

The Evolution of Computing Education Research: A Meta-Analytic Perspective



Lauri Malmi, Jane Sinclair, Judy Sheard, Simon, and Päivi Kinnunen

1 Introduction

In this book, the main perspective for analysing the development of computing education research (CER) is scientometric analysis, which focuses on authors and their collaborations, publication venues, and research topics in terms of keyword analysis. There are sophisticated algorithms and tools that perform this analysis automatically on the basis of publication metadata. However, there are several other perspectives for analysis that can provide deeper and more refined information. Scientometric methods fall short of answering questions such as what theoretical frameworks have been used in research, what research methodologies, data collection tools, or analysis methods have been used, and how the publications match various criteria for research quality or nature. Keyword analysis can give some information in these areas, but a more detailed analysis requires manual inspection, at least of the abstracts and sometimes of the full papers.

L. Malmi (✉)
Aalto University, Espoo, Finland
e-mail: Lauri.Malmi@aalto.fi

J. Sinclair
University of Warwick, Coventry, UK
e-mail: J.E.Sinclair@warwick.ac.uk

J. Sheard
Monash University, Melbourne, VIC, Australia
e-mail: judy.sheard@monash.edu

Simon
Wadalba, NSW, Australia

P. Kinnunen
University of Helsinki, Helsinki, Finland
e-mail: paivi.kinnunen@helsinki.fi

Over the past 20 years, many researchers have carried out significant work in various meta-studies, mapping studies, and systematic literature reviews in CER, addressing, for example, use of theories [37, 38, 40, 74], research methods [32, 41, 42, 63], learning tools and environments [2, 46, 54, 65], or specific topics, such as introductory programming [36]. In addition, many studies have investigated computing education publications in specific venues, such as the Australasian Computing Education conference [66], Koli Calling [59, 68], and the SIGCSE Technical Symposium [6].

In this chapter, we focus on what these reviews can reveal about the history and development of the CER field. We acknowledge that our perspective is limited to the more recent developments, as there are very few studies that have explored the early years of computing education literature. Valentine [83] explored 20 years of CS1 papers in the SIGCSE Technical Symposium (1984–2003), broadly categorising the type of work that was reported. Becker and Quille [6] explored almost 50 years of research on CS1 in SIGCSE (1970–2018), focusing on the topics of the research. However, we are not aware of any reviews that cover papers before 2000 and discuss the nature and process of the research, or that reflect on the richness and maturity of the field, and the quality and variety of evidence presented.

The absence of such analyses is not surprising considering how the field began. Computer science itself is a new research field, with roots in electrical engineering, mathematics, and science [76], which gained its own identity between 1940s and 1960s. It had to struggle to establish its own position in the academic world, with its own departments and methods [13, 76]. Since then it has grown and greatly expanded its scope, and today its applications are used in practically all fields of science and in society more broadly.

Computing education has naturally followed this development [75]. In 1950s, training of programmers was conducted mainly in computer companies, while mathematics departments within academia focused more on formal aspects of programming and logic. The first ACM Computing Curricula, published in 1968 and 1977 [3, 4], helped universities to set up their computer science degree programs in a more uniform way. In this phase, it became natural to organise conferences where pedagogical practices were discussed and novel educational innovations were presented. The SIGCSE Technical Symposium, launched in 1970, was the first leading venue to focus solely on computing education, followed later by other venues where computer science teachers could meet one another, present their work, and exchange ideas and experiences. This exchange was the prevalent idea in the beginning, as the whole concept of ‘computing education research’ or ‘computer science education research’ was still vague, and teaching practitioners placed a high value on acquiring new ideas and sharing experiences. In the meantime, educational scientists were conducting research, addressing challenges emerging in more established disciplines and investigating teaching, learning, and studying in generic terms.

While most computing educators in 1970s and 1980s focused on developing pedagogical approaches and adequate learning resources, significant research was already being carried out in CER, especially investigating the work of professional

programmers and the learning of programming. Weinberg's book, *Psychology of Programming*, was published in 1971 [85]. Papert's book addressing children's learning of programming, turtle graphics, and LOGO was published in 1980 [51]. Soloway and Ehrlich carried out important research on experts' programming plans in 1980s [72] as well as comparing the ways that novice and expert programmers work. Important early venues for presenting research were the Empirical Studies of Programmers (ESP) conferences and Psychology of Programming Interest Group (PPIG) workshops, which both started in 1986. The ESP conferences ceased in 1990s, but PPIG is still active. Guzdial and du Boulay [22] present a more comprehensive history of the early years of research in computing education.

In this chapter, we focus on later developments, mainly during the past 20 years, when there emerged a growing awareness and interest in the quality of evidence of the impact of educational innovations, as well as building deeper theoretical understanding of the factors involved in teaching and learning computing. We discuss these developments in the light of reviews and meta-analyses of CER that have been carried out during this period. We begin by looking at the development of CER using Fensham's framework [17], which was originally developed to analyse the growth of a neighbour discipline, science education research. In Sect. 3, we discuss a number of categorisation schemes that have been developed as tools to analyse various aspects of computing education literature. In Sects. 4 and 5 we look at collected evidence of the use of theoretical frameworks and methodological approaches. We continue in Sect. 6 by looking at findings of analyses which have focused on specific publication venues. Finally, we discuss what these findings holistically report on the development of CER, concluding with some recommendations.

