. And I admire that. But we can't enforce it. What I have tried to enforce in a couple of conferences is a minimum bid. You must bid, yes or maybe, on at least 30 papers.

Similarly, Rodriguez et al. [112] argue that “referee fatigue” can compromise the bidding process with many submissions receiving no bids or some bidders deeming themselves experts in every domain to the point that the bidding process is pointless. Moreover, a general concern with the bidding process is that reviewers opt for submissions that they are interested in reading, rather than what they are qualified to evaluate.

Nevertheless, despite such concerns and limitations around the bidding process, there are clear advantages in reviewer bidding as participation in venues grows, and it's not clear that a better alternative exists.

4.3.6 Number of reviews per paper. The number of reviewers assigned to a submission varies from venue to venue (e.g., Freyne et al. [54] report that peer-review usually involves three to five external reviewers), and the number is often not explicitly justified, or even discussed, by many of the core computing science education research venues in their documentation. This is driven in part by the number of reviews that are considered to be required for each submission and its type, and in part by the reliability of reviewers in providing timely and useful reviews. The number of reviewers assigned per submission in turn affects the number of submissions a reviewer is assigned.

The optimal number of reviews per submission is the number that ensures each submission is considered fairly and thoroughly. Among the venues that declare the desired number of reviews per submission (e.g., ICER, ToCE, ICSE SEET), three is common, although ITiCSE specifies four (previously six). The survey received similar responses: 74.4% of respondents considered 3 to be an acceptable minimum for the number of reviews per paper (and 22% considered 4 to be an acceptable minimum number of reviews.)

How many is 'optimal' depends in part on the expertise of the reviewers:

Editor 1: [T]his is also the reason why we can base an accept decision based on two reviews most of the time, because the people are selected based on their experience and familiarity with the field. And, if I have two experts saying in unison this paper should be accepted, I don't need a third person.

The argument could be made that a set of two reviews from experienced reviewers with appropriate discipline knowledge, experience of relevant research methodologies, and a track record of effective reviewing is more desirable than a set of six reviews from inexperienced reviewers that have no such knowledge or experience. Further, the number of reviews considered 'sufficient' is generally lower in cases where the reviewers agree.

Even if the optimal number of reviews can be determined, there is the concern that not all reviews will be obtained in a timely manner, resulting in acceptance decisions for submissions being determined on a varying number of reviews per submission.

Chair 3: [Y]ou look back at the forewords of the [ITiCSE] proceedings, and they say things like every paper had

at least three reviews and some had as many as six. Because what's been happening with a conference such as ITiCSE is a number of reviewers aren't as interested as they think they are, or they're interested, but they don't get the time, or something happens, some illness in the family or something. And a lot of reviewers simply don't deliver their reviews.

A primary concern is obtaining a sufficient number of high-quality reviews in a relatively short period of time to support an accurate decision process. Unfortunately, this can result in some venues seeking more reviews, effectively "over-recruiting," in order to ensure that the minimum number of reviews is attained.

4.3.7 Emergency Reviewing. ITiCSE provides a clear articulation of emergency reviewing [91]:

Despite the best of intentions, some reviewers find that they are not able to review the papers assigned to them and others simply drop out of contact for various reasons. The program chairs often discover this only at the end of the reviewing period. They then need to find reviewers who have completed their assigned reviews and are able to review one or two more papers within a day or two.

Journals, as well as conferences, may need to compensate for reviewers who drop out, although (except in the case of scheduled special issues) there is usually capacity to extend the reviewing period. For most venues (both conferences and journals), emergency reviewing is invisible to authors and not explicit in the guidance; it is typically handled quietly, through direct invitation to reviewers, although some venues (such as ITiCSE) may put out a general call.

4.3.8 Meta-review. Fundamentally, a meta-review is an overview of all the reviews provided for a submission. It may include a recommendation, and it typically draws on the experience and expertise of the meta-reviewer and hence is more than just a summary. Meta-reviewing is built into journal review processes that include an Associate Editor (who acts as the meta-reviewer and also handles assignment and reviewer management); it has become increasingly common for conference venues, to support a timely decision process, especially where the number of submissions is high.

Chair 3: So the reviewers and the meta-reviewer together put a case for the chairs to help the chairs make a decision of accept or reject.

Chair 1: [I]f you find meta-reviewers that you can trust to do their job well, then they can decide how to filter, include, exclude, and editorialize the reviews that were written.

The meta-review process is fairly well established in computing education venues. SIGCSE 2014's documentation [117] outlines the use of meta-reviewing to support the program chair. SIGCSE's meta-reviewers (called Associate Program Chair (APC)) were expected to:

- summarize the case for accepting or rejecting the submission that emerges from the submitted reviews;
- state a clear recommendation for the submission to support the decision process; and

- ensure that the reviews meet the standards expected of the conference.

ITiCSE introduced meta-reviews and the use of Associate Program Chairs later than SIGCSE, in 2017 [88].

ICER's meta-reviewing role has evolved over the years. Its first version, in 2014, named a "meta-reviewing group," was quite different from the process SIGCSE adopted that year. The group was made up of ICER's three Program Chairs, plus two additional Associate Program Chairs (APCs). Each member of the meta-reviewing group was responsible for reading approximately 30 papers (and their reviews), enough so that each paper would be read by two members of the group. The group then met to discuss all the papers and make the final decisions. In effect, rather than adding an extra meta-reviewing phase to the process, ICER 2014 expanded the number of Program Chairs [65].

By 2017, ICER had developed a separate meta-reviewer phase, similar to SIGCSE's. It expanded the number of meta-reviewers and assigned each of them to write written reviews, with the goal of supporting both decision-makers and authors, and to manage a discussion period among the reviewers [67]. In 2020, the meta-reviewers began meeting with the Program Chairs to discuss papers, as in the 2014 model, but after managing the reviewers' discussions, ensuring that reviews satisfied the review criteria, and producing written reports [70, 72].

The meta-reviewing role is discussed further in Section 5.2.

Coping with volume. SIGCSE 2014's use of meta-reviewing is not unexpected given the scale of the venue. Where there is agreement among reviewers and meta-reviewer, program chairs can consider a single meta-review per submission, instead of reading all the reviews:

Chair 3: I would be inclined to guess that, for most of the sizable conferences I'm involved with, the chairs don't read all the reviews. For example, if a paper has strong reject, strong reject, strong reject ... and ... the meta-reviewer says we don't need to discuss this any further – clearly, nobody thinks that it's worth looking at. I wouldn't be surprised if the chairs don't read those reviews. ... It's even plausible that if a paper says strong accept, strong accept, strong accept, the chairs don't read those reviews.

Borderline or controversial cases are typically discussed, with both the meta-review and the reviews (and perhaps further input from the meta-reviewer) informing the decision.

ICER 2014's addition of a "meta-reviewing group" was also driven by a high volume of submissions [65]:

ICER originally had relatively small numbers of submissions, and the three chairs could comfortably look over all submissions, along with the reviews, and make a sensible choice of papers for the programme. With 68 submissions last year, and maybe more this year, even if each paper was looked over by 2 of the 3 chairs, that would be over 40 each.

Discussion within meta-reviewing. By 2017, the version of meta-reviewing used by ITiCSE [88], ICER [67], and SIGCSE [120] included the expectation that meta-reviewers would encourage discussion among reviewers. (Prior documentation for ICER [66] and SIGCSE [119] does not suggest meta-reviewers are expected to motivate discussion between reviewers.)

Chair 3: [T]hat particular purpose of the meta-reviewer is to see whether the really positive reviewers, having read the other reviews, might say, ‘Oh yeah, that is a bit of a drawback, I hadn’t thought of that’, or whether the really negative reviewers will say, ‘Oh yeah, I suppose it does have some good points; I just overlooked some because it was such a bad paper overall’. So the meta-review discussion can bring the reviewers closer together – or can at least make them a little more receptive to the thought that reviewers can validly have different opinions.

Note that discussion *within* meta-reviewing is distinguished from program committee discussions that *follow* meta-reviewing (see Section 4.3.9).

Different perceptions of discussion within meta-review. The outcomes noted in the quotation above highlight different goals for discussion within meta-review: reviewer reflection (with potential improvement to reviews), convergence on a recommendation, and recognising legitimately different opinions. Our survey respondents perceived meta-review as having multiple purposes, typically (83.3%) making multiple selections from the options offered (e.g., facilitate discussion among reviewers (77.5%), summarize and filter reviews (76.3%), help reviewers homogenize scores (51.2%), and exclude poor-quality reviews (38.8%). In written comments, they also suggested: raise points that have been missed by reviewers; allow reviewers to see and learn from other reviews; identify perspective differences/outliers; provide accept/reject advice to the program chairs.

The tension between encouraging convergence and promoting an effective discourse is highlighted:

Chair 1: I’m strongly... opposed to meta-reviewing. I think it causes reviewers to converge on a single view of a paper. And I think the authors would be better served by seeing variation. ... I think it’s disrespectful to the reviewers because it dis-empowers the minority reviewer in terms of the validity of their point of view. And I think it’s unfair to the author, because – if there was one positive review, [that] thought that there was an aspect of the paper that was really, really cool, could be developed further – by the time we normalized and converged on something, maybe that positive input and impetus is lost.

Hence, requiring meta-reviewers to encourage discussion and reflection among reviewers could have unintended and undesirable consequences.

SIGCSE 2020 guidance targets consensus [124]:

As an APC, we expect you to lead the discussion among the reviewers to reach consensus on a recommendation about whether the paper should be accepted or rejected.

but *not* at the expense of reviewer (or meta-reviewer) autonomy [124]:

It is important that at no point reviewers should feel forced to change their reviews, scores, or viewpoints in this process. The APC can disagree with them and communicate that to the Program Co-Chairs as needed, but the APC should NOT force reviewers to change their review because of a difference in viewpoint.

The key aspect is that the meta-reviewers, and reviewers for that matter, have autonomy and are not required to change reviews. This guidance suggests that meta-reviewing is not about enforcing convergence, but about encouraging reviewers to reflect, with a view to conducting an effective evaluation of a submission.

4.3.9 Decision. Decisions are typically made by the journal Editor or conference Program Chairs, informed by the recommendations of the Associate Editors or meta-reviewers, and by the reviews.

Chair 3: So the meta-reviewers should understand that they write the meta-review. They make the recommendation. But the chairs might decide other than that recommendation.

Traditionally, computer-science conferences relied on the Program Committee to discuss disparities between reviewers and reach consensus on disputed or “borderline” submissions [See, e.g., 125]. ICER has introduced this model in 2020 [72]:

The role of the meetings is to reach consensus (by vote if required) on the outcome for each borderline submission. After a decision is made in each case, the responsible PC member (meta-reviewer) will add a summary of the discussion at the end of their meta-review, explaining the rationale for the final decision.

Such PC discussions may be instead of, or in addition to, the meta-reviews.

4.3.10 Revision and re-submission. The review process for journals is typically iterative, allowing a period for authors to respond to reviewer commentary and revise their submission as appropriate, followed by either additional review (usually by the same reviewers) or an editorial decision, depending on the venue and the scope of the revisions required. The number of iterations is limited; the ideal is that the paper improves with each iteration and moves from major to minor revisions (or acceptance), and from minor revisions to acceptance.

Editor 2: [A] paper that’s been recommended for major revisions in the first submission ... our new policy in the last two years is that that paper must then advance to minor revisions or accept the second revision in order to stay in the review process. ... if a paper is not advancing, ... I want to allocate those resources, those AE resources and those reviewing resources, to another paper.

Hundhausen [64] reported that more than 90% of articles resubmitted to TOCE in 2016 were accepted.

None of the core CER conferences has a rebuttal or re-review phase. However, they all have a period during which authors of accepted papers can revise/improve their submissions in response to reviewer comments, as part of the preparation of the camera-ready version of the paper. Some other CS conferences (e.g., ACM CHI) do offer a rebuttal/review phase for selected papers.

4.4 The impact of scale

The increase in the volume of submissions is having an impact on all aspects of the review process, including sub-processes, recruitment, roles, and criteria.

Chair 3: So even when we've reduced the number of reviews per paper, the number of submissions is climbing rapidly from year to year. And we are always struggling to get as many reviewers as we need.

Birman and Schneider [28] focus on this challenge and argue that it is caused, in part, by computing's reliance on conferences rather than journals: instead of publishing a longer article containing several results, authors publish smaller incremental chunks in separate conference papers, each of which requires reviewers. In sum, Birman and Schneider argue that this emphasis on conferences results in a larger number of submissions, a heavier burden on reviewers, and – a concern for the field itself – work that is less polished (due to the absence of an enforced revision phase), but at the same time more conventional and less original.

However, we note that the computing education journals are also experiencing a significant increase in submissions. One editor attributes this to growth in computing education research:

Editor 2: From my perspective as as Editor in Chief, I'm more concerned about scale ... than others, because I see the volume of papers, and I understand ... how much of an investment it is to send a paper out for review. So my ongoing concern with this journal is that it's not going to keep up with the growth in the field. I've seen a proliferation of papers and trying to keep up with that has been challenging. ... how I'd like to change that is by having more reviewers, or more associate editors...

Nevertheless, there is also concern that many submissions are less polished, resulting in more editorial filtering:

Editor 1: ...with the increase in scale ... also the number of papers which didn't make it past our desk actually increased ... disproportionately. So we're getting a lot more papers than we used to get. But I think the number of papers that are being sent out to review has been stable over the past three years.

4.5 Preserving knowledge of the process

Kumar et al. [102] argue that conferences have limited, if any, memory and that as a consequence lessons are not learned and problems persist. The computing research community appears to have more continuity. For example, Koli Calling prioritizes its community and chooses chairs who have engaged with the conference. SIGCSE

has addressed the continuity issue with the (staggered) Junior and Senior Program Chairs [72].

Chair 6: So there's...a nice balance with trying to find...a few ways that you can make improvements every year, and then making sure that the people who follow have the knowledge they need, that they can start making those changes as well. Which is why I really like that we moved to the junior/senior role within a particular Program Chair or symposium chair. So that knowledge isn't lost from year to year.

ICER also has two junior and senior Program Chairs, but for ICER, this represents a loss of institutional memory: through 2016, ICER had three Program Chairs with staggered terms. The move from three Program Chairs to two was accompanied by the creation of two new positions – junior and senior Site Chairs – who took over significant responsibilities that had been handled by the Program Chairs.

Another way to ensure long-term institutional memory of the conference's policies and practices is to create a Steering Committee. A number of conferences have adopted this model, including ITiCSE, ICER, and CompEd.

5 WHO IS INVOLVED

The research questions addressed in this section are:

What are the job titles involved in the peer review process at computing education conferences and journals, and what are the associated roles and responsibilities?

For purposes of this discussion, we define “job title” as any title found on a conference or journal's website or in the literature that appears to be related to the peer review process, and “role” as a set of responsibilities related to the peer review process.

The distinction between “job title” and “role” is important, because there is a multitude of different titles for the roles involved in the peer review process, both in the literature and in the online documentation. A given title can refer to different roles, and different titles can be used to describe the same role. Different venues use different titles, and a given venue may change job titles from year to year. As a result, it is not always possible to deduce from a title what the real responsibilities are.