While each of the review papers in our data pool covers a limited span of years, perhaps focusing on certain aspects of papers in a small number of publication venues, together these reviews cover a very significant share of papers published in central CER publication venues during the past 20 years, as CER has emerged as an independent field. They thus provide an interesting additional perspective to complement the scientometric analyses presented elsewhere in this book.

2 Emergence of CER as an Independent Field

Discussion of the need for more rigorous research in computing education gained wider attention in the early 2000s. Fincher and Petre published their seminal book, *Computer Science Education Research* [18], in 2004. The book discussed the nature and scope of work in the field and presented a comprehensive tutorial for conducting empirical research in computing education, with a number of case studies illuminating different types of research in the area. The Koli Calling conference (Koli), which had been launched in 2001, took steps towards becoming an international research conference in 2004. The International Computing Education Research workshop (ICER) was launched in 2005, clearly laying out in its call for papers what would be expected from submissions as research papers: a clear theoretical

basis building on literature, a strong empirical basis, drawing on relevant research methods, and explicitly explaining how the paper contributes to existing knowledge in computing education. At the end of the decade, the Journal of Educational Resources in Computing (JERIC), which had published many experience reports, curricular and course descriptions, and learning resources in various subareas of computing, transformed into ACM Transactions on Computing Education (TOCE), which sought to publish high quality research, and has joined the longer-standing Computer Science Education (CSE) as a premier journal in the field.

It is likely that several factors contributed to this increasing interest in research quality. While experiences of various pedagogical approaches and observations of their impact are highly valuable for teaching practitioners seeking to develop their teaching, their value as evidence of the impact is unclear. The quality of data collection and analysis matters, as does the chain of inference behind the findings. If computing education researchers wish colleagues in other disciplines to acknowledge that they are undertaking serious research, publication quality matters. Computing education research is by its very nature an interdisciplinary field which draws on several disciplines, especially education, psychology, and the computing sciences. Thus, comparison with work in these areas is also to be expected.

Moreover, PhD theses, which are evaluated based on academic standards, have to pass the evaluation process in faculties, schools, or departments where academics from other disciplines have an opportunity to assess their quality. High-quality publications are also needed for appointment and promotion in academic positions. Finally, many academic institutions explicitly specify how teaching quality is evaluated and how relevant it is when making decisions on academic promotions, and publication of research papers in education-related conferences and journals has thus become stronger evidence for candidates.

There were consequently a growing number of people who enjoyed carrying out computing education research and for whom the question of quality was very obviously relevant, and these people had to start considering structures and resources that would support quality improvement not only in their own work but also more broadly in the computing education community. As most computing educators teach their classes and carry out their main research in specific topic areas of computing itself, specific venues were needed where the focus of work would be the pedagogical methods and tools in teaching and learning computing across the board, as well as studies on various aspects of students, recruitment, retention, etc.

This development fits well with Fensham's analysis on the development of science education as an independent field of research [17]. Fensham identified two sets of criteria for an independent research field. *Structural criteria* include conferences and journals dedicated to publishing work in the area, established professorships in the area, professional associations, and centres for research and organised research training. All of these elements have existed in CER for a considerable time, but they are not relevant for this chapter, so we shall not discuss them further.

Intra-research criteria focus on the content of the work. Fensham defined seven criteria for evaluating how a field is conducting research. The first criterion,

scientific knowledge, refers to basic knowledge and skills in the field, without which studies cannot be carried out. CER certainly has a massive literature covering several decades of work, which cannot be ignored when building new knowledge. A closely related criterion, *asking questions*, means that the discipline asks questions that other disciplines do not. For example, CER has directed much focus on how students learn programming, and how this learning can be supported by the application of various pedagogical approaches and tools.

The third criterion, *conceptual and theoretical development*, is of particular interest to us. This development is embodied, for example, in the extensive empirical work that researchers have carried out to investigate the impact of various pedagogical innovations on students' learning outcomes and study practices. These works often build on theoretical constructs from educational sciences and psychology, such as self-efficacy [5], fixed/growth mindset [16], and goal orientation [25], seeking to identify relevant factors and their roles in students' learning outcomes, motivation, and study practices. Numerous statistical models have been built to describe these complex relationships, as reviewed and described in the work of Malmi et al. [37, 38]. Moreover, computing education researchers have invested a great deal of work in qualitative analyses to investigate students' conceptions of different computing concepts, resulting in numerous phenomenographical outcome spaces [9, 77, 78] and grounded theories [28, 88]. Computing education researchers have also developed various taxonomies and classification schemas to analyse the relationships and structures among computing concepts, tools, or pedagogical methods [26, 31, 44, 52].