For example, in the early years, ICER had a *review committee* whose members reviewed the papers, and the *conference chairs* selected papers based on the reviews. A few years later, although the tasks were the same, the name of the review committee was changed to *program committee* and the conference chairs became *program co-chairs*. In 2017, a new level of actors was established for shepherding discussions among reviewers and writing meta-reviews, and they were called the *program committee*, while the previous year's *program committee* became the *reviewers*. Then in 2018, these meta-reviewers were called the *senior program committee* and the reviewers were called the *program committee* again. Such changes are not unique to ICER; they reflect the dynamic nature of conference organizations.

Since the titles are ambiguous, we define the following names for the key roles: *EIC*, *AE*, *C*, *PC*, *M*, and *R*, as shown in Table 7. These names are deliberately artificial, to distinguish them from the actual titles used. Roughly speaking, however, they correspond

to Editor and Associate Editor (for journals); Chair, Program Chair, and Meta-Reviewer (for conferences); and Reviewer (for both) ².

In the following subsections, we explore these key roles, the associated titles, and their responsibilities (Sections 5.1 and 5.2); the available information about how individuals are recruited for these roles (Section 5.3); different ways to reward reviewers (Section 5.4); and monitoring review quality (Section 5.5). These are the main topics – from the roles point of view – in the literature. In addition, we examine various ways in which conferences can be structured, based on their roles (5.2) and the increasing workload associated with these roles (5.6).

Our discussion is based on the literature, the documentation provided by the venues listed in Table 1, and our interview data. There is a significant body of literature concerning the peer review process. For purposes of this section, we draw on literature from other fields of science to supplement the papers concerning computer science or computing education; much of the broader literature is of a generic nature, and similar types of challenges occur in most fields of science. In addition, because the literature does not explicitly focus on roles as such, we capture the following information from papers that address the review process. Most of the information relates to reviewers, but some information about other roles is presented as background. Finally, most of the information concerning conferences is drawn from the documents and interviews, because almost all papers outside computing focus on journals, the main publication venue in other sciences. A few papers address computer science conferences, however, such as Shah et al. [115], who present a very detailed analysis of the Neural Information Processing Systems conference’s review process.

5.1 Journal roles

All the journals we examined share the three key journal roles defined in Table 7: *EIC*, *AE*, and *R*. We discuss the titles and responsibilities associated with these roles one by one, followed by the responsibilities associated with a few more unusual titles.

The *EIC* role is associated with the titles “Editor” and “Editor-in-Chief”; all the journals we examined had exactly one of these two titles. Whatever their names, the responsibilities described for *EIC* and *R* are consistent with those discussed in the literature. Parberry [109], a comprehensive description of the review process in the context of theoretical computer science, lists the following responsibilities associated with *EIC*s (which Parberry calls “Managing Editors”):

- making policy decisions,
- selecting *AE*s (which Parberry calls “Editors”),
- acting as an intermediary between the publisher and the Editorial Board (a group of *AE*s), and
- resolving any disputes between *AE*s, authors, and *Rs*.

The *AE* role was usually associated in our documents with the title “Associate Editor” [14] and sometimes “Member of the Editorial Board” [97]. The term “Editorial Board” is ambiguous, however. Sometimes it was used to refer to the *AE*s collectively [41]; sometimes it referred to the *EIC*, the *AE*s, and (optional) miscellaneous additional titles [15, 82, 83, 96]. Sometimes it was used in both ways,

to refer both to the *AE*s and also to the longer list of which the *AE*s were a subset [97].

The transparency of the *AE* role varies widely. When the responsibilities of the *AE* were listed, they were consistent with those described in the literature, but the information was not always available. Sometimes *AE*s are named, but we found no information about their responsibilities [128]. In one case, the Editorial Board included the *EIC*, the *AE*s, and others, and responsibilities were assigned to the “Editorial Board” collectively, making it difficult to tell which responsibilities actually belong to which roles [82, 83].

An exemplary description, by far the clearest one we found, is given on the TOCE website [14]. It describes the *AE*’s role relative to reviewers and the *EIC* as follows [14]:

- (1) “Assign at least 3 reviewers with relevant expertise within 1 week of receipt”
- (2) “Ensure that reviewers complete these reviews within a reasonable timeframe, optimally within 30 days of assignment”
- (3) Make a written recommendation to the *EIC* “along with a detailed rationale” and “within 1 week of receiving the reviews”
- (4) “clearly state what changes need to be made” before the paper can be considered for re-review (for major revisions) or before it is published (for minor changes)
- (5) Discuss the paper with the *EIC* as needed to reach a decision
- (6) By default, continue as *AE* if a paper is revised and re-submitted, handling those reviews in the same way, except for one additional task: they must also ensure that the authors provide the required cover letter thoroughly explaining their response to each of the action items suggested by the *AE* and reviewers.

Parberry’s list of *AE* responsibilities [109] is consistent with the TOCE description. Smith, in a guide for computer-science reviewers, gives a similar description and adds that the *AE*s need to resolve conflicting review reports and “tell the authors to what extent they must comply with the referees’ comments when making changes” [125, p. 70]. In theory, an *AE* can overrule even a unanimous reviewers’ recommendation. This is rare, but it is not unusual for an *AE* to take the side of the minority of reviewers when selecting and justifying a recommendation. Another important responsibility for Smith’s *AE*s is looking for possible plagiarism.

The third key role, the *R* role, is associated with the titles “Reviewer” and “Referee” in the journals we examined; sometimes both are used interchangeably on the same webpage (see, e.g., [20]). Their responsibilities are similar in both journals and conferences; in both cases, they are the foundation of the review process.

With regard to the *Rs*’ responsibilities, much more information is available than for any of the other peer-review roles. All conferences and journals have some form of instructions for reviewers. Their primary responsibility is to evaluate the paper(s) they have been assigned and submit anonymous reports to the *AE*, *M*, or *PC*, depending on the venue. A good summary for the character of the review report is that “The audience of the formal report therefore consists of the editor and the author, who desire the same information but have different perspectives” [109, p. 97]. Their task is to

²Another term frequently used is Referee. We use Reviewer and Referee as synonyms.

Table 7: Names used in this paper for the key roles in the peer review process, along with definitions. These names are deliberately artificial, but roughly speaking, they correspond to (E)ditor-(i)n-(C)hief, (A)ssociate (E)ditor, (C)onference Chair, (P)rogram (C)hair, (M)eta-Reviewer, and (R)eviever. Any of these roles may be performed by more than one person: for example, a journal may have two *EIC*, a conference may have one, two, or several *PCs* working together, and there are usually several *Rs* associated with any venue.

Venue type	Role	Definition
Journal	<i>EIC</i>	responsible for journal’s whole review process; makes final decisions on paper acceptance
	<i>AE</i>	manages review process of individual papers under <i>EIC</i> ’s instructions; invites <i>Rs</i> and coordinates their work; makes justified recommendations to <i>EIC</i> concerning paper acceptance
Conference	<i>C</i>	general oversight of conference; has indirect effect on review process, for example by determining how many rooms and days will be available
	<i>PC</i>	makes final decisions on paper acceptance (perhaps in consultation with a committee or by chairing a committee and recording the committee vote)
	<i>M</i>	reads papers and reviews of those papers; encourages discussion among reviewers (if discussion is part of process); submits synthesis of reviews and justified recommendation concerning acceptance of papers to <i>PC</i>
Both	<i>R</i>	scientist who analyzes assigned papers and writes justified judgement of their pros and cons

“read the paper carefully, with an open mind, ... [with] no presumption as to its quality or accuracy when checking and evaluating the paper” [125, p. 65].

One difference between conference and journal *Rs* is that journal *Rs* are often held responsible for re-reviewing the same paper, if it is revised and re-submitted. They might, then, think longer term about the development of the work than a conference reviewer, who is only required to make a yes/no decision. It is not unknown for authors to submit their papers to one conference after another, however, so conference reviewers may find themselves evaluating the same paper more than once as well.

For more about reviewers’ responsibilities, see Section 3, which discusses the nature of a good review, and Section 6, which covers ethical issues raised by the reviewers’ work.

In addition to the three key roles, we found several additional titles that each appear in one journal only, with little or no information about the associated responsibilities. Wiley’s *Journal of Engineering Education*, for example, has a Deputy Editor, an Assistant Editor, a Senior Associate Editor, and an Advisory Board. *IEEE TSE* has an Associate Editor-in-Chief.

Although these titles appear rarely in our sample, a quick Google search reveals that they are used by other journals. They might be ways in which editors delegate some of their responsibilities or they might simply be honorary titles. In general, the use of such titles is probably part of the unwritten culture of a particular journal, perhaps shared with other journals from the same publisher.

5.2 Conference roles

The ten conferences we examined had, among them, 12 titles for roles potentially related to the peer review process. Organized by the roles they correspond to, they are:

- *C*: Conference General Chair (ITiCSE); General Chair (ACM CHI, LaTiCE); General and Organizing Chair (ISSEP); Symposium Chair (SIGCSE).
- *PC*: Program Chair (ITiCSE, SIGCSE); Program Committee Chair (LaTiCE, ISSEP); Subcommittee Chair (ACM CHI), Track Chair (ICSE SEET).
- A set of *PCs*: Subcommittee (ACM CHI)
- *M*: Associate Program Chair (ITiCSE, SIGCSE), Associate Chair (ACM CHI)
- A set of *Ms*: Program Committee (ICER).
- *R*: Reviewer (ICER, ITiCSE, SIGCSE), Associate Chair (ACM CHI), External Reviewer (ACM CHI)
- A set of *Rs*: Program Committee (Koli, ACE, ICSE SEET, ICER, ISSEP, LaTiCE); Review Committee (ICER). Also Additional Reviewers (ISSEP).
- *C* combined with *PC*: Program Chair (ICER, Koli), Chair (ACE)

Not only does each of these roles have more than one name, but some of the names correspond to more than one role. “Program Chair” can mean either *PC* or a combination of *C* and *PC*. “Program Committee” can mean a set of *Ms*, a set of *Rs*, a set of *PCs* who work together to decide on paper acceptance (as in [125]), or a set of roles that are not related to reviewing research papers at all [122]. At CHI, some Associate Chairs write reviews (*R*), and some write meta-reviews (*AE*) [36].

The responsibilities associated with *PC* and *M* are discussed in the following sections. The responsibilities of *Rs* are discussed above in Section 5.1, and further in Sections 3 and 6.

5.2.1 *PCs*. The *PCs*’ main responsibilities include preparing the call for papers and deciding which papers to accept. Preparing the

call for papers, in turn, involves setting the scope of the conference and defining the review criteria:

Chair 3: The criteria in all of the conferences I work with are determined by the chairs. Sometimes they are simply copied from the preceding year, but other times the chairs say: ‘I am not at all happy with these criteria. I want to reword them in this manner.’ So it’s always up to the chairs to choose the criteria. Almost always they are some kind of improvement on those from the previous year.

Another one of the *PC*’s major responsibilities is to decide on the program:

Chair 3: I don’t remember ever working for a conference that has meta-reviewers where every meta-reviewer recommendation has been accepted without question. So the meta-reviewer ... should understand that they write the meta review. They make the recommendation. But the chairs might decide other than that recommendation.

While reviews are, of course, the main instruments for selecting the program, there are sometimes other criteria, too, especially concerning selection among borderline papers.

Chair 3: When the paper hasn’t scored as well as others in the reviews, then the reviewers will effectively be told, ‘Yeah, you weren’t really enthusiastic about this paper, but the chairs have seen reason to include it in the program’.

A conference may have one, two, or three *PC*s making the paper decisions, or may involve a larger group. At ICER in 2020, the *PC*s shared part of their decision-making responsibility with the *Ms*. *Ms* were required to attend one or more virtual meetings, in which they participated in the accept/reject decisions in borderline cases: “the goal is to collectively reach consensus, rather than relying on the program chairs alone to make final decisions” [70].

This is similar to the model described by Smith, except that Smith describes the process in a time before the large scale-up in the number of submissions. In Smith’s version, the *PC*s could meet face-to-face and discuss all the submissions and their reviews, though not in as much depth as a journal could. To make things faster, a numerical ranking of papers based on reviewers’ scores was often used as a starting point. Decisions were made by majority vote. The committee chair’s vote might have been weighted more heavily than that of other committee members, but generally not enough to alone out-vote a majority of the committee [125, pp. 67, 70].

5.2.2 *Ms*. ITiCSE describes the *Ms*’ responsibilities as follows: “Associate program chairs (APCs) lead the discussion among reviewers and metareview each paper, providing a recommendation and feedback to the program chairs” [91]. They can expect a workload of 8-12 six-page papers, and the discussion and meta-reviewing periods combined last ten days.

ICER’s instructions are more detailed: instead of “leading the discussion”, they must ask reviewers “to read all the reviews of their assigned papers and discuss any disagreements” [72]. The recommendation the *Ms* make is “whether to accept or reject the

work, or whether to discuss it at the *PC* meeting”, and must be based on the reviews, the discussion, and their own evaluation of the paper [72].

Moreover, ICER’s *Ms* have some additional tasks: they must make sure that the reviews are both constructive and “aligned with the review criteria” (which correspond to our “paper criteria”), asking reviewers to revise their reviews if necessary [72]. The *Ms* “will be expected to attend” one or more video-conference meetings to discuss borderline cases. If appropriate, they can nominate one of their assigned papers for the Chairs Award [72]. The ICER *Ms* can expect to review 8-10 papers, each up to 10 pages long, as well as working with 24-30 reviewers [72].

In some ways, the *M*’s role is the most challenging. While *Rs* can learn how to write reviews, in part, by reading others’ reviews, learning how to write good meta-reviews is not so easy:

Chair 7: At some point you can see other people’s reviews and you can use that to model behavior. You can’t do that with meta-reviews, because there’s no access to other people’s meta-reviews.

As another interviewee pointed out, *Ms* have to juggle responsibilities to three different audiences:

Chair 8: When I started meta-reviewing, we were told that we had three audiences: for the program chair, we had to make a recommendation and support that recommendation with evidence from the paper and the reviews; for the authors, we had to synthesize the reviewers’ comments and prioritize any suggested changes; and for the reviewers themselves, we were supposed to manage the discussion period and, in our meta-review, provide an example of a good review. ... I now think that it’s also important to acknowledge the work done by the reviewers. So I look for something in each of the reviews that I can quote in the meta-review.

Finally, the *M* must be prepared to be overruled:

Interviewer: [H]ow much power does the ... meta-reviewer have ... Say you’ll make these judgment calls between [a] two to one split. Are you allowed to go the other way or should you just be summarizing ... ?

Chair 8: I assume that the program chair can overrule my recommendation. I don’t know if you ever saw the old Siskel and Ebert TV show about movies. ... Well, what they would do is they would give you thumbs up or thumbs down. But they also gave you these really generous clips so you could make up your own mind. And that’s kind of the meta-reviewer job in my mind. ... I give you [the program chair] my recommendation, but I also try to give you all the information.

5.2.3 *Roles and conference structure*. By translating the multitude of conference titles into our role names, we can observe something about a conference’s structure. The conferences we looked at illustrated the patterns shown in Table 8.

Each pattern corresponds to a different conference structure, reflecting the ways in which these conferences have evolved. SIGCSE and ITiCSE, both fairly large and established (and sponsored by

Table 8: Roles patterns in the 10 conferences. The roles correspond roughly (but not exactly) to Conference Chair, Program Chair, Meta-Reviewer, Associate Editor, and Reviewer. (For more details about the roles, see Table 7). A hyphen means that two different roles are performed by the same individual(s).

Roles	Conferences
{C, PC, M, R}	SIGCSE and ITiCSE
{C-PC, M, R}	ICER-19
{C-PC, M-PC, R}	ICER-20
{C, PC, R}	LaTiCE and ISSEP
{C-PC, R}	ACE and Koli
{PC, R}	ICSE SEET
{C, {PC, AE, R}, {PC, AE, R}, ...}	ACM CHI

the same organization), have four separate titles, each corresponding to one of our four roles. ICER has all four roles, but assigns *C* and *PC* to the same title. The notation also captures a change that happened between 2019 and 2020: ICER’s *C-PC*s shared some of their decision-making responsibility with the *Ms*. LaTiCE and ISSEP omit *M*, and have three separate titles for the other roles. ACE and Koli have three roles – like LaTiCE and ISSEP, they omit *M* – but they go a step further and assign *C* and *PC* to the same title. And finally, ICSE SEET, as a track of ICSE, only has *PC* and *R*.

ACM CHI is an order of magnitude larger than any of these CER conferences (it received 3126 submissions in 2020 [37], as opposed to the 544 that SIGCSE received [137]). Its organization is correspondingly complex, but briefly, in addition to the General Chair (*C*), they have numerous Subcommittees (each for a different topic). Each Subcommittee has a Subcommittee Chair, some additional Subcommittee members (together performing the *PC* role), and some Associate Chairs. Each paper is assigned more than one Associate Chair. One of the Associate Chairs writes a meta-review and recruits “External Reviewers” (in our terms, an *AE*); the other writes a review (acting as an *R*) [36]. In our notation, this becomes *C* (the general conference organizer(s)), plus several instances of $\{P, AE, R\}$. The *M* associated with other conferences is classified as an *AE* at ACM CHI because *AE*s, unlike *Ms*, recruit their own reviewers. Thus, ACM CHI incorporates features of both conferences and journals.

5.3 Recruiting

In general, the literature gives very little detailed information about the recruitment process for various roles. The mentions are on a general level, for example that the *AE* solicits reviewers for papers, or that the *EIC* invites people for *AE* tasks.

Shah et al. [115], however, discusses the practices in the Neural Information Processing Systems conference in considerably more detail. The sheer size of the conference – like CHI, it receives thousands of submissions – causes significant pressures for good organization. The program chairs at the Neural Information Processing Systems conference lead the process and recruit around 100 *Area chairs* who largely correspond to *AE*s and coordinate reviews in their own topical area. To recruit new *AE*s, the program chairs

invited nominations from two previous years’ *AE*s as well as from the conference’s Board, to enable covering all different topical areas of the conference. The proposed names were discussed in the board, which voted for the list of *AE*s. After an analysis of the number of papers in the different areas in the previous year’s conference, the final *AE* list was compiled. Thereafter, each *AE* was accompanied by a “buddy *AE*” to support in the final decision making in selecting papers in their topical area.

In the next step, each *AE* should invite 30 “senior reviewers” with a PhD, and when accepted, these reviewers were asked to “clone themselves”, i.e., invite another senior reviewer to the area. This pool of senior reviewers was then complemented with a pool of “junior reviewers” who were gathered by asking authors of each submitted paper to nominate one reviewer among themselves. When such pools were available, each *AE* would manually assign one “senior, highly qualified reviewer” for each paper, after which two senior and 3 junior reviewers were automatically assigned to each paper. After reviews were completed, the *AE*’s task was to combine reviewers’ decisions into the final decision, instead of combining the reviews. Buddy *AE*s helped them in the process of making the decisions.

The interviews helped to fill in more details about recruiting. For *PC* and *EIC*, the larger computing education venues now have formal application processes. One interviewee discussed the process of being selected as *EIC*:

Editor 2: [T]hey [previous *EIC*s] stepped down and there was an application process that was run by a board... And so when I saw the opening, I applied. So there’s an application process. There’s a committee that decides. The editor in chief and presumably the committee consists of people who are experts in the field.

Another described the experience of becoming a *PC* as follows:

Chair 4: For our community, the way that Program chairs get assigned is that the ... board runs a process where people can nominate themselves or nominate other people to be the program chair. And then the board deliberates and chooses. ... As Program chair, I applied to be the program chair when they put out a request for volunteers and they chose me.

On the other hand, some smaller conferences have a more informal tradition. For example, in Koli Calling the current chairs discuss together to identify a potential candidate for a new junior chair³. Typically the potential candidate(s) are asked in advance, whether they might be willing to accept the invitation and if a positive response is received, the nomination is discussed and confirmed at the Program Committee meeting during the conference. In some cases, the potential candidates are discussed only in the meeting, followed by the chairs contacting the selected candidate. A clear prerequisite for being a candidate is that the person has attended the conference, knows “its character” and has been a member of the Program Committee for some years. There has been a practice, because the conference is always organized at the same site, and

³Both ICER and Koli currently have a practice that chairs are nominated for two years, so that each year the senior chair steps down, while the junior chair continues till the next year becoming senior, and a new junior chair is nominated.

the chair needs to coordinate actions with the local organization committee, that one of the chairs is selected from Finland while the other one is always from abroad.

The selection process for journal *AEs* and conference *Ms* and *Rs* depends on the venue. Many venues use invitations. One interviewee described the process of selecting *Rs* for a particular conference:

Chair 3: I don't remember ever seeing a call for expressions of interest in reviewing for [our conference]. I imagine that the chairs ask most of the reviewers from preceding years. I imagine they also look at the authors of successful ... submissions in recent years and say, ... that was a pretty good paper. Maybe we should ask them to review ... next year. But so far as I'm aware and as best I can recall, over the many years I've been involved ... the program chairs, the conference chairs have always gone into a huddle and then said ... 'Here's the list of reviewers.'

In sum, the *PCs* keep many of the reviewers from previous years (as can be seen from the public documentation). In addition, authors of good papers that have appeared in that conference (and indeed, other individuals with a strong research record) are likely candidates to be invited.

For journals, sometimes the *EIC* simply selects the *AEs*.

Editor 2: [The publisher] provides a lot of flexibility regarding how the editor in chief chooses the editorial board. I'm given the responsibility of choosing the editorial board. And I really don't get any feedback on that. I just have to make the process acceptable to the research community.

Of course, these invitations are based on careful consideration.

Editor 2: [A]s a reviewer, I learned a great deal from other reviewers. And so if I come across somebody who does a really nice job, I tend to take notice of their names, so that I can invite them in the future because they've done a good job...

For *Ms*, as for other positions, the length of service for a venue is a clear plus.

Chair 7: In the case of [venues I've worked with], I became a meta-reviewer having done the reviewing for many, many years.

On the other hand, poor reviewing or frequently submitting reviews late may well result in not being invited again.

In large conferences, simply inviting *Rs* can seem unworkable:

Chair 1: For [our conference], the pipe, the volume is so huge, that we have no alternative, essentially, but to invite every person who submitted a paper to be a reviewer.

As seen earlier in the discussions of *SIGCHI*, *ISSEP*, and the Neural Information Systems conference, however, there are other models. A conference can define track chairs who act as *PCs* for their own subject area, or even have the *Ms* act as *AEs* and recruit their own reviewers.

Recruiting has implications across the peer review process as a means of addressing some of the key challenges as the field expands. The increasing volume of submissions and hence the need for more reviewers was noted earlier (section 4.4). The need for differing expertise – with respect to topics, methods, and different paper types – increases as the field expands. There is considerable interest in drawing early-career academics into the process, reinforced by the perception that the most 'efficient' reviewers may not be the most senior.

Editor 2: I have made some interesting observations about ... senior versus junior and ... years of experience in the field. There's no correlation between, it seems to be no correlation between that and their efficiency and integrity in doing their jobs. ... some of my best *AEs* are young junior faculty. ... they're go-getters. They ... want this to look good on their CVs. They want to show that they've made an impact, and maybe down the road, then, they can become an editor.

There is also the need to preserve and disseminate knowledge of the review process (as discussed in 4.5).

5.4 Rewarding reviewers

As Kelly et al. [100] note, reviewing is a part of scientists' work. Parberry suggests as a reasonable guideline that researchers should contribute as many reviews to the community as they receive: roughly two times the number of papers they submit [109, p. 105].

But each individual reviewing assignment is voluntary and time-consuming. Tite and Schroter [130], Egghe [53], and Ware [134] all reported workload as a big problem in recruiting qualified reviewers. Our survey of reviewers, although it was not a random sample, did indicate a substantial time investment by the respondents: 65.6% reported spending 45–120 minutes on average to review a 6–8 page conference paper. (16.7% spend less than 45 minutes; 16.6% spend more than 120 minutes.) In addition, as Golden and Schultz [59] noted, reviewers also have other responsibilities like family, salaried jobs, community service etc., which limit their availability for reviewing tasks.

Given all this, when asked "For a conference, how many papers is it reasonable to ask a reviewer to review?", 41.5% of our respondents considered 3 papers a reasonable load; 26.8% considered 4 papers a reasonable load. Computing education conferences are now telling reviewers to expect a substantially larger assignment [see, e.g., 72, 91].

What motivates people to take on this task? And how should they be recognized and rewarded? There is a compelling need to answer these questions, given the growing need for reviewers. In computing education venues, reviewers are not paid. This is consistent with most journals [27, p. 50], and, in response to a survey of reviewers for five different biomedical journals, those surveyed reported that modest financial incentives would not be effective given the time pressure they had. Consulting fees reflecting the value of the reviewers' time might make a difference, but few journals would find that a practical option [130].

There are powerful intrinsic motivations inherent in the task itself. These include learning about current work (even though what a reviewer learns must be kept confidential); gaining perspective

on the review process; the satisfaction of providing a service to the community; and the opportunity to help others to develop their work. One important incentive, however, the opportunity to discuss the current submissions and future research directions in the field, may be disappearing with the scale-up in submissions: [28, p. 35]:

If submissions are read by only a few PC members then there will be fewer broad discussions at PC meetings about the most exciting new research directions. Yet senior PC members often cite such dialogue as their main incentive for service.

Some interesting aspects for the motivation to accept invitations (as in ICER) or to voluntarily enroll as an *R* (as in SIGCSE and ITiCSE) were discussed in our interviews. An obvious motivation is that academic service is among promotion criteria:

Chair 4: But if you have done no academic service to your academic communities, such as reviewing papers, that would be a red flag in a promotion process.

Some argue that the credit given towards promotion and tenure is insufficient, and more credit might make a difference (Street and Ward [127]). Publishing papers, getting funding, teaching, and other work-related responsibilities are competing for the same researchers' time, however. The amount of credit is not likely to justify diverting a lot of time from those other activities.

A more significant item in a tenure and promotion folder would be a position as *M* or *PC*. As noted in our interviews, this kind of service is seen as important:

Chair 4: I was invited to join that program committee, and that was considered, in software engineering, a very, very strong signal that I was a highly respected researcher ... You couldn't just join it no matter what, you had to be invited because you were trusted for your expertise.

A clear career path from experience as an enrolled *R* (at SIGCSE or ITiCSE) to an invited *R* to an *M* or *AE* position might motivate people to review [109, 130].

Social motivations are also powerful. Enhanced reputation and the goodwill of editors are cited by Parberry [109]. The reviewers surveyed in Tite and Schroter [130] responded that being acknowledged in print might influence their decision to review. This is common for journals in other fields. In our documents, *ACM Transactions on Software Engineering and Methodology* (TOSEM) and *IEEE Transactions on Software Engineering* both name their reviewers on their websites [20, 84]. TOSEM also lists members of a Board of Distinguished Reviewers on its website [19]. Computing education conferences routinely list the names of reviewers in their proceedings [2, 33, 40, 73, 74, 85, 92, 116] and often on their websites as well [39, 71, 89, 101, 103]. It has not been the custom for computing education journals to list their reviewers, but CSE is currently considering doing so [42].

Other potentially effective non-financial rewards suggested by the respondents to Tite and Schroter's survey include a free subscription to the journal and more feedback about the outcome of the submission and quality of the review [130]. In addition, Benos et al. report that the American Physiological Society holds a yearly banquet for reviewers at one of its major conferences [27, p. 50].

Some superficial quantitative measures have also been suggested [see, e.g., 56, 59, 99, 127]. These reflect quantity, however, and do not address the quality of the reviews done.

Ghosh et al. [56] proposed motivating reviewers by transforming the peer-review process to one of open reviewing. For further discussion of open reviewing, see Section 6.

5.5 Monitoring review quality

Monitoring review quality was mentioned as one responsibility for people acting in the roles *P*, *M*, and/or *AE*.

Chair 4: How do we evaluate whether reviewers have done their job well? I think that you can certainly look at surface features of their work. ... Did they respond to these different facets that we were asking them to evaluate. ...

I think that this year at [our conference] we tried to ensure that the meta-reviewers were checking the reviews to make sure they were humane and constructive. ... And there were other things like, if they noticed a reviewer that didn't do their job, ... we could delegate the task of nagging that reviewer to finish their work to the meta-reviewer rather than to us.

Chair 7: But one of the things we can do is identify the reviewers who are doing a great job, and those reviewers who are not doing a great job, and provide them with a little bit of feedback afterwards.

As noted above, this is listed by ICER 2020 as one of the responsibilities of their meta-reviewers.

5.6 Workload

One of the effects of the increase in submissions to computing education conferences, in addition to the meta-reviewer role being added to many conferences, is the steady increase in the workload of both reviewers and meta-reviewers.

This increase includes both larger and larger numbers of papers to review and also new tasks. As described in Section 4.3.8, ICER's meta-reviewers evolved from reading papers and meeting with the program chairs; to reading papers, writing reports for the program chairs (but also for the authors), and managing discussion among the reviewers; to all of the above, plus ensuring the quality of reviews. Reviewers at ICER, ITiCSE, and SIGCSE have all been asked to engage in a discussion period, in addition to reading and evaluating papers.

Regardless of the merit of these changes – and they all have positive aspects – they definitely increase both the workload of the individuals involved and the collective time investment of the community in this process. One of our interviewees commented on this situation:

Chair 1: But this whole meta-review and adjunct Program Committee ... we're investing an enormous amount of our community's time in these reviewing processes. In fact, in many of the conferences more recently, I would say that the amount of effort that goes into reviewing a conference paper is greater than what would go into reviewing a journal article.

Chair 1: Well, let's take ITiCSE last year. I did a quick back-of-the-envelope calculation based on the idea that a reviewer would be able to read the paper and write the expected review at the expected level of detail – and let's face it, those expectations are fairly high in those ... communities – in fifteen minutes. So ... based on an assumption that you could read the paper and write the review in 15 minutes, we invested six months of working time, for one individual. Just in producing the reviews. Six months of ... full-time work went into that, in total investment of the community. And let's face it, ... that's an underestimate. ... the likely investment is somewhere between one and two years of effort. Just for the reviewers, in terms of ... the time we've taken from our colleagues, asked them to contribute / donate to this process.