A closely related criterion, *research methodologies*, describes how the field has adopted methods used in other fields and adapted them to its specific needs, as well as developing its own methods within the field itself. For example, CER has adapted several general psychological instruments to measure such matters as students' attitudes and self-efficacy in computing or programming [14, 30, 80]. Moreover, the field has developed several concept inventories to analyse students' conceptions and misconceptions, for example in programming and in data structures and algorithms [12, 42].

The fifth criterion, *progression*, describes how the field builds on previous work to further accumulate scientific knowledge. While obviously all scientific papers (should) incorporate literature reviews, genuine progression is more difficult to measure. Many research publications cite related work without actually building on it in any way [37]. On the other hand, there are streams of research that have systematically explored a single area over a long period, such as Guzdial's studies on media computation [21], the long-term research around the Jeliot program visualisation tool [7], and research in developing BlueJ and analysing data collected with it [10, 29].

As part of this progression, some publications become *seminal* or *model publications*, opening new avenues for research, and therefore become widely cited (criteria 6 and 7). As examples, Hundhausen in his meta-study on algorithm visualisation [24] identified students' interaction with the visualisation as an important factor for learning; an ITiCSE working group report from 2002 [48]

built on this observation and proposed the engagement taxonomy as a hypothesis of the impact of various engagement methods, and this taxonomy led to substantial research on evaluating the impact of visualisations [82]. Another example is the idea of programming puzzles, where students construct a program by ordering a given set of program statements into a working program [53]. These—often called Parsons puzzles—have been further developed and extended in many ways and have led to substantial empirical work [15].

3 Classifying Papers

Research publications can be classified in many ways beyond that involving scientometric publication metadata. We can categorise papers based on the research topics addressed, such as topical areas in a curriculum, on the pedagogical techniques covered in the paper, on the perspectives or properties of stakeholders (students, teachers, the organisation. . .) [27, 36, 66], or on the educational level of the target group (primary, secondary, or tertiary) [60]. We can classify papers based on their type, such as empirical research papers, experience reports, learning resources, tool descriptions, or opinion pieces [83], or based on whether their main purpose is to describe, evaluate, or formulate something such as concepts, activities, resources, novel methods etc. [84]. We can also look at various aspects of the documented research process, such as what theoretical frameworks have been used, if any, or what research methodologies, data collection, and analysis methods have been used [41, 60]. In this section, we discuss some categorisation schemes that are most relevant to our goal of addressing research processes in CER publications, as well as the nature of their contributions.

Simon [66, 71] presented a systematic categorisation scheme comprising four dimensions. The *Context* dimension describes the curricular context in which the research is carried out, such as programming, data structures and algorithms, security, operating systems, etc. If there is no such context, the paper is classified as broad-based. *Theme* (originally Topic) categorises what the research is about: for example, teaching and learning techniques, teaching and learning tools, assessment techniques, ability and aptitude, etc. *Scope* describes the breadth of community involvement in the research, which might be a single subject (course), a program of study, a whole institution, or more than one institution. Finally, *Nature* seeks to differentiate types of paper. A *study* paper reports a research question, the data collection and analysis used to address it, and the findings. An *experiment* paper does the same, but in a manner that would typically be recognised as at least a quasi-experiment. An *analysis* paper investigates existing data such as course or program results or literature. A *report* describes something, such as a new pedagogical approach, learning tool, or learning resource, which has been implemented and possibly used in practice, perhaps with some initial experiences of using it. Finally, a *position/proposal* paper presents the authors' beliefs on a particular matter or a proposal for work yet to be done. Simon considers the first three categories to be

research papers, contrasting them with the other types of papers. Simon's system has been used to analyse papers in multiple publication venues [66–68, 71].

Malmi et al. [41] developed a classification scheme focusing on theoretical constructs, research goals, and various aspects of research process. Their scheme has several dimensions, building on and extending the work by Vessey et al. [84], who had classified computing research literature in several dimensions. From their work, Malmi et al. adopted the dimensions of reference discipline, research approach, and research method and extended the classification scheme into seven dimensions: (1) the theories, models, theoretical frameworks or instruments that had been used; (2) the disciplines from which these had been adopted; (3) the technologies or tools that were used and reported in the work; (4) the general purpose of the work: descriptive, evaluative, or formulative, each with several subcategories; (5) the overall research framework that had been used in the empirical work; (6) the data sources that had been used; and (7) the data analysis methods that had been used. The Malmi et al. classification scheme has been used in a couple of extensive analyses of the computing education literature [40, 41].

Randolph [60] analysed a wide selection of CER literature using a broad scheme, which addressed authors, reporting elements, research topic or content, and a comprehensive analysis of research designs, collected data, and generated evidence.

In the following sections, we analyse what previous studies have revealed about the overall state and trends of CER publications in different venues, beginning with reviews of the theoretical background to work and of the methods used in research, and proceeding to survey some work analysing publications from specific venues. We appreciate that there are many reviews that focus on specific topic areas in computing, such as recursion [43] or event-driven programming [35], on specific educational tools, such as program visualisation [73] or algorithm visualisation [64], or on pedagogical approaches, such as pair programming [11, 81]. However, we do not discuss these reviews, as their perspective is too narrow for a consideration of the overall development of CER as a research field.

On the other hand, reviews that focus on analysing the research presented in specific venues are relevant for this chapter, as they reflect on the development of the CER community's preferences and approaches in research and illuminate the holistic development of the field. Moreover, they complement the scientometric analysis results presented in other chapters of this book.

4 Theoretical Development in CER

In educational research there is a symbiosis between theory and practice. Purely academic theory without connection to practice is unhelpful to practitioners, whereas pure description of the practice without a framing theory—or an aim to develop a more abstract model or theory of the practice—limits the researchers' and practitioners' potential to see beyond immediate empirical observations. This symbiosis between theory and practice is crystallised by Kant: “theory without

practice is empty; practice without theory is blind” (cited by Morrison and Werf [47]). In addition to their potential to help us improve our understanding of the practice, theories have a more immediate practical affordance for researchers. Theories may guide the research process in many ways. Theories can point to relevant phenomena to study and suitable methods with which to study them, and can provide a distinctive viewpoint from which the data can be interpreted [61, 62]. Finally, conceptual and theoretical development of a research field is one of the criteria used by Fensham [17] to describe and evaluate development of science education. The use of existing theories and the development of field-specific theories are signs of a mature research field. In the past decade, there has been an increasing interest in understanding how and to what degree CER is maturing as a research field. This has resulted in several publications that aim to describe and map out what, how, and based on which theories computing education has developed as a research field.

These recent CER papers have approached the definition of theory from three viewpoints. Szabo et al. [74, p. 92] define theory by stating what a theory is: “we use an inclusive definition of theory as a generalisation, abstraction, explanation or prediction of a phenomenon”. Malmi et al. [40, p. 29] approach the definition of theory from another point of view by emphasising what theories provide or enable us to do: “we define ‘theory’ to mean a broad class of concepts that aim to provide a structure for conceptual explanations or established practice”. Finally, a later definition used by Malmi et al. [37, p. 188] focuses on the quality of the process through which theory is formed: “we defined the concept *theoretical construct* as a theory, model, framework, or instrument developed through application of some rigorous empirical or theoretical approach”. Combining these three approaches, we could define theory—or theoretical construct—as

- a structure for conceptual explanations or established practice, a generalisation, abstraction, explanation, or prediction of a phenomenon
- that is developed through application of some rigorous empirical or theoretical approach
- and can be expressed in the form of a model, framework, or instrument.¹

One sign of CER as a maturing research field is that increasing numbers of papers report the use of theories to guide their research. Malmi et al. [40] found on the basis of data from 2005 to 2011 that just over half of the papers published in CSE, JERIC/TOCE, and ICER explicitly used theories. Often the theories used are borrowed from other fields. The studies by Malmi et al. [40] and Szabo et al. [74] both investigated which theories CER is building on. The findings of these two studies suggest that CER is heavily borrowing theories from other fields, especially from education and psychology. The findings of Szabo et al. [74] suggest that, for

¹ While instruments are generally considered methodological tools to measure something, we take here the perspective that they are theory-informed constructs which support the implementation of some specific theory or theories in research.

example, flow theory, chunking theory, and learning styles were much referenced theories of learning in CER publications. The motivation for borrowing theories from outside the field of CER is readily understandable as many phenomena in the teaching and learning of computing can be explained through non-subject-specific theories such as expectancy-value theory [86].

However, there has been a rising interest in understanding what field-specific theories or field-specifically tuned versions of more generic learning-related theories have been developed within the CER community. Malmi et al. [37, 38] studied which field-specific theories, models, and instruments had been developed within the preceding 10–15 years. The results suggest that the CER community has developed theories of its own, but that new CER-specific theories are still rare. Only 12% of the papers published in ICER, CSE, and TOCE in 2005–2015 proposed a new CER-specific theory. An example of one of the most cited new CER-specific theories in learning is Lopez et al.'s hierarchical model of programming skills [34].

A recent study by Malmi et al. [39] shows that the great majority (65%) of the new CER-specific theories are developed as a result of quantitative research. Qualitative research designs and a combination of literature analysis and argumentation are each used to develop new CER-theories in almost one third of cases. The same study classifies the papers according to the main purpose of the developed theories [20]. The results reveal that only a fraction (5%) of the newly developed CER-specific theories aim at prediction or design and action. The purpose of the great majority of the new CER-specific theories is analysis (rich description of the phenomena), explanation (rich description with explanations but no predictions), and explanation and prediction (predictions and causal explanations).

In light of the recent surveys focusing on theories, there are signs of CER maturing as a research field. Many research publications are based on some existing theory or are presenting a new field-specific theory that the authors have developed. The purpose of many new CE-specific theories is rich description and explanation of what, why, and when something happens; but few publications are able to propose theories that are capable of offering a basis for predictions. This distribution of different kinds of newly developed theory makes sense for a relatively young research field, as we need to begin by gaining a wide understanding of the nature of the phenomena, interactions, and processes in different settings in computing education. As Gregor [20] suggests, analysis theories are the basis of all other types of theory. It takes time for the knowledge to accumulate enough to inform further theoretical development.