Chair 1: The other thing that I've raised with the SIGCSE committee is what is the purpose of the process? ... no reasonable human being with any other type of academic job can read the number of pages that is produced in that process in the time available to make decisions. Typically, those conferences allow one to two weeks, to make the decisions. Now, let's suppose, though, that the reviews come in two weeks prior to that. So you have four weeks in total. And so for 300-ish, 200 and let's say 250-ish papers that we had for ITiCSE last year, times four reviews, and we got all four reviews for every paper, ... plus a meta-review. That's five. So that is conservatively one to one and a half thousand A4 pages of text. Can you read that in a week? Probably. Can you use it in a high-value way to make decisions? I doubt it.

5.7 Discussion - roles

There is a great confusion on various titles of roles and what they actually mean. The same titles are used for different purposes. This is clearly associated with the independence to organize venues in different ways. Also the size of the conference is important; as conferences grow, that change drives changes in organizational structure. The titles are not as important as the actual role and the responsibilities.

However, to understand the big picture is not trivial. Little public information is available which would tell more of the actual meanings and purposes. All conference websites include titles for those involved in organizing the conference, but few add any information about the responsibilities associated with those titles. For example, we found no description of Conference General Chair, General Chair, General and Organizing Chair, and Symposium Chair for any of the ten conferences we examined.

Sponsoring organizations provide broad guidelines, leaving a lot of details to be filled in. For example, ACM SIGCSE, which sponsors ICER, ITiCSE, and SIGCSE, publishes a fairly abstract statement of "Program Chair Responsibilities" [12]. Despite the page's title, the responsibilities it lists are those of "the Program Chair(s) and Committee". The term "Committee" is not defined, and

the relationship between the Committee and the Program Chair(s) is not specified.

The limitations of the published information became particularly clear when we examined the two conferences outside of CER in our dataset: ACM CHI and ICSE. Our analysis of the CER conferences was supplemented by our own knowledge. On the ACM CHI website, we found extensive information about roles and process [36]. But in sharp contrast, ICSE, like the CER conferences, provides little more than titles, and as a result we were confronted with the gaps in our knowledge.

Moreover, while there is a significant need for people in the various roles, there is little if any accessible information concerning how people are recruited for the roles. Thus, there is a lot of hidden internal knowledge which newcomers in the field might find strange and hard to learn. The lack of transparency and the common practice of using invitations to recruit people can also create suspicions among newcomers that there is a hidden internal community, which might have its own rules, even invisible privileges.

One choice that venues make is how to reward reviewers. While journals usually publish guidelines for reviewers on their websites, they generally do *not* mention them by name. When considering this question, journal editors may find it helpful to see examples of other strategies, such as listing reviewers' names on their websites or recognizing "Distinguished Reviewers". Several other ideas for motivating reviewers are discussed in Section 5.4. Because the field is growing rapidly, especially now that K12 computing education research gains more and more visibility, it would be good to consider what kind of strategies would be appropriate for us as a research community.

Finally, looking more broadly at the role of conferences for the whole field, an interesting perspective was given by one conference chair, when discussing whether high selectivity is a value for a conference:

Chair 1: I think the purpose of conferences is to ... ventilate ideas and new thoughts, and to engage in a dynamic academic discourse with the leading and developing minds in the discipline.

Computer science as a field has its own tradition, in which conference papers are submitted and evaluated as full papers, and they are considered very important merits in academic life. In other fields, journals are the forum for publishing, and conferences are for meeting people, presenting new ideas, and getting feedback. While we cannot dig into this tradition and the reasons behind it in this paper, it worth considering whether we should aim more at exclusion (low acceptance rate) or inclusion (higher acceptance rate) in conferences.

6 ETHICS AND ETIQUETTE

The research question addressed in this section is:

What are the ethical issues associated with peer review, and what is the customary code of behavior?

As noted elsewhere in this report, peer review is crucial in an academic discipline; the results are important both for authors of manuscripts and the development of the discipline itself. Peer review should evaluate work objectively on its scientific merit and

clarity of presentation, and it is important that stakeholders believe that is the case.

There are a number of ethical issues surrounding this process that have been identified in the literature; we will focus on these: bias, conflicts of interest, confidentiality, civility, timeliness, transparency, qualification (of reviewers), and redundancy; these are identified in the literature as potential problems. Souder [126] and Rockwell [111] provide broader overviews.

The following terms will guide our analysis of the general research literature and of the concrete practices in computing education research:

Bias is when a reviewer uses factors irrelevant to the quality of the manuscript in making their decisions; for example, a reviewer may tend to give lower scores based on gender or geographic origin.

Conflicts of interest are when a reviewer has reason to not be objective when a certain result provides him or her some benefit. These benefits might be tangible, like favorably reviewing a project for which you are a consultant, or intangible when reviewing something by a relative, friend, or collaborator.

Confidentiality means that a reviewer cannot use any of the findings of a paper under review, nor discuss that paper with anyone but the program chair or editor that assigned the review.

Civility is an expectation that reviewers communicate professionally and politely in their reviews.

Qualification is the expectation that the reviewer has the technical knowledge to assess the quality of the manuscript and its contribution to its field.

Redundancy concerns the practices of submitting the same work to multiple venues, or submitting papers that are very small extensions of previously published work. It also includes plagiarism, submitting someone else's work as your own. While this issue deals with author ethics, it may be part of a reviewer's job to detect it.

Timeliness is the expectation that the reviewer finish the review on time.

Transparency is the degree to which the review process is "visible" to the author, or in some cases, to the readers.

The following sections discuss these ethical issues (and some ways in which they may be avoided). Each section first analyzes relevant general research literature, and then summarizes the policies and practices that are documented for computing education research venues, and considers how they compare to the literature.

6.1 Bias

6.1.1 Literature on bias. In their review of a small number of studies on biases in reviewing, Benos et al. [26] note inconclusive findings; see also Souder [126]. Even though institutional bias, i.e., bias based on the reputation of the authors or their institution, appears to be present, no statistically-significant difference was found in the acceptance rates for regular papers studied in the literature reviewed. However, comparing single-blind and double-blind reviewing, Tomkins et al. [131] found evidence that reviewers in single-blinded conferences not only tended to be more likely to bid for papers authored by researchers with high standing or from highly-regarded institutions, but also were more likely to recommend such papers for acceptance; such a tendency was not found in double-blind reviewing. The authors raise the question of how to

weigh the potential benefit of using prior knowledge to better assess the quality of a research paper, over the dangers of disadvantaging other work. They suggest that, despite the lack of a deep understanding of the implications of either reviewing practice, one should "seriously consider the advantages of double-blind reviewing" [131, p. 12712].

Similar results were reported for gender-based bias or the lack thereof. Gilbert et al. [57] conducted a retrospective study within the context of a major medical journal. While certain aspects of the reviewing process appeared to be gender-biased, such as assignment of papers to editors of the same gender as the corresponding author, and male reviewers accepting review requests from male editors more frequently, no significant difference in acceptance was found with respect to the genders of corresponding authors.

On the other hand, Lloyd [105] reported on an experiment in which she produced a fabricated paper, and sent it to reviewers of five behavioral journals, asking them to review it as they would with other submissions. She sent each reviewer one of two versions of the manuscript that differed only in the gender of the authors. Her analyses showed that female reviewers were significantly more likely to accept the female-authored paper than the male-authored paper, that the likelihood of the female-authored paper being accepted by female reviewers was significantly higher than the likelihood of it being accepted by male reviewers, and that the likelihood of male reviewers accepting the female-authored paper was not significantly different from the likelihood of them accepting the male-authored paper.

Researchers have attempted to formalize the modeling and analysis of bias. Avin et al. [22] modeled co-authorship between PC members of several high-ranking ACM and IEEE conferences and authors of papers accepted to the respective conferences by means of a (social) network. They then investigated the so-called *coverage bias*: What is the ratio of the number of authors connected to at least one PC member and the number of authors? The authors then compared these against benchmark ratios and contrasting numbers over a variety of randomly-generated graphs. Acknowledging the challenges of creating an "appropriate" set of random graphs for comparison, the authors concluded that, under their measure, most conferences were not biased toward accepting papers written by collaborators of PC members, and that those for which a bias could be computed exhibited a very minor tendency.

Building on assumptions about the cognitive load incurred by reviewing a paper, García et al. [55] hypothesize that reviewers may suffer from confirmatory biases for papers that are particularly hard to parse. They used conditional probabilities to model the reviewing process as the optimization of a trade-off between the value and the cost of obtaining review information. Based on this model, they conclude that reviewers who, for whichever reason, start with a low prior assessment of the quality of a manuscript are less likely to respond to positive signals or even actively search for such signals, the more the cognitive complexity of obtaining such information increases. Similarly, their theoretical model suggests that, in situations in which the loss of accepting a poor-quality paper is less than the gain of accepting a high-quality paper, low prior expectations and high cognitive reviewing load also tend to lead to a reduced probability of positive reviews.

Bias appears to play a major role in authors' perceptions of the reviewing process. Responding to an anonymous survey by Resnik et al. [110], roughly half of the survey population reported having experienced bias in peer reviewing. Even when keeping in mind the subjective nature of such an opinion survey, negative opinions regarding the reviewing process can severely affect a journal's or conference's standing. Thus, the literature agrees on strongly recommending that reviewers guard themselves against bias [27, 48, 109, 111, 125].

6.1.2 Policies and practice on bias. As discussed above, actual or perceived bias may severely affect both authors and publication venues. All computing education research conference series considered in this study employed at least a double-blind reviewing process in order to guard themselves against the most obvious sources of institutional or gender-based biases. Nonetheless, reviewer guidelines point out explicitly the importance of unbiased reviews and require that reviewers excuse themselves from reviewing if they feel they cannot give an unbiased review [123]. The same holds for journal publications, with the additional safeguard that editors are given the responsibility "for checking for potential reviewer biases, rather than relying solely on declarations" [128]. Acknowledging the subtleties of discerning between different educational contexts, ITiCSE attempts to alleviate biases based on educational or national contexts by geographically balancing reviewers [87]. Finally, the reviewer training material offered for ICER covers other types of biases that may arise from the use of different research orientations or methods from other scientific communities [70].

6.2 Conflicts of interest

There is almost complete unanimity that, where there is a (perceived) conflict of interest (COI), a reviewer must raise the issue to the editor [24, 27, 48, 109, 111, 125] and should not review the paper [27, 111, 125].

6.2.1 Literature on COIs. Conflicts of interest can occur in various forms. Some are due to professional relationships, including current and recent collaborators; some are due to personal relationships, including friends, family members, and "people you detest" [111]. Souder, discussing conflicts of interest in his review of the literature on peer-review ethics, notes that "Financial conflicts of interest were front and center" [126].

Another conflict may arise when asked to review work that is close to one's own. Rockwell [111] points out that it puts the reviewer in an untenable position, as rejecting the paper may give the appearance that the reviewer was trying to gain an advantage in publishing the reviewer's own results. Parberry [109] discusses the ethical dilemma caused by reviewing a manuscript that proves something that the reviewer has recently proved, but not yet written up and submitted for publication. In Resnik et al.'s survey of scientists about the ethics surrounding peer review, 9.6% of the respondents believed that "A reviewer delayed the review so that he/she could publish an article on the same topic" [110].

Duchesne and Jannin [52] propose a reviewing workflow that has checking for potential conflicts of interest as the first action by potential reviewers.

6.2.2 Policies and practice on COIs. We investigated a number of major computing education research and a few non-computing education research conferences and journals for the past couple of years, when the data was available online, to garner an understanding of their policy on conflicts of interest. The policies highlight how the potential implications of conflict of interest have led to constraints on submission by those who make publication decisions (i.e., chairs and editors), in some venues. The policies for the computing education research venues are described in turn:

ACM-sponsored conferences. The ACM policy on conflict of interest states [10]:

Possible conflicts should be reported to the individual in charge of the venue involved (e.g., editor or program chair). In the case that the individual in charge is also conflicted, then conflicts can be reported to a predesignated alternate or to the ACM Director of Publications. A best practice would be for there to be one or more predesignated alternates identified and publicized in advance of any potential need for reporting [...]

In all cases of identified review COIs, it is best practice for the alternate decision-maker to "blind" any conflicted parties (e.g., reviewers, editors) as to who has assumed their roles. If the conflicted party is in a lead decision-making role (e.g., EiC, PC chair), s/he should designate an appropriate colleague (e.g., Associate Editor, co-chair) to appoint another person to the decision-making role. Using this process, the third person remains anonymous from all conflicted parties, when feasible.

The SIGCSE Board's policy is more restrictive: it states [13]:

Program chairs and Conference chairs may not submit any scholarship (papers, posters, special sessions, panels, lightning talks, working groups, etc) to their conference, but their students are permitted to do so.

ICER reviewers cannot review a paper if either they or their students co-authored the paper, or they identify themselves to have conflict of interest. Reviewers, meta-reviewers, or program chairs with a conflict of interest cannot participate in any reviewing activities such as evaluation, discussion, and the decision process. It is the responsibility of program chairs to ensure that no person with a conflict of interest is involved in the reviewing process [72]. From 2017-2019, ICER program chairs were prohibited from submitting papers, a restriction that was relaxed in 2020 [67, 69, 72].

We were able to find information for the SIGCSE symposium for 2015, 2019, and 2020

For the SIGCSE Symposiums in 2015, 2019, and 2020, we were able to find information that stated clearly that conference chairs and program chairs were not permitted to submit any type of the work to the conference. Further, the guidance required reviewers to contact the program chair if they thought they could not provide a fair and unbiased review [13, 118, 121, 123].

ITiCSE [90] refers to ACM's conflict of interest policy [10] (quoted earlier), which gives a comprehensive definition of what constitutes a conflict of interest and recommends an "alternate" to be identified

when a decision-maker has a conflict of interest.

Other venues. *Computer Science Education* requires reviewers to declare conflicts of interest to the editor [128]. CSE is published by Taylor and Francis, which does not exclude anyone from submission to its journals [129].

ACE reviewers are required to declare conflicts of interest. There is no exclusion for submission due to a conflict of interest. If the conference chair, however, declares a conflict of interest for a paper, it is the responsibility of the co-chair to manage the review process for the paper in question [1].

Journals and conferences have different policies on whose responsibility it is to declare and deal with conflicts of interest, often sharing responsibility among all involved. For instance, IEEE requires editors and reviewers to disqualify themselves if they have a COI [78]; ACE defines this as the responsibility of the reviewers [1]; TOSEM expects the editor-in-chief to identify possible conflicts of interest before assigning a paper to a reviewer [18]; while JLS requires the authors to disclose any COI or “competing interest” that could have influenced their research [98].

6.3 Confidentiality

6.3.1 Literature on confidentiality. The papers we read that discussed confidentiality were fairly consistent in their views. Most of the papers [27, 48, 109, 111, 125] agreed that a reviewer is not allowed to use ideas or data from a paper that he or she is reviewing before it is published. Most of these agreed that it is inappropriate to share the manuscript with anyone else without editor permission [27, 48, 111]. Some mentioned that a reviewer was not allowed to reveal the result of the review [109, 111], or even that a paper had been submitted [27, 109]. Smith [125] provided an exception, namely that a reviewer could use the ideas and distribute the paper if it was already publicly available (for example, as a Technical Report).

The COPE Council wrote guidance for editors, in which they suggested that editors make reviewers aware of their confidentiality responsibilities, and that editors have systems in place that guarantee the confidentiality of manuscripts that have been submitted [45]. Bailey et al. [24], in their proposed code of conduct for peer review, presented (different) confidentiality responsibilities for authors, reviewers, and editors, with an emphasis on the blindedness of the review process.

In Resnik et al.’s survey about ethical problems in peer review [110], 6.8% of researchers reported that they had experienced a reviewer breaching confidentiality, and 4.5% experienced a reviewer using their ideas, data, or methods without permission.

The focus in these papers is strongly on the reviewer, less on the editor, and nearly absent on the author. Most of the considerations regarding confidentiality are about inappropriately using ideas in the manuscript, which clearly exempt the author, or sharing the manuscript, which largely exempt the author, unless that sharing is being done to ‘game’ the review process, for example by distributing their manuscript for comments to potential reviewers, as described by Bailey et al. [24].

6.3.2 Policies and practice on confidentiality. The online materials associated with our venues do not discuss confidentiality associated

with peer review, but they provide (sometimes multi-hop) links to materials provided by their parent organizations that do – ACM for SIGCSE, ITiCSE, ICER, and TOCE; IEEE for *IEEE Transactions on Education*; Taylor and Francis for *Computer Science Education*. These organizations present the following policies.

The ACM policies are given in its *Policy on Roles and Responsibilities in ACM Publishing* document [11]. This document discusses the roles and responsibilities as a set of expectations: what ACM expects of authors, reviewers, and editors, and what authors, reviewers, and editors can expect from ACM. The expectations of confidentiality are:

- ACM expects reviewers to not use any results from the submission, to not disclose its existence or status, and to not share it with anyone without the editor’s permission;
- authors can expect ACM to keep the submission confidential;
- reviewers can expect ACM to not disclose that they reviewed this manuscript, but to “Acknowledge their efforts in the publication process”, and
- editors can expect ACM to make sure that manuscript tracking systems provided by ACM ensure confidentiality.

The main IEEE document regarding publication is the *IEEE Publication Services and Products Board Operations Manual 2020* [78], which covers many aspects of IEEE publication in detail. This manual states that IEEE expects anyone having access to submitted manuscripts to not “make any inappropriate use of the special knowledge that access provides”, nor to share it with others. Referees and editors should treat manuscripts as confidential documents. Specifically, reviewers should not use “non-public information contained in an article to advance their own research or financial interests”.

The IEEE Computer Society’s *Reviewer Information for Journals* also provides reviewer conduct guidance [80], that reviewers should assume submitted manuscripts are not meant to be public, to not use or share material from a manuscript they are reviewing, and to not distribute copies of said manuscript.

The Taylor & Francis *Peer Review Guide* [129] provides ethical guidelines for peer reviewers, including that they keep the process confidential and “not share information or correspondence about a manuscript with anyone outside of the peer review process”. Taylor & Francis’s *An editor’s guide to the peer review process* [128] includes a section, “The ethics of peer review” with best practices for editors, including that they ensure confidentiality and not provide details about submissions to anyone other than the reviewers without author permission, and that they make sure that reviewers know what to expect in the review process.

6.4 Civility

6.4.1 Literature on civility. Even though civil discourse should be a scholarly norm, the blindedness of most review processes appears to tempt some reviewers – and, as pointed out by Hadjistavropoulos and Bieling [61], also dissatisfied authors – to direct attacks *ad hominem* or to raise unsubstantiated accusations; see Hadjistavropoulos and Bieling [61] and Souder [126] and the references therein.

Unsurprisingly, there is a unanimous consensus in the surveyed literature that reviews need to be courteous and helpful [21, 24, 27,

109, 111, 125]. A study by Gosden [60] found reviews that recommended “acceptance pending minor revisions” to almost always be helpful regarding the particular type of revisions needed, indicating that reviewers were genuinely trying to help the authors improve the paper. On the other hand, less than half of the more negative reviews showed the similar levels of mentorship and helpfulness.

Signing the reviews or, more generally, open peer review, can be seen as one approach to mitigate issues with civility. In a randomized control trial comparing two groups of reviewers, the editors of the *British Journal of Psychiatry* found reviewers in the groups that had agreed to sign their reviews to be more courteous in their tone but noted that “the majority of reviews were at the courteous end of the scale in both groups” [133, p. 49]. The authors noted, however, that moving to a system exclusively consisting of open peer review would likely result in loss of access to a non-trivial amount of reviewers who felt uncomfortable signing their reviews.

Hadjistavropoulos and Bieling [61] point out that the relevant professional organizations – in their case, the Canadian Psychological Association and the American Psychological Association – already have institutionalized ethical standards. They conjecture that reviewers, while being aware of professional norms in their domain when it comes to research, need to be reminded that the same norms need to be upheld for reviewing. In short: “Your review should be directed at the paper, not at the author” [125, p. 66].

6.4.2 Policies and practice on civility. Given the consensus about the importance of civil discourse and of addressing reviews at the work itself and not *ad hominem* (see Section 6.4.2), it is unsurprising that almost all of the major computing education research publication venues we considered emphasize these requirements for reviews [8, 14, 70, 75, 123, 129]. SIGCSE reminds reviewers explicitly that constructive reviews “help the authors, and in the long run, the conference” [123]. SIGCSE also explicitly discusses how to deal with recalcitrant reviewers [124]. In contrast to conferences, journals may have several rounds of review. Both CSE and TOCE ask their reviewers and associate editors to not only engage in civil discourse but to make sure that comments are actionable (defined as “providing concrete suggestions for how to improve the paper before it is resubmitted”) [14, 129]. In conclusion, the computing education research venues considered in this study have policies and recommendations in place that are in alignment with the general literature.

6.5 Qualification

6.5.1 Literature on qualification. As can be seen from the term “peer review” itself, the process assumes that reviewers have sufficient technical knowledge to assess the contribution of a submission. Reviewers are usually selected and approached by the editor or chairs [45]. The ultimate responsibility, however, of judging whether or not they are competent to review a paper lies with the reviewers themselves [21, 24, 27, 48, 111]. Dispute of a reviewer’s competence to properly assess the paper has been cited as the most frequent reason for unhappiness among authors [110]. While Resnik et al. concede that this “may simply reflect scientists’ frustration with having manuscripts rejected or critiqued” [110, p. 308], repeated accusations of reviewer incompetence are likely to affect the standing of a journal or conference.

6.5.2 Policies and practice on qualification. In general, the documents considered for this study refer to the reviewers’ qualifications only implicitly, e.g., by stating that “papers shall be reviewed by at least three *qualified*, independent reviewers” [emphasis added] [6]. As part of the reviewing process for journals, the handling editors invite reviewers based on their experience [8, 78, 129]. Similarly, for conferences where there is a bidding process (e.g., ICER), attempts are made to match paper topics and reviewers’ expertise [72]. Formal requirements, such as having a doctorate in a relevant discipline, are not mentioned explicitly, with the notable exception of ICER, for which reviewers are required to have “a PhD in CS, CS Ed, Ed, or related field, or equivalent research experience” [69]. Our interviews with editors and chairs, showed that research experience is effectively a necessary condition for becoming a reviewer.

6.6 Redundancy

6.6.1 Literature on redundancy. There are multiple forms of redundancy, both those based on one’s own work (e.g., multiple simultaneous submissions, or unacknowledged or unauthorized re-use) and those based on misuse of others’ work (i.e., plagiarism). The COPE Council [45] includes “encouraging reviewers to comment on the originality of submissions and to be alert to redundant publication and plagiarism” in their best practices for journal editors, and further suggests that editors should use plagiarism detection software and support authors who have been victims of plagiarism.

Plagiarism is the practice of taking someone else’s work and presenting it as one’s own, and is considered a major ethical violation. Benos et al. also use the term to describe the situation in which a reviewer takes advantage of the idea or data that is presented in the submitted paper [27] (a violation of confidentiality discussed in Section 6.3).

While plagiarism is the use of someone else’s work, there are forms of redundancy based on an author’s own work. Such practices include *shotgunning*, submitting a paper simultaneously to more than one journal or conference without notifying all the program chairs or editors, and *duplicate publication*, a “paper which reports the same data as another previously published paper to answer the same or similar research question or test the same hypothesis” [108]. Such practices have been referred to as *self-plagiarism*, although that term is used differently by different authors [104]. *Salami-slicing*, breaking down a study into as many papers as possible (also known as *least publishable units*), can involve redundancy as well, as there may be text shared by multiple papers. Jefferson [94] describes such redundancy as legitimate as long as the related papers clearly cite the others, and the authors make no attempt to mislead the editor and readers.

Redundant publication can have a number of bad effects. Norman and Griffith [108] note that it can exaggerate the weight of redundant results in meta-analyses, reduce the space available for other papers, and waste reviewer time. Schulzrinne [113] adds a few more, including authors receiving “double credit” for the same work, and listener boredom from re-runs at conferences. There might also be copyright violations [50].

Lin [104] recommends a formal definition and criteria for self-plagiarism, including the level at which self-plagiarism should be handled (e.g., by editors). Similarly, Schulzrinne [114] recommends

that publication venues should provide clear guidelines for double (or concurrent) submissions.

It should be noted that redundancy is not necessarily unethical. Smith [125] notes that simultaneous submission is permissible given that the editors or program chairs have given their permission. Norman and Griffith [108] point out that it is not always possible to report all of the results of a large study in a single paper. COPE Council [46] provides advice to editors on redundant submissions in the form of a flow chart that illustrates levels of severity and alternative responses.

6.6.2 Policies and practice on redundancy. ACM provides a comprehensive guide on plagiarism, misrepresentation, and falsification consequences, how to report it and how to appeal the decision [9]. This guide provides an explicit definition of self-plagiarism:

Self-plagiarism is defined as the verbatim or near-verbatim reuse of significant portions of one's own published work without citing the original source. Note that self-plagiarism does not apply to publications based on the author's own previously published work (e.g., appearing in a journal or conference proceedings) if an explicit and appropriate reference is made to that prior publication.

ACM requires that a publication based on previously published work must have at least 25% new material, and that the author must notify the editor about the other paper [4].

ACM uses plagiarism detection software to ensure that papers that are sent out for reviewing are free from plagiarism [8].

The ACM's authorship policy was changed in 2018, to provide greater accountability. According to this new policy, being listed as an author requires that an individual "made substantial intellectual contributions to some components of the original work described in the manuscript", "participated in drafting and/or revision of the manuscript", "[is] aware the manuscript has been submitted for publication", and "agree[s] to be held accountable for any issues relating to correctness or integrity of the work" [5].

While ACM, as a publisher, requires the work to be 25% original, some of its venues follow the same policy such as the *ACM CHI Conference on Human Factors in Computing Systems* [35], and some override it. For example, TOCE requires that 30% of the content be original [16]. SIGCSE requires reviewers to let program committee know if the work is not original, but it does not specify the degree to which the paper may contain duplicate material and still be considered for publication [121]. ITiCSE seems to be silent about this issue, as we were not able to find any information online in this regard.

IEEE requires the manuscript to be original and neither have been published and simultaneously submitted somewhere else, nor be plagiarized or falsified [77]. It also allows for the publication of evolutionary work, providing that all the previous and current works have been peer-reviewed, and the current work is substantially different from the previous work. However, it does not define what accounts as a "substantial" enhancement of the paper [76]. While ACM [8] and IEEE [76] screen submissions for plagiarism with a software tool, Taylor & Francis requires reviewers to inform the editors if they detect similarity between the paper that they are reviewing and other papers that they know [129].

6.7 Timeliness

6.7.1 Literature on timeliness. Delayed reviews can negatively affect the publication process and the dissemination of knowledge [26] and have been found to be a major source of unhappiness among authors [110, 135]. In consequence, there is general consensus that reviewers should only accept review requests if they are confident that they can meet the reviewing deadline [27, 48]. After having accepted a review, "reviewers should complete their review agreements in a timely fashion" [24, p. 70]. The COPE Council [48] additionally asks that reviewers should respond promptly to review requests even if they cannot honor the request.

6.7.2 Policies and practice on timeliness. All conferences considered have fixed timelines for the submission and reviewing process (see Figure 2) which are communicated clearly in the respective calls for papers. Journals as well ask for reviews to be completed on time [8, 14]. Acknowledging the length and depth of journal submissions [109] as well as the scarceness of exceptional reviewers, journals are also more flexible to accommodate constraints on the side of reviewers [18, 48]. Nonetheless, reviewers are asked to accept invitations for reviews only if they are confident that they can complete the task in the agreed-upon amount of time [75, 129].

6.8 Transparency

6.8.1 Literature on transparency. Transparency in peer review is generally considered good, but it can mean a variety of things. At one level, explaining the review process on a web page makes peer review more transparent, as the author knows something about the decision process. Most of what we discuss here concerns how public the reviews and reviewers are.

The COPE Council [47] defines a number of peer review models. A subset of these, in order of increasing transparency is

double-blind, where neither the author nor the reviewer knows the other's identity;

single-blind, where the author does not know the reviewer's identity but the reviewer knows the author's identity;

open, where both author and reviewer know the other's identity, and

transparent, which makes the reviews (but not the reviewers' names) publicly available when the paper is published.

These uses of the terms *open* and *transparent* peer review are not the same as used by all authors, so we shall compare features: the difference between single-blind and open review here is whether the reviewer is identified to the author (hence open review is also known as "signed reviews"), and the difference between open and transparent reviews is whether the reviews are publicly available (hence transparent review is also known as "open reports").

Jamali et al. [93] surveyed a diverse set of early-career researchers about their beliefs concerning peer review, and found that, when writing reviews, 47.9% of the respondents preferred double-blind review, 18.1% preferred triple-blind review (where the editor, reviewers, and authors are each unknown to the others), and 7.3% preferred single-blind review. On the other hand, only 10.7% favored having the reviewer's name published, and 2.9% favored having their reviews (but not names) published. The strong preference for anonymity may be due in part to these people being relatively junior.

While most of the research literature studied was found to be in favor of double-blind review, quite a few papers suggest that the effectiveness of blinding should not be overestimated: if it is not done carefully, the identity of the author can be revealed by self-citing [26] or proprietary software used for the research [113].

There has been a good deal published on the effects and desirability of signed reviews. Walsh et al. [133] conducted a randomized trial of signed vs. unsigned reviews for the *British Journal of Psychiatry*. They found that signed reviews were of higher quality, that they took longer to complete, and that they tended to accept papers at a higher rate. Godlee et al. [58] ran a randomized trial based on signed vs. unsigned and double-blind vs. single-blind reviews (resulting in four possible conditions). The authors then introduced eight errors into a manuscript that had been accepted to the *British Medical Journal* and sent the altered manuscripts to 211 regular reviewers for the *British Medical Journal* that had been randomly assigned to groups corresponding to the four conditions. They concluded that the performance of all the groups (measured by the number of identified errors) was about the same, so that neither changing to double-blind nor changing to signed reviews would have much effect at the *British Medical Journal*. In an overview, Souder [126] reports that the studies comparing signed and unsigned reviews have mixed results.

Ethically, the case can be argued in either direction: signed reviewing is more transparent and can make reviewers more accountable; unsigned reviews can protect junior researchers reviewing senior researchers and may reduce bias. Benos et al. [26] summarize the results in the literature: signed reviews make authors feel better, but the quality of the reviews is no better, and it is more difficult to find reviewers. Aleksic et al. [21] propose an “Open Science Peer Review Oath”, the first item of which is “I will sign my name to my reviews”. The rationale is that reviewers should work with authors to improve their research in a collaborative fashion.