Finally, there is one aspect that may hold back the further theoretical development of CER. The recent studies suggest that the community is not building widely on the published theoretical contributions. For instance, over 90% of papers just briefly describe the theoretical construct from the paper they cite, not using the construct or developing it further in their own paper [37, 38]. This same trend, of new theoretical constructs not being cited, used to inform further research, or further developed by others, was also noted in a recent study that analysed 85 papers on six broad topics from ICER, CSEd, and TOCE published between 2005 and 2020 [39].

5 Methodological Development in CER

One specific class of papers identifiable within the body of CER meta-analyses is that of methodological review. An often-cited measure of maturity for an emergent research area is the strength of the methodological underpinning for its research activities and the corresponding judgement of what is considered acceptable for publication. Metareviews that take a structured approach to analysing publications in the field and the methods they employ can contribute to our understanding of areas of strength and weakness in research practice and of trends and developments over time. Literature reviews generally focus on particular topics or themes within a subject area. A methodological review is concerned not with the specific subject-related outcomes that are reported but with how the work has been conducted, what research practices are reported, and on what basis the conclusions have been drawn.

A 2004 review by Valentine [83] that covered 444 SIGCSE papers focusing on CS1 published between 1984 and 2003 is often cited as an early exploration of the research approaches used in computing education. This is not specifically a methodological review (and its own methodological basis has been questioned [66]) but it nevertheless provides an interesting snapshot of CER publications in one particular venue. Valentine distinguished between experimental papers that employ some type of “scientific analysis” (either qualitative or quantitative, but somewhat loosely defined) and those that do not (which are purely descriptive or discursive and which he assigned to one of five other classifications). Valentine found that 94 papers (21%) could be classed as experimental. His conclusions challenged authors to go a step further in their work, not just reporting but also evidencing outcomes.

At about the same time as this challenge, seminal work in CER methodological review was being carried out by Randolph et al., later published in Randolph’s 2007 PhD thesis [57] and in several related studies conducted using a similar approach [56, 58–60]. Randolph’s thesis is the first detailed review of CER methodology that has a broad coverage and that itself employs a carefully articulated methodological approach. The novelty of the work was demonstrated by the findings of the thorough literature review in Randolph’s thesis, which identified just three previous CER methodological reviews: two involving his own work [58, 59], and the third being Valentine’s paper [83]. In the thesis, a stratified random sample of 352 papers was drawn from 1306 articles published between 2000 and 2008 in the CER publications of five conferences (SIGCSE Technical Symposium, Innovation and Technology in Computer Science Education (ITiCSE), ICER, Australasian Computing Education Conference (ACE), and Koli Calling) and three journals (Computer Science Education, SIGCSE Bulletin, and Journal of Computer Science Education Online). The overarching purpose of the research was to determine the methodological characteristics of the articles surveyed. This objective was broken down into specific sub-questions to determine the proportion of articles with human participants and the types of method they involved, the measures and instruments they used, the independent/dependent/mediating variables used, and the characteristics of the paper’s structure; the types and proportions of articles not

involving human participants; the types of research design used in articles taking an experimental approach; and the statistical practices used in articles taking a quantitative approach.

Randolph's analysis was approached from a behavioural sciences perspective using both quantitative and qualitative methods to investigate how (and why) individuals and groups act. This influences the structure of the research questions and explains the differentiation between articles involving human participants and those addressing other types of data. Randolph takes the reasonable stance that when conducting research on human participants "the conventions, standards and practices of behavioural approach should apply". The work looks most closely at papers of this type, which constitute roughly a third of the papers sampled.

The findings of Randolph's work are presented with detailed breakdowns across many variables and comparisons. However, the data overall present an interesting picture regarding the structure, methodology, and reporting in CER articles published between 2000 and 2005 through the lens of a number of accepted measures of robust research. Of the articles without human participants, over 60% were purely descriptive reports of interventions, innovations, etc. Nearly 40% of papers involving human participants provided only anecdotal evidence for the claims made. Of those that were evidenced, nearly two thirds used (quasi-)experimental (quantitative) approaches and over a quarter employed an explanatory descriptive (qualitative) methodology. Within the (quasi-)experimental group, roughly half used a single group post-test method only. That is, there was no pre-test and no control group or comparison. Without such measures, an analysis is much less likely to accurately discover a causal relationship. Dependent variables related most frequently to attitude (60%), with attainment the second most prevalent (52%).

Randolph also noted other issues with research design, such as inadequate amounts of data and results not supporting the conclusions drawn. Questionnaires were the most common measurement instrument, but measures of validity or reliability were given in only one article. Issues were also observed with presentation across all papers in the sample. For example, over 28% did not review existing literature; 78% did not state research questions or hypotheses; and over 63% did not state the purpose of the work.