While signed reviews provide reviewer names to the authors, some advocate that even more information go to the public. Wolfram et al. [136] discuss what they call open peer review as peer review that has either *open identities*, where reviewer names for accepted papers are made public, or *open report*, where the reviews for accepted papers are made public. They report the growth in journals using open peer review across a number of disciplines.

Bravo et al. [31] analyzed a pilot study about the effects of open report in journal review. In this study, five Elsevier journals agreed to publish all of the reviews for each accepted article; reviewers could choose to make their names public with their reviews. Data were collected both before and during the years that the reviews were published, based on 9,220 submitted manuscripts and 18,525 reviews. What they found was that making the reviews public had no significant effect on reviewers’ willingness to review, turnaround times, or recommended decisions, and that male reviewers tended to write more constructive reviews. Only 8.1% of the reviewers, however, were willing to make their names public with their reviews.

6.8.2 Policies and practice on transparency. There is a strong preference among computing education research venues for double-blind reviewing. All of the core conference venues use double-blind

reviewing, information that is readily available in calls for participation and conference web sites [1, 34, 72, 90, 123].

In terms of journal venues, TOCE and CSE use double-blind reviewing [16, 43], while *IEEE Transactions on Education* uses single-blind reviewing [81]. ACM Transactions and journals generally use single-blind reviewing [7], as do IEEE Computer Society journals [79], and as did TOCE until April 2017 [64].

6.9 Training reviewers: An effective remedy?

6.9.1 Literature on training reviewers. Given not only the variety of ethical issues that may arise during peer review but also their subtle nuances and their potential impact, it does not come as a surprise that Benos et al. recommend to institute formal reviewer training at least for journals [26]. The COPE Council sides with this view, and encourages mentoring junior reviewers [48].

Houry et al. [63] performed a randomized trial study with reviewers for *Annals of Emergency Medicine* to determine the effects of mentoring on junior reviewers. They found that providing a formal mentoring process for new reviewers did not lead to statistically significant improvements in the quality of subsequent reviews. They concluded that, rather than mentoring, “careful and permanent monitoring of reviewer performance” [63, p. 6] should be established as standard practice.

6.9.2 Policy and practice on training reviewers. Several computing education research publication venues considered in this study have some mechanism in place to help train reviewers: SIGCSE lists examples of “good” and “substandard” reviews in its reviewer guidelines [123], ICER provides an online document for self-study [70], and Taylor & Francis, the publisher of CSE, offers both examples of “good” and “substandard” reviews and formal reviewer training [129].

6.10 Ethics issues raised in interviews

The ethical issue raised most in the interviews concerned transparency.

Chair 4 considers the tradeoffs between transparency and civility:

Chair 4: Social forces that I’ve talked about only function if there’s transparency to reveal the identity of people. And so things like double-blind review remove some of those social forces... That’s good in some ways, right, it can remove some bias, but it’s bad in others, because it removes incentives to be humane.

Chair 1 sees a downside with anonymous reviews:

Chair 1: I see, as a general chair and as an editor of some journals, very destructive reviewing practices and very destructive reviews, where people have written things that they would never, ever write if they thought that the person receiving that review would know who they were. And I think there’s a real lack of professionalism and accountability in that system that would be improved, in fact, by reviewers having to say who they were.

Chair 7 discusses his own experience with public signed reviews. On the positive side,

Chair 7: I have done some reviews for journals where their review has been public and your name is attached to it. And it made me much more careful about the quality of the review that I did because I didn't want to have my name associated with something of low quality that was now visible publicly. And that's a good thing. It made my review better.

However, the visibility had other effects,

Chair 7: But one of the consequences of that is that I was also perhaps hedging what I said more than I normally would, because you never know where these authors will end up.

Finally, public signed reviews may be problematic for the reviewer:

Chair 7: And if they know and associate that review with an individual person, that could potentially damage [someone's] career. ... And I think what would happen is the natural consequence that people would not give honest reviews. So I think the double-blind, while it's flawed and it has some problems ... On the balance, I think it's a better way to go.

This is consistent with the survey results - in which 70.2% of respondents agreed that, "Overall, the trade-off between workload and benefits makes double-blind valuable",

7 SUMMARY AND FUTURE WORK

The goal of this working group was to map the landscape of peer-review practice in computing education research and to seek insights about what influences decisions about the process and criteria. This report has drawn on a number of sources (documents, literature, interviews with program chairs and journal editors) to characterize current practice.

A number of themes have emerged and been discussed throughout the paper, including:

- *The relationship between the venue identity/purpose/values and the review process and criteria.* For example, a venue that identifies itself as archival may have a higher expectation for rigor, whereas one that identifies itself as community-building may have a higher expectation of the likelihood that the paper will prompt discussion.
- *The impact of scale* (i.e., volume of submissions) on all aspects of the review process, including sub-processes, recruitment, workload, and criteria.
- *The perception of reviewers as a valuable (and finite) resource.* Editors and chairs want to treat good reviewers well, so they have the capacity to write good reviews. Nevertheless, their attitudes about reasonable workload vary. The workload assigned to reviewers, and particularly meta-reviewers – both the number of papers and the number of tasks related to reviewing those papers – has increased sharply and continues to increase. Reviewer workload is complicated further by the variable reliability of reviewers in returning reviews, potentially prompting increased review requests.

- *The audiences for whom reviews are written:* Reviews have multiple readers (author, editor, meta-reviewer, other reviewers...) who read with different purposes, making it challenging to write a good review. The multiple readers also influence how the review form is structured, how much time is allocated for decisions, and perceptions of who reads what. A related issue is the way that most venues provide multiple descriptions of paper and review criteria, each aimed at a different audience, and the challenge of keeping all the descriptions consistent.
- *The multitude of names that are used for the different roles involved in peer review,* particularly at conferences, and the way that translating those names into a canonical form can help to reveal the structure of a given venue, and the commonalities of structure between venues.
- *The value of reviewing to reviewers.* Traditionally, computing-education research venues have relied on motivations such as prestige, contribution to promotion, paying back to the community, insight into the way in which the community assesses research, awareness of novel work, and inherent satisfaction in the contribution. Few reviewers are compensated directly, and some argue that academics are already paid, as service is an element of their job description. We discuss several other possible rewards that might be worth considering, especially given the growing need for reviewers.
- *Computing education research is a relatively small community,* with the same people involved in/taking roles in multiple venues - and hence transferring practices between venues and homogenizing the review processes.

The focus of the working group was on providing an informed basis for making choices about peer-review (*not* providing recommendations). Some of the things we observed that we found particularly interesting or informative have implications for the research community. Some have to do with not assuming common knowledge, but instead making systems explicit, for example:

- *Recruiting:* Information about the (sometimes mysterious) recruiting process for the various roles involved in peer review – suggests that the routes into peer review, and the way the various roles are conceived for a particular venue, could be made clearer, particularly for people new to the community.
- *Meta-review:* The evolving role of the meta-reviewer suggests that the research community should recognise that there is a variety of interpretations of meta-review, and each venue should consider which interpretation it might choose, and why (given its particular goals and priorities), and make that explicit.

Some observations concern challenges to take into account when making decisions about how to conduct peer review, for example:

- *Discourse:* Peer review is part of a discourse, and hence disagreement can be constructive and useful. Hence reflection and discussion during meta-review are valuable, but convergence on a recommendation might not be appropriate.
- *Reliance on informed judgment:* Peer review relies fundamentally on informed (if not expert) judgment and on an

engaged discourse, and this cannot be replaced with procedures. Hence, checklists are helpful tools, but they should not become mechanisms or obligations, and they should not replace the recognition of informed judgment.

- *Fit to venue*: The relationship between the purpose and identity of a venue and how it conducts peer review is crucial. Hence, it makes sense for venues to avoid adopting practices just ‘because they’re what other venues are doing’, and to look instead at the fit to the venue’s mission (e.g., in terms of goals, priorities, costs).

Some observations concern broader issues that need more attention.

- *Cost/benefit*: Many interviewees referred to the investment of human resources in peer review, and questioned whether too much is expected in some contexts. Hence, we need to understand better the cost-benefit balance of different peer-review practices, and how to strike the right balance for a given venue.
- *Open review*: The tension between anonymization (to reduce bias and empower less-established academics) and open review (to build discourse and hold reviewers accountable) was highlighted. This intersects with the challenges of recruitment, and of engaging reviewers with a venue’s values, goals, and priorities. Hence, we need to understand better both the goals and the costs of anonymization and to engage with the potential of open review.

This work provides a strong foundation for community discussion, for developing appropriate guidance and training for reviewers, and for a deeper exploration of the reviewer experience; these remain as future work.

ACKNOWLEDGMENTS

Our thanks to the editors and chairs who engaged so frankly in the interviews, and to the members of the community who responded to the survey. Thanks also to the Princeton University Department of Computer Science for funding to assist with transcription, to the reviewers for their detailed comments, and to the ITiCSE organizers who made the Working Groups possible in a difficult time.

REFERENCES

- [1] ACE. 2020. Call for Participation. Available at <https://aceconference.wordpress.com/call-for-participation/>. Accessed 2020-06-17.
- [2] ACE. 2020. Program Committee. In *Proceedings of the Twenty-Second Australasian Computing Education Conference (Melbourne, VIC, Australia)*. viii. Available at <https://dl.acm.org/doi/proceedings/10.1145/3373165> (Accessed 2020-09-04).
- [3] ACE. 2020. Reviewers. Available at <https://aceconference.wordpress.com/reviewers/>. Accessed 2020-06-17.
- [4] ACM. [n.d.]. ACM Publications Policies and Procedures. Available at <https://www.acm.org/publications/policies>. Accessed 2020-06-18.
- [5] ACM. 2018. Criteria for Authorship. Available at <https://www.acm.org/publications/policies/authorship>. Accessed 2020-09-06.
- [6] ACM. 2018. Pre-Publication Evaluation. Available at <https://www.acm.org/publications/policies/pre-publication-evaluation>. Accessed 2020-06-18.
- [7] ACM. 2018. Reviewer Anonymity. Available at <https://www.acm.org/publications/policies/reviewer-anonymity>. Accessed 2020-06-18.
- [8] ACM. 2018. Welcome ACM Associate Editors. Video, available at https://www.youtube.com/watch?v=_svVVjCS0. Accessed 2020-06-16.
- [9] ACM. 2019. ACM Policy on Plagiarism, Misrepresentation, and Falsification. Available at <https://www.acm.org/publications/policies/plagiarism-overview>. Accessed 2020-06-18.
- [10] ACM. 2019. Conflict of Interest Policy. Available at <https://www.acm.org/publications/policies/conflict-of-interest>. Accessed 2020-06-17.
- [11] ACM. 2020. Policy on Roles and Responsibilities in ACM Publishing. Available at <https://www.acm.org/publications/policies/roles-and-responsibilities>. Accessed 2020-06-14.
- [12] ACM SIGCSE. [n.d.]. Program Chair Responsibilities. Available at <https://sigcse.org/sigcse/policies/pcr.html>. Accessed 2020-06-19.
- [13] ACM SIGCSE. 2017. Conflict of Interest Policy. Available at <https://sigcse.org/sigcse/policies/COI.html>. Accessed 2020-06-13.
- [14] ACM Transactions on Computing Education. [n.d.]. Associate Editor Guidelines. Available at <https://dl.acm.org/journal/toce/associate-editors>. Accessed 2020-06-13.
- [15] ACM Transactions on Computing Education. [n.d.]. Editorial Board. Available at <https://dl.acm.org/journal/toce/editorial-board>. Accessed 2020-06-17.
- [16] ACM Transactions on Computing Education. [n.d.]. Guidelines for Authors. Available at <https://dl.acm.org/journal/toce/author-guidelines>. Accessed 2020-06-17.
- [17] ACM Transactions on Computing Education. [n.d.]. Review Form. unpublished.
- [18] ACM Transactions on Software Engineering and Methodology. [n.d.]. Author Guidelines. Available at <https://dl.acm.org/journal/tosem/author-guidelines>. Accessed 2020-06-15.
- [19] ACM Transactions on Software Engineering and Methodology. [n.d.]. Distinguished Reviewers. Available at <https://dl.acm.org/journal/tosem/distinguished-reviewers-board>. Accessed 2020-06-15.
- [20] ACM Transactions on Software Engineering and Methodology. 2019. TOSEM Referees. Available at <https://dl.acm.org/journal/tosem/past-reviewers>. Accessed 2020-06-15.
- [21] Jelena Aleksic, Adrian Alexa, Teresa K. Attwood, Neil Chue Hong, Martin Dahlö, Robert Davey, Holger Dinkel, Konrad U. Förstner, Ivo Grigorov, Jean-Karim Hériché, Leo Lahti, Dan MacLean, Michael L Markie, Jenny Molloy, Maria Victoria Schneider, Camille Scott, Richard Smith-Unna, and Bruno Miguel Vieira. 2015. An Open Science Peer Review Oath. *F1000Research* 3 (Jan. 2015), 271. <https://doi.org/10.12688/f1000research.5686.2>
- [22] Chen Avin, Zvi Lotker, David Peleg, and Itzik Turkel. 2015. Social Network Analysis of Program Committees and Paper Acceptance Fairness. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*. ACM Press. <https://doi.org/10.1145/2808797.2809305>
- [23] Samy A. Azer, Subha Ramani, and Ray Peterson. 2012. Becoming a peer reviewer to medical education journals. *Medical Teacher* 34, 9 (May 2012), 698–704. <https://doi.org/10.3109/0142159x.2012.687488>
- [24] Charles D. Bailey, Dana R. Hermanson, and Timothy J. Louwers. 2008. An examination of the peer review process in accounting journals. *Journal of Accounting Education* 26, 2 (Jan. 2008), 55–72. <https://doi.org/10.1016/j.jaccedu.2008.04.001>
- [25] Mordechai Ben-Ari. 2000. How to Get a Good Review. *ACM SIGCSE Bulletin* 32, 2 (June 2000), 4–6. <https://doi.org/10.1145/355354.571920>
- [26] Dale J. Benos, Edlira Bashari, Jose M. Chaves, Amit Gaggar, Niren Kapoor, Martin LaFrance, Robert Mans, David Mayhew, Sara McGowan, Abigail Polter, Yawar Qadri, Shanta Sarfare, Kevin Schultz, Ryan Splittgerber, Jason Stephenson, Cristy Tower, R. Grace Walton, and Alexander Zotov. 2007. The ups and downs of peer review. *Advances in physiology education* 31, 2 (2007), 145–152. <https://doi.org/10.1152/advan.00104.2006>
- [27] Dale J. Benos, Kevin L. Kirk, and John E. Hall. 2003. How to review a paper. *Advances in physiology education* 27, 2 (2003), 47–52. <https://doi.org/10.1152/advan.00057.2002>
- [28] Ken Birman and Fred B. Schneider. 2009. Program Committee Overload in Systems (Viewpoint). *Commun. ACM* 52, 5 (May 2009), 34–37. <https://doi.org/10.1145/1506409.1506421>
- [29] Georges Bordage. 2001. Reasons reviewers reject and accept manuscripts: the strengths and weaknesses in medical education reports. *Academic Medicine* 76, 9 (2001), 889–896. <https://doi.org/10.1097/00001888-200109000-00010>
- [30] Lutz Bornmann, Irina Nast, and Hans-Dieter Daniel. 2008. Do editors and referees look for signs of scientific misconduct when reviewing manuscripts? A quantitative content analysis of studies that examined review criteria and reasons for accepting and rejecting manuscripts for publication. *Scientometrics* 77, 3 (Sept. 2008), 415–432. <https://doi.org/10.1007/s11192-007-1950-2>
- [31] Giangiacomo Bravo, Francisco Grimaldo, Emilia López-Iñesta, Bahar Mehmani, and Flaminio Squazzoni. 2019. The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications* 10, 1 (2019), 1–8. <https://doi.org/10.1038/s41467-018-08250-2>
- [32] Guillaume Cabanac and Thomas Preuss. 2013. Capitalizing on order effects in the bids of peer-reviewed conferences to secure reviews by expert referees. *Journal of the American Society for Information Science and Technology* 64, 2 (2013), 405–415. <https://doi.org/10.1002/asi.22747>
- [33] Koli Calling. 2019. Program Committee. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research (Koli, Finland)*. iv (in Front matter). Available at <https://dl.acm.org/doi/proceedings/10.1145/3364510> (Accessed 2020-09-04).