Randolph's work also evidenced some development of methodology during 5 years spanned by the sampled articles. The proportion of papers with claims backed by purely anecdotal evidence decreased consistently from 58% in 2000 to 27% in 2005. While this may be seen as a welcome indication that published results were becoming more evidence-based, other apparent trends, such as a seeming decline in the use of qualitative research methods, are harder to interpret.

Randolph concluded from his findings that CER was at a crossroads. The high proportion of papers failing to adopt robust research practices could be viewed as helpful in generating hypotheses but not in confirming them. Either the current approach could continue or (as might be considered a mark of developing maturity in a research area) the balance could shift towards more application of rigorous methods. The challenge for CER researchers was clear. Randolph's data and analysis provide a good baseline against which to compare future findings.

However, the behavioural science perspective of the work means that quantitative methods are given greater attention, and baseline information on the robustness of qualitative methods used in CER is lacking [33]. Further, this division does not allow for consideration of mixed methods approaches, whose combination of different perspectives has the potential to deliver rigour through triangulation and insight. This point was taken up by Thota et al. [79], who document a structure for “paradigm pluralism” and present a case study of its application in a computing education research project.

Aspects of methodology have also formed part of the classification criteria used in other meta-studies. In 2010, Malmi et al. [41] sought to establish a broader picture of the kind of work being carried out in CER against a categorisation scheme with seven aspects—theory/model/framework/instrument (TMFI), technology/tool, reference discipline, research purpose, research framework, data source, and analysis method. Several of these dimensions clearly relate to the methodology used. The sample consisted of all 79 ICER papers published over the first 5 years of the conference. This work set out not to critique the application of the methods (for example, to consider control variables used or validation of instruments) but to map the field more generally. It was found that the approach was descriptive in 11% of papers, evaluative in 71%, and formulative (of a novel concept, model etc.) in 18%. Sixty percent of the sample used one or more TMFI, 86% provided some kind of empirical evaluation, and 79% used an identifiable research framework (most commonly a survey). Although the sample is from just one conference, these findings indicate a significant reduction in purely anecdotal evidence. The authors note that the ICER requirements for papers make stipulations on the theoretical and empirical basis of the work and research design. Findings such as these and those of Simon [67] suggest that conference and journal policy plays a major role in promoting a shift towards more rigorous approaches.

Survey instruments and quantitative analysis remain the two main ways in which data is collected and analysed in CER, and several metareviews have considered the ways in which these have been developed and applied. Margulieux et al. conducted a review of 197 CER papers to determine the variables and instruments used and the analysis performed [42]. In this case, the sample consisted of all papers with empirical evaluation involving human participants published in CSE, TOCE, or ICER from 2014 to 2017. The authors note the benefit of standardisation where this is possible, to allow reuse of validated instruments and greater comparability between different studies. It was found that 37 standard instruments had indeed been used, but noted the lack of standardised CER survey instruments for further commonly assessed variables, such as perceptions of the computing field. Overall the authors find that the proportion of papers assessing indicators of learning in CER is in line with that observed in general educational research. However, as with Randolph’s earlier findings, they note that reporting of data collection and analysis often falls short of commonly accepted good practice. For example, 51% of the papers surveyed did not provide information on learner characteristics and nearly 15% did not state the number of participants. Recommendations of aspects of reporting to improve are set out by McGill and Decker [45].

Where survey instruments are used, the level of confidence in results obtained is related to the reliability (consistency of measurement) and validity (assurance that survey measures the intended constructs) of the instrument itself. In a study of 297 CER papers published from 2012 to 2016 across a number of venues, Decker and McGill found that of the 47 different instruments used, 94% were obtainable (either published or available on request), 60% provided some measure of reliability, and 51% provided some rationale for validity (although this was often by expert opinion of face validity) [12]. However, the findings of Heckman et al. [23] paint a less positive picture. They observed that most of a sample of 427 papers from five major CER venues in 2014 and 2015 did not publish the survey questions used. In some cases, there may be good reason for this, such as space restrictions in publications or the desire to prevent future survey participants becoming familiar with the questions in advance. However, this has led to the same concept being measured by multiple different instruments and a tendency to ‘reinvent the wheel’ rather than the community making use of standard instruments that allow comparison of results and for which reliability and validity could be established [12]. One useful resource for CER researchers is the CER database (<https://csedresearch.org>), which includes (in September 2022) a collection of 140 computing-focused instruments (as well as many others) that are classified by topic and can be reused. However, while information on the reliability and validity of these instruments would be a useful guide for researchers, it is not provided.

Where quantitative analysis is performed, the robustness of the work reported also depends not just on using (particular types of) statistical tests but on the appropriateness and correct application of those tests. Further, for both quantitative and qualitative analysis, full information needs to be reported to the reader in order for the method to be transparent and the work to be replicable. Randolph’s work showed a baseline position that allowed considerable scope for improvement in these areas. More recently, research by Sanders et al. [63] looked at the use and reporting of inferential statistics in all 270 ICER papers published between 2005 and 2018. They found that 51% of the sample provided statistical analysis beyond purely descriptive, while 28% either had no data or did not describe it numerically. The authors describe the reporting as “not encouraging”, with very few papers giving sufficient information to adequately describe the analysis. Amongst many deficiencies noted, hypotheses were rarely stated, the precise variant of the statistical test was not given, and there was no indication that the assumptions for the test were met. Sadly, a cautious comparison (given the different venues surveyed) with Randolph’s work from over 10 years earlier showed no improvement in this regard. It seems that when it comes to the robust application of statistical methods, CER has not advanced. This picture is confirmed by the findings of Heckman et al. [23]. While over 82% of the papers published in five major CER venues in 2014 and 2015 were found to provide some empirical evidence, norms for reporting were often not being met, and only a quarter of the papers reported survey results in a manner that was strongly replicable.