- [34] CHI. 2020. Anonymisation Policy. Available at <https://chi2020.acm.org/authors/papers/chi-anonymisation-policy/>. Accessed 2020-06-17.
- [35] CHI. 2020. Papers FAQ. Available at <https://chi2020.acm.org/authors/papers/papers-faq/>. Accessed 2020-06-18.
- [36] CHI. 2020. Papers Review Process. Available at <https://chi2020.acm.org/authors/papers/papers-review-process/>. Accessed 2020-06-18.
- [37] CHI. 2020. Technical Program Chairs' Welcome. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, vi (in Front matter). <https://doi.org/10.1145/3313831>
- [38] Daryl E. Chubin and Edward J. Hackett. 1990. *Peerless science: Peer review and US science policy*. SUNY Press.
- [39] CompEd. 2019. Papers Review Process. Available at <https://www.acmcomped.org/reviewers-2/>. Accessed 2020-09-11.
- [40] CompEd. 2019. Reviewers. In *Proceedings of the ACM Conference on Global Computing Education (Chengdu, Sichuan, China)*, pp. xi–xiv (unnumbered). Available at <https://dl.acm.org/doi/proceedings/10.1145/3300115> (Accessed 2020-09-04).
- [41] Computer Science Education. [n.d.]. Editorial Board. Available at <https://www.tandfonline.com/action/journalInformation?show=editorialBoard&journalCode=nscse20>. Accessed 2020-06-15.
- [42] Computer Science Education. [n.d.]. Review Form. unpublished.
- [43] Computer Science Education. 2020. Instructions for Authors. Available at <https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=nscse20>. Accessed 2020-06-17.
- [44] CONSORT. 2010. CONSORT home page. Available at <http://www.consort-statement.org>. Accessed 2020-06-15.
- [45] COPE Council. 2011. Code of conduct and best practice guidelines for journal editors. <https://publicationethics.org/files/Code%20of%20Conduct.pdf>, Accessed on 6/3/2020.
- [46] COPE Council. 2015. Suspected redundant publication in a submitted manuscript. <https://publicationethics.org/resources/flowcharts-new/what-do-if-you-suspect-redundant-duplicate-publication>, Accessed on 9/3/2020.
- [47] COPE Council. 2017. COPE Discussion document: Who "owns" peer reviews. https://publicationethics.org/files/Who_owns_peer_reviews_discussion_document.pdf, Accessed on 6/3/2020.
- [48] COPE Council. 2017. Ethical guidelines for peer reviewers. https://publicationethics.org/files/Ethical_Guidelines_For_Peer_Reviewers_2.pdf, Accessed on 6/3/2020.
- [49] CORE: Computing Research and Education Association of Australasia. [n.d.]. Welcome to CORE. Available at <http://www.core.edu.au>. Accessed 2020-06-28.
- [50] Robert P. Dellavalle, Marcus A. Banks, and Jeffrey I. Ellis. 2007. Frequently asked questions regarding self-plagiarism: How to avoid recycling fraud. *Journal of the American Academy of Dermatology* 57, 3 (Sept. 2007), 527. <https://doi.org/10.1016/j.jaad.2007.05.018>
- [51] A. R. Dobeles. 2015. Assessing the quality of feedback in the peer-review process. *Higher Education Research & Development* 34, 5 (2015), 853–868. <https://doi.org/10.1080/07294360.2015.1011086>
- [52] Simon Duchesne and Pierre Jannin. 2008. Proposing a manuscript peer-review checklist. *Neuroimage* 39, 4 (2008), 1783–1787. <https://doi.org/10.1016/j.neuroimage.2007.09.065>
- [53] Leo Egghe. 2011. Problems with "natural selection of academic papers". *Scientometrics* 88, 2 (2011), 663–667. <https://doi.org/10.1007/s11192-011-0395-9>
- [54] Jill Freyne, Lorcan Coyle, Barry Smyth, and Padraig Cunningham. 2010. Relative status of journal and conference publications in computer science. *Commun. ACM* 53, 11 (Nov. 2010), 124–132. <https://doi.org/10.1145/1839676.1839701>
- [55] J. A. García, Rosa Rodríguez-Sánchez, and J. Fdez-Valdivia. 2020. Confirmatory bias in peer review. *Scientometrics* 123, 1 (2020), 517–533. <https://doi.org/10.1007/s11192-020-03357-0>
- [56] Satrajit S. Ghosh, Arno Klein, Brian Avants, and K. Jarrod Millman. 2012. Learning from open source software projects to improve scientific review. *Frontiers in Computational Neuroscience* 6 (2012). <https://doi.org/10.3389/fncom.2012.00018>
- [57] Julie R. Gilbert, Elaine S. Williams, and George D. Lundberg. 1994. Is There Gender Bias in JAMA's Peer Review Process? *Journal of the American Medical Association* 272, 2 (July 1994), 139–142. <https://doi.org/10.1001/jama.1994.03520020065018>
- [58] Fiona Godlee, Catharine R. Gale, and Christopher N. Martyn. 1998. Effect on the Quality of Peer Review of Blinding Reviewers and Asking Them to Sign Their Reports. *Journal of the American Medical Association* 280, 3 (July 1998), 237. <https://doi.org/10.1001/jama.280.3.237>
- [59] Mary Golden and David M. Schultz. 2012. Quantifying the Volunteer Effort of Scientific Peer Reviewing. *Bulletin of the American Meteorological Society* 93, 3 (March 2012), 337–345. <https://doi.org/10.1175/bams-d-11-00129.1>
- [60] Hugh Gosden. 2003. 'Why not give us the full story?': functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes* 2, 2 (2003), 87–101. [https://doi.org/10.1016/S1475-1585\(02\)00037-1](https://doi.org/10.1016/S1475-1585(02)00037-1)
- [61] Thomas Hadjistavropoulos and Peter J. Bieling. 2000. When Reviews Attack: Ethics, Free Speech, and the Peer Review Process. *Canadian Psychology/Psychologie Canadienne* 41, 3 (Aug. 2000), 337–345. <https://doi.org/10.1037/h0086865>
- [62] Joseph Y. Halpern and David C. Parkes. 2011. Journals for Certification, Conferences for Rapid Dissemination. *Commun. ACM* 54, 8 (Aug. 2011), 36–38. <https://doi.org/10.1145/1978542.1978555>
- [63] Debra Houry, Steven Green, and Michael Callahan. 2012. Does mentoring new peer reviewers improve review quality? A randomized trial. *BMC Medical Education* 12, 1 (Aug. 2012). <https://doi.org/10.1186/1472-6920-12-83>
- [64] Christopher D. Hundhausen. 2017. From the Editor's Desk: TOCE continues on a positive trajectory in 2016. *ACM Transactions on Computing Education* 17, 2 (June 2017), 1–7. <https://doi.org/10.1145/3078193>
- [65] ICER. 2014. Review Process. Available at <https://archive.icer.acm.org/icer-2014/paper-submissions/>. Accessed 2020-09-10.
- [66] ICER. 2016. Review Process. Available at <https://archive.icer.acm.org/icer-2016/review-process/>. Accessed 2020-09-11.
- [67] ICER. 2017. Review Process. Available at <https://archive.icer.acm.org/icer-2017/review-process/>. Accessed 2020-09-11.
- [68] ICER. 2018. Review Form. unpublished. With permission of Conference Chair.
- [69] ICER. 2018. Review Process. Available at <https://icer.acm.org/icer-2018/review-process/>. Accessed 2020-06-17.
- [70] ICER. 2020. How to review for ICER. Available at <https://docs.google.com/presentation/d/1rs9DREfWH6ok4UKxBbsJQ3A2OhRRQIKGZTAOXxcSps>. Accessed 2020-06-18.
- [71] ICER. 2020. Papers Review Committee. Available at <https://icer.acm.org/committee/icer-2020-papers-review-committee>. Accessed 2020-09-11.
- [72] ICER. 2020. Research Papers. Available at <https://icer.acm.org/track/icer-2020-papers>. Accessed 2020-06-28.
- [73] ICER. 2020. Review Committee. In *Proceedings of the 2020 ACM Conference on International Computing Education Research (Virtual Event, New Zealand)*, x–xi. Available at <https://dl.acm.org/doi/proceedings/10.1145/3372782> (Accessed 2020-09-04).
- [74] ICSE SEET. 2019. Program Committee. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering Education and Training (Montréal, Québec, Canada)*, xii–xiii. Available at <https://dl.acm.org/doi/proceedings/10.5555/3339216> (Accessed 2020-09-04).
- [75] IEEE. [n.d.]. Become an IEEE Reviewer. Available at <https://journals.ieeeauthorcenter.ieee.org/submit-your-article-for-peer-review/become-an-ieee-reviewer/>. Accessed 2020-06-16.
- [76] IEEE. [n.d.]. IEEE Author Center. Available at <https://journals.ieeeauthorcenter.ieee.org/become-an-ieee-journal-author/publishing-ethics/ethical-requirements/>. Accessed 2020-06-29.
- [77] IEEE. [n.d.]. IEEE Author Peer Review Responsibilities. Available at <https://conferences.ieeeauthorcenter.ieee.org/wp-content/uploads/Peer-Review-Policies-for-the-Author.pdf>. Accessed 2020-06-18.
- [78] IEEE. 2020. IEEE Publication Services and Products Board Operations Manual 2020. Available at <https://pspb.ieee.org/images/files/files/opsmanual.pdf>. Accessed 2020-06-16.
- [79] IEEE Computer Society. [n.d.]. Journal Peer Review. Available at <https://www.computer.org/publications/author-resources/peer-review/journals#double-blind>. Accessed 2020-06-17.
- [80] IEEE Computer Society. [n.d.]. Reviewer Information for Journals. Available at <https://www.computer.org/publications/author-resources/peer-review/journals/reviewer>. Accessed 2020-06-16.
- [81] IEEE Transactions on Education. [n.d.]. Home Page. Available at <https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=13>. Accessed 2020-05-21.
- [82] IEEE Transactions on Software Engineering. [n.d.]. TSE Author Information. Available at <https://www.computer.org/tse-author-information>. Accessed 2020-06-15.
- [83] IEEE Transactions on Software Engineering. [n.d.]. TSE Editorial Board. Available at <https://www.computer.org/csdl/journal/ts/about/107385>. Accessed 2020-06-15.
- [84] IEEE Transactions on Software Engineering. 2020. 2019 Reviewers List. *IEEE Transactions on Software Engineering* 46, 01 (Jan 2020), 114–117. <https://doi.org/10.1109/TSE.2019.2962566>
- [85] ISSEP. 2019. Program Committee and Additional Reviewers. In *Informatics in Schools. New Ideas in School Informatics*. Springer International Publishing, vii–viii. <https://doi.org/10.1007/978-3-030-33759-9>
- [86] ITiCSE. 2011. Reviewers. Available at <http://www.iticse2011.tu-darmstadt.de/authors/reviewers.html>. Accessed 2020-06-17.
- [87] ITiCSE. 2016. Reviewers. Available at <https://ucsp.edu.pe/iticse2016/reviewers/>. Accessed 2020-06-17.
- [88] ITiCSE. 2017. Reviewers: new and improved review process for ITiCSE 2017. Available at <http://iticse.acm.org/ITiCSE2017/reviewers/>. Accessed 2020-06-17.
- [89] ITiCSE. 2020. Conference Committee. Available at <https://iticse.acm.org/conference-committee/>. Accessed 2020-06-17.
- [90] ITiCSE. 2020. Important Information for Authors. Available at <http://iticse.acm.org/important-information-for-authors/>. Accessed 2020-06-17.