Other comparisons with Randolph’s work come to similar conclusions. Lishinski et al. [32] reviewed papers published in CSE and ICER from 2012 to 2015. Of 136

papers found, 110 were judged to be empirical in nature and only eight of these relied on purely anecdotal evidence. However, no significant move was observed towards more rigorous qualitative and quantitative approaches. Results did not suggest any difference in methodological rigour between the journal and conference publications.

These methodological metareviews conducted on publications now spanning more than 20 years show some developments, but also point to other areas where little progress is observed. It is certainly difficult to make direct comparisons, given the different venues from which the samples of papers are drawn and the different interpretations of classifications such as ‘experimental’. However, the broad picture shows that CER has moved away from publications with purely anecdotal evidence, largely meeting Valentine’s challenge to make the extra effort in providing evidence. Journal and conference policies appear to have had a considerable influence in this. On the other hand, repeated studies show that the design methods used are at the weaker end of the spectrum (such as post-test only) and the analysis reported still falls short of expectations in many respects.

6 Analyses of CER Publication Venues

With maturity comes the potential for introspection. Literature from key publication venues for a discipline provides lenses through which to assess the evolution of the discipline and possible future directions.

A number of metareviews have focused on analysing the publications from one or two computing education venues. We found reviews of five prominent computing education conferences (SIGCSE TS, ITiCSE, ACE, NACCQ (a New Zealand conference), and Koli Calling) and the journal *Informatics in Education* (InfEdu). These reviews typically classify papers published at the venues based on their general characteristics and addressed research topics, and investigate any trends over time. Sometimes the metareviews were timed to mark milestones of the particular venue [6, 70]. A summary of the metareviews we discuss is presented in Table 1.

We note that a number of premier publication venues, ICER, CSE, and TOCE, are not among these venue-specific reviews; we have not found any such reviews for them. It is of course conceivable that such reviews have been written, submitted to the venues, and not accepted for publication. As discussed in the previous sections, papers published in these three venues have been analysed in several reviews that have focused on the use and development of theoretical frameworks as well applied research methods. The five venue-specific reviews considered in this section focus on other aspects. A sixth review, on InfEdu, is included because the review considered it as a journal with strong parallels to the Koli Calling conference despite its broader coverage of topics, thus permitting some interesting comparison between the two venues.

Table 1 Summary of venue-focused metareviews

Venue	Acronym	Year of first conference/issue	Metareview time frame	Papers analysed
Conferences				
Special Interest Group in Computer Science Education Technical Symposium	SIGCSE TS	1970	1970–2018	481
Innovation and Technology in Computer Science Education	ITICSE	1996	1996–2019	1295
Australasian Computing Education Conference	ACE	1996	2004–2006 1996–2008	129 328
Koli Calling International Conference on Computing Education Research	Koli	2001	2001–2006 2001–2007	102 130
National Advisory Committee on Computing Qualifications (NZ)	NACCQ	1988	2004–2006 2000–2007	46 157
Journal				
Informatics in Education	InfEdu	2002	2002–2007	121

A series of metareviews of literature published in ACE, NACCQ, InfEdu, and Koli were reported from 2007 to 2009. The analyses of the literature at these venues were conducted using a classification scheme for computing education literature developed by Simon [66]. As mentioned in a preceding section, Simon's scheme comprises four dimensions (*context*, *theme*, *scope* and *nature*) and has subsequently been applied in a number of reviews of the computing education literature, particularly studies of venues.

The first review in this set is an early review of computing education papers published at the ACE and NACCQ conferences in 3 years from 2004 to 2006 [66]. The 175 papers (129 from ACE and 46 from NACCQ) gave insights into the focus of interest of researchers and practitioners in the Australasian computing education

community. A major outcome of this work was the development and trialling of Simon's scheme for classification of computing education literature.

A review by Simon et al. [71] extended this work but focused on the NACCQ conference. In this review, the 157 computing education papers published at NACCQ from 2000 to 2007 included 3 years from the Simon study [66]. As the analysis involved multiple classifiers, the inter-rater reliability of classifications was tested. A system where classifiers worked individually and then in pairs produced fair to good agreement for three dimensions (context, theme, and scope) and excellent agreement for the fourth (nature). Analysis of the NACCQ papers found that the most frequent themes of work reported were teaching, learning, and assessment, with 20% of papers having a theme of teaching/learning techniques, 10% with a theme of teaching/learning tools, and 10% concerning assessment techniques and tools. Curriculum was a strong feature of work at NACCQ, with 15% of papers having this theme. The most common context was programming (13%); however, a high number of papers (30%) had no specific context. In regard to the nature of the research, the most frequent type of paper was report (40%) and almost the same number of papers were classified as research (37%). Almost half the work (45%) was conducted in single courses. The longer time frame of this review allowed for analysis of trends. The main trend reported was for the nature of the work, with a decreasing proportion of report papers and a corresponding increase in research papers.