- [91] ITiCSE. 2020. Paper review process. Available at <https://iticse.acm.org/paper-review-process/>. Accessed 2020-06-17.
- [92] ITiCSE. 2020. Reviewers of Full Papers. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (Trondheim, Norway)*. xvii – xx. Available at <https://dl.acm.org/doi/proceedings/10.1145/3341525> (Accessed 2020-09-04).
- [93] Hamid R. Jamali, David Nicholas, Anthony Watkinson, Abdullah Abrizah, Blanca Rodríguez-Bravo, Cherifa Boukacem-Zeghmouri, Jie Xu, Tatiana Polezhaeva, Eti Herman, and Marzena Swigon. 2020. Early career researchers and their authorship and peer review beliefs and practices: An international study. *Learned Publishing* 33, 2 (2020), 142–152. <https://doi.org/10.1002/leap.1283>
- [94] Tom Jefferson. 1998. Redundant publication in biomedical sciences: Scientific misconduct or necessity? *Science and Engineering Ethics* 4, 2 (1998), 135–140. <https://doi.org/10.1007/s11948-998-0043-9>
- [95] Tom Jefferson. 2002. Measuring the Quality of Editorial Peer Review. *Journal of the American Medical Association* 287, 21 (June 2002), 2786. <https://doi.org/10.1001/jama.287.21.2786>
- [96] Journal of Engineering Education. [n.d.]. Editorial Board. Available at <https://onlinelibrary.wiley.com/page/journal/21689830/homepage/editorialboard.html>. Accessed 2020-06-16.
- [97] Journal of the Learning Sciences. [n.d.]. Editorial Board. Available at <https://www.tandfonline.com/action/journalInformation?show=editorialBoard&journalCode=hlms20>. Accessed 2020-06-16.
- [98] Journal of the Learning Sciences. 2020. Instructions for Authors. Available at <https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=hlms20>. Accessed 2020-06-17.
- [99] Sushil Ghanshyam Kachewar and Smita Balwant Sankaye. 2013. Reviewer index: A new proposal of rewarding the reviewer. *Mens Sana Monographs* 11, 1 (2013), 274. <https://doi.org/10.4103/0973-1229.109347>
- [100] J. Kelly, T. Sadeghih, and K. Adeli. 2014. Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. *EJIFCC* 25, 3 (2014), 227–243. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975196/pdf/ejifcc-25-227.pdf>
- [101] Koli Calling. 2020. Program Committee. Available at <https://www.kolicalling.fi/pc.php>. Accessed 2020-09-20.
- [102] Amruth Kumar, Patricia A. Joseph, Michael Goldweber, and Paul J. Wagner. 2008. Reviewing the SIGCSE reviewing process. *ACM SIGCSE Bulletin* 40, 2 (June 2008), 84–89. <https://doi.org/10.1145/1383602.1383641>
- [103] LaTiCE. 2020. LaTiCE 2020 Organization. Available at <http://www.laticence-conference.org/organization.html>. Accessed 2020-09-11.
- [104] Wen-Yau Cathy Lin. 2020. Self-plagiarism in academic journal articles: from the perspectives of international editors-in-chief in editorial and COPE case. *Scientometrics* 123, 1 (Feb. 2020), 299–319. <https://doi.org/10.1007/s11192-020-03373-0>
- [105] Margaret E. Lloyd. 1990. Gender factors in reviewer recommendations for manuscript publication. *Journal of Applied Behavior Analysis* 23, 4 (Winter 1990), 539–543. <https://doi.org/10.1901/jaba.1990.23-539>
- [106] Monica M. McGill, Adrienne Decker, and Zachary Abbott. 2018. Improving research and experience reports of pre-college computing activities: A gap analysis. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 964–969. <https://doi.org/10.1145/3159450.3159481>
- [107] Bertrand Meyer, Christine Choppy, Jørgen Staunstrup, and Jan van Leeuwen. 2009. Research Evaluation for Computer Science (Viewpoint). *Commun. ACM* 52, 4 (April 2009), 31–34. <https://doi.org/10.1145/1498765.1498780>
- [108] Ian Norman and Peter Griffith. 2008. Duplicate publication and ‘salami slicing’: Ethical issues and practical solutions. *International Journal of Nursing Studies* 45, 9 (2008), 1257–1260. <https://doi.org/10.1016/j.ijnurstu.2008.07.003>
- [109] Ian Parberry. 1994. A Guide for New Referees in Theoretical Computer Science. *Information and Computation* 112, 1 (1994), 96 – 116. <https://doi.org/10.1006/inco.1994.1053>
- [110] David B. Resnik, Christina Gutierrez-Ford, and Shyamal Peddada. 2008. Perceptions of Ethical Problems with Scientific Journal Peer Review: An Exploratory Study. *Science and Engineering Ethics* 14, 3 (2008), 305–310. <https://doi.org/10.1007/s11948-008-9059-4>
- [111] Sara Rockwell. 2006. Ethics of peer review: a guide for manuscript reviewers. Available through U.S. Health and Human Services Office of Research Integrity website at <https://ori.hhs.gov/sites/default/files/prthetics.pdf>. Accessed 2020-06-13.
- [112] Marko A. Rodriguez, Johan Bollen, and Herbert Van de Sompel. 2007. Mapping the bid behavior of conference referees. *Journal of Informetrics* 1, 1 (2007), 68–82. <https://doi.org/10.1016/j.joi.2006.09.006>
- [113] Henning Schulzrinne. 2009. Double-blind reviewing: more placebo than miracle cure? *ACM SIGCOMM Computer Communication Review* 39, 2 (2009), 56–59. <https://doi.org/10.1145/1517480.1517492>
- [114] Henning Schulzrinne. 2009. Double submissions. *ACM SIGCOMM Computer Communication Review* 39, 3 (June 2009), 40–42. <https://doi.org/10.1145/1568613.1568622>
- [115] Nihar B. Shah, Behzad Tabibian, Krikamol Muandet, Isabelle Guyon, and Ulrike Von Luxburg. 2018. Design and analysis of the NIPS 2016 review process. *The Journal of Machine Learning Research* 19, 49 (2018), 1–34. Available at <https://jmlr.org/papers/volume19/17-511/17-511.pdf> (Accessed 2020-09-11).
- [116] SIGCSE. 2020. Paper Reviewers. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (Portland, OR, USA)*. xlix – lvii. Available at <https://dl.acm.org/doi/proceedings/10.1145/3328778> (Accessed 2020-09-04).
- [117] SIGCSE Symposium. 2014. The Meta-Review Process for Paper Submissions. Available at <http://sigcse2014.sigcse.org/reviewers/metareviews.php.html>. Accessed 2020-06-17.
- [118] SIGCSE Symposium. 2015. Guide for Reviewers of Papers. Available at <http://sigcse2015.sigcse.org/reviewers/guidepapers.html>. Accessed 2020-06-17.
- [119] SIGCSE Symposium. 2016. Meta Review Process. Available at <http://sigcse2016.sigcse.org/reviewers/metareviews.html>. Accessed 2020-06-17.
- [120] SIGCSE Symposium. 2017. APC Responsibilities. Available at <https://sigcse2017.sigcse.org/reviewers/apc-responsibilities.html>. Accessed 2020-06-17.
- [121] SIGCSE Symposium. 2019. Guide for Reviewers of Papers. Available at <http://sigcse2019.sigcse.org/reviewers/paper-review-guidelines.html>. Accessed 2020-06-17.
- [122] SIGCSE Symposium. 2020. Conference Committee. Available at <https://sigcse2020.sigcse.org/info/committees.html>. Accessed 2020-06-17.
- [123] SIGCSE Symposium. 2020. Paper Review Guidelines. Available at <https://sigcse2020.sigcse.org/reviewers/paper-review-guidelines.html>. Accessed 2020-06-17.
- [124] SIGCSE Symposium. 2020. Paper Review Process. Available at <https://sigcse2020.sigcse.org/reviewers/paper-review-process.html>. Accessed 2020-06-17.
- [125] A. J. Smith. 1990. The task of the referee. *Computer* 23, 4 (1990), 65–71. <https://doi.org/10.1109/2.55470>
- [126] Lawrence Souder. 2011. The ethics of scholarly peer review: a review of the literature. *Learned Publishing* 24 (2011), 55–74. <https://doi.org/10.1087/20110109>
- [127] Chris Street and Kerry W. Ward. 2019. Cognitive Bias in the Peer Review Process: Understanding a Source of Friction between Reviewers and Researchers. *SIGMIS Database* 50, 4 (Nov. 2019), 52–70. <https://doi.org/10.1145/3371041.3371046>
- [128] Taylor & Francis. [n.d.]. An editor’s guide to the peer review process. Available at <https://editorresources.taylorandfrancis.com/managing-peer-review-process/>. Accessed 2020-06-15.
- [129] Taylor & Francis. [n.d.]. A guide to becoming a peer reviewer. Available at <https://editorresources.taylorandfrancis.com/reviewer-guidelines/>. Accessed 2020-06-30.
- [130] Leanne Tite and Sara Schroter. 2007. Why do peer reviewers decline to review? A survey. *Journal of Epidemiology & Community Health* 61, 1 (2007), 9–12. <https://doi.org/10.1136/jech.2006.049817>
- [131] Andrew Tomkins, Min Zhang, and William D. Heavlin. 2017. Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences* 114, 48 (2017), 12708–12713. <https://doi.org/10.1073/pnas.1707323114> arXiv:<https://www.pnas.org/content/114/48/12708.full.pdf>
- [132] Moshe Y. Vardi. 2009. Conferences vs. journals in computing research. *Commun. ACM* 52, 5 (May 2009), 5–5. <https://doi.org/10.1145/1506409.1506410>
- [133] Elizabeth Walsh, Maevé Rooney, Louis Appleby, and Greg Wilkinson. 2000. Open peer review: A randomised controlled trial. *British Journal of Psychiatry* 176, 1 (Jan. 2000), 47–51. <https://doi.org/10.1192/bjp.176.1.47>
- [134] Mark Ware. 2008. Peer review in scholarly journals: Perspective of the scholarly community—Results from an international study. *Information Services & Use* 28, 2 (2008), 109–112. <https://doi.org/10.3233/ISU-2008-0568>
- [135] Ellen J. Weber, Patricia P. Katz, Joseph F. Waeckerle, and Michael L. Callahan. 2002. Author perception of peer review: impact of review quality and acceptance on satisfaction. *Journal of the American Medical Association* 287, 21 (2002), 2790–2793. <https://doi.org/10.1001/jama.287.21.2790>
- [136] Dietmar Wolfram, Peiling Wang, Adam Hembree, and Hyoungjoo Park. 2020. Open peer review: promoting transparency in open science. *Scientometrics* (May 2020). <https://doi.org/10.1007/s11192-020-03488-4>
- [137] Jian Zhang and Mark Sherriff. 2020. SIGCSE 2020 Recap. *SIGCSE Bull.* 52, 2 (April 2020), 4–7. <https://doi.org/10.1145/3397568.3397570>

A INTERVIEW SCRIPT

The overall structure of the interviews was sent to informants in advance:

A.1 ITiCSE 2020 — WG9 CONVERSATION IN A NUTSHELL

Introduction:

- What is the unique selling point of your venue compared to others?
- How does this affect reviewing?

Our 4 principal themes:

- *Criteria*: The factors that determine whether a submission is accepted.
- *Ethics/etiquette*: Conventions/decisions keeping the process fair and constructive.
- *Process*: The organizational/logistical apparatus from CFP to acceptance decision.
- *Roles*: What responsibilities are associated with editors, associate editors, reviewers, (or chairs, meta-reviewers, reviewers) and how are they assigned/ audited?

Three opening questions on each of these themes:

- How/why did the current state of affairs come to be?
- How—and how often—does the venue evaluate/re-assess/update what it does?
- How does this compare to other venues?

— *Augmented with further, follow-up questions.* —

Closing:

- Are there currently things that you'd like to change?
- Are there issues of ongoing concern/debate?
- Anything we've missed?
- Is there someone else you think we should interview?

A.2 Follow-up questions

Depending on the content and flow of the interview, a variety of follow-up questions was asked on each of the themes. For example:

- Are review criteria of equal weight? If not, which criteria are prioritized?
- How is [a particular criterion] interpreted?
- What is the conflict of interest policy for your venue?
- What if a reviewer recognizes the author of an anonymous submission?
- Do authors have a responsibility to review? (for your venue and in general)
- How are editors or reviewers compensated or rewarded for their work?
- Has the review process evolved over time? If so, how and why?
- Does your process reinforce the review criteria (e.g., through the structure of the review form)?
- What influence does the publisher have on the process / decision-making?
- How has the rapid growth in the number of conferences and the number of submissions per conference affected the roles involved and the nature of each role?
- What information do people in different roles have access to (e.g., names of reviewers)?
- How (if at all) Do people in different roles participate in decisions outside of reviewing, such as planning future conferences or special issues, or instructions to reviewers or authors?
- How much time do you spend on this work?
- How much time does a typical chair/AE/reviewer spend?

B SURVEY QUESTIONS

- (1) How many papers have you reviewed for C.S. Education conferences/journals THROUGHOUT YOUR CAREER?
 - Options: 0; 1; 2; 3; 4-5; 6-9; 10+.
- (2) How many C.S. Education conferences/journals have you reviewed for THROUGHOUT YOUR CAREER?
 - Options: 0; 1; 2; 3; 4-5; 6-9; 10+.
- (3) What proportion of your reviews IN THE PAST 5 YEARS have been for non-C.-S. Education venues?
 - Options: None; 1-25%; 26-50%; 51-75%; >75%.
- (4) How many C.-S. Education venues (*i.e.*, conferences/journals) have you reviewed for in the PAST YEAR?
- (5) On average, how many MINUTES does it take you to review a 6–8 pages conference paper?
- (6) For a conference like ITiCSE, how many papers is it reasonable to ask a reviewer to review?
- (7) For a conference like ITiCSE, what should the minimum number of reviews be for a full paper?
- (8) If you reviewed for ITiCSE 2020, did you know that ITiCSE 2020 published guidelines for reviewers and the review process? (Those guidelines can be found at: <https://iticse.acm.org/paper-review-process/>)
 - Not relevant, I did not review for ITiCSE 2020
 - No, I did not know about them
 - No, but I read a previous iteration
 - Yes, I knew about them but didn't read them
 - Yes, I skimmed them briefly
 - Yes, I referred to them while reviewing
- (9) Having submitted a review, have you ever changed your score? (Check all that apply.)
 - No
 - Yes, because I made a slip
 - Yes, because other reviews influenced my thinking
 - Yes, as a result of the meta-review process
- (10) In your experience, how is meta-review used? (Check all that apply.)
 - Facilitate discussion among reviewers
 - Summarize + filter reviews
 - Exclude poor-quality reviews
 - Help reviewers homogenize scores
 - Other: _____
- (11) What statements best characterize your perception of "double-blind" reviewing? (Check all that apply.)
 - It reduces bias in reviewing
 - It reduces confirmation bias in reviewing that unfairly advantages senior authors
 - It reduces cognitive biases in reviewing that negatively affect new authors
 - I don't think it is useful
 - Authors are still easy to guess
 - Redacting identity is too big a hassle
 - Overall, the trade-off between workload and benefits makes double-blind valuable
 - Other: _____
- (12) When you wrote your first review, did you ask for—or receive—any guidance about how to write a good review? (Check all that apply.)
 - I did not have any instructions except the review form
 - I read the review form, and other instructions for reviewers
 - I asked for additional guidance from a more experienced reviewer but did not learn much
 - I asked for and received useful guidance from a more experienced reviewer
 - As a new reviewer I was mentored by someone from the conference/journal
 - Other: _____

C REVIEW PROCESSES

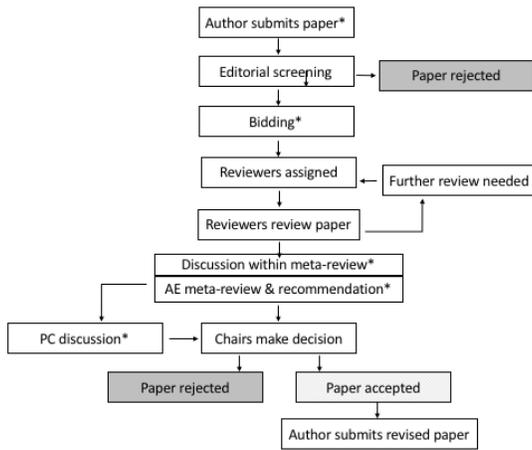


Figure 3: A fairly typical conference review process. Asterisk (*) indicates variations (i.e., not all conferences use bidding; not all meta-review includes discussion; not all conferences hold Program Committee discussions; not all conferences allow revision of papers prior to submission of camera-ready material) as discussed in Section 4.3.

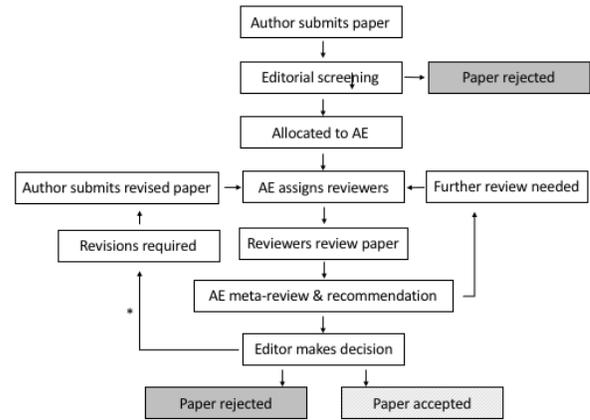


Figure 4: A fairly typical journal peer-review process. Note that not all journals use Associate Editors (AEs). Note also that the 'revision' loop (*) is often limited.