A similar review by Simon [69] again extended his previous work [66], this time focusing on the ACE conference. The review covered all 10 years of papers published at ACE from 1996 to 2008 (the conference was not held each year during this time), including 3 years from the previous study [66]. Overall, 328 papers were analysed using Simon's scheme. As with NACCQ, the most frequent themes concerned teaching, learning, and assessment; however, the percentages for ACE were higher in each case. At ACE, 34% of papers had a theme of teaching/learning techniques, 15% teaching/learning tools, and 12% assessment. The 11% of papers with a theme of curriculum was lower than was found at NACCQ. At ACE almost a third of the papers (32%) were in the context of programming, considerably higher than at NACCQ (13%). The nature and scope of papers at ACE showed different profiles from NACCQ, with ACE having a higher percentage of reports (70%) and a lower percentage of papers classified as research (23%). Almost two thirds of the work at ACE (64%) was conducted in single courses. The analysis of trends found an increasing proportion of research papers, from 10% in 1996 to nearly 50% in 2008, a similar but stronger trend to that at NACCQ [71].

A further review by Simon [67] applied his scheme to papers published at the Koli conference from 2001 to 2006. The 102 papers comprised the complete set of full papers published since the conference began. As with NACCQ and ACE, the most frequent themes at Koli concerned teaching, learning, and assessment. The profile was similar to that of ACE, with 28% of papers having a theme of teaching/learning techniques, 20% teaching/learning tools, and 13% assessment. The theme of curriculum accounted for 9% of the papers. At Koli almost a quarter of the papers were in the context of programming (25%). The nature and scope of

papers at Koli Calling showed a similar profile to NACCQ, with 47% classified as reports and 35% as research. Around half of the work at Koli (53%) was conducted in single courses. An analysis of trends in the nature of the papers showed increasing proportions of research papers, as was found in the NACCQ [71] and ACE [69] studies. For the first 3 years, research papers were 14%, 10%, and 7% respectively of the full papers published; this then jumped to 47%, 44% and 59% for the next 3 years, reflecting the change of the conference goals as reported in chapter “Computing Education Research in Finland” of this book.

Simon’s scheme was also applied to a comparative analysis of all the papers published in the *Informatics in Education* journal (InfEdu), based in Lithuania, and the Koli conference, held in Finland [68]. The analysis comprised 121 papers in six volumes of InfEdu from 2002 to 2007 and 130 papers at Koli from 2001 to 2007. Although the goals of *Informatics in Education* are broader than those of Koli, encompassing the use of computers in all education rather than just in computing education, the report found many similarities between the themes and contexts of the papers. The most frequent themes in both were teaching/learning techniques (Koli 30%, InfEdu 26%), teaching/learning tools (Koli 19%, InfEdu 12%) and assessment (Koli 12%, InfEdu 13%). The main difference was that InfEdu had higher proportions of papers in the theme of educational technology. The prominent context for Koli was programming (37%), whereas broad-based contexts were most common at InfEdu (26%), with programming the next most frequent (20%). The venues showed some difference in the nature of the papers, with Koli having 52% reports compared with 44% for InfEdu. The main differences with the scope were that Koli had more than half its papers reporting work in a single course (53%) whereas more than half the papers at InfEdu had no applicable scope (57%).

This series of metareviews between 2007 and 2009 provides a unique snapshot of these computing education venues over the preceding decade. There were many similarities between the venues. Each venue had a strong focus on work in the context of programming and themes relating to teaching and learning techniques or tools and assessment. Each venue had an increasing proportion of research papers. However, there were also key differences that indicated the particular focus of each venue; for example, NACCQ’s focus on curriculum and InfEdu’s focus on educational technology.

Recently there have been two larger reviews of leading computing education conferences. A review by Simon and Sheard [70] analysed 24 years of literature from ITiCSE. Motivated by the 25th anniversary of the conference, they applied Simon’s scheme to analyse all full papers and working group reports published at ITiCSE from its inception in 1996 through to 2019. During this time there were 1295 full papers and 129 working group reports published at the conference. The analysis considered working group reports separately from full papers. Analysis of the papers shows that the most common context was programming (38%), with increasing focus on school computing. Teaching and learning techniques (28%) and tools (22%) were the focus of more than half the papers, but there has been an increasing focus on ability, aptitude, and understanding. More than half the papers (53%) report work done in a single course of study. The most common nature of

papers is report (46%), with study (29%) as the next most common. There has been steady growth in work with natures of analysis, study, and experiment.

The 50th year of the SIGCSE Technical Symposium inspired a review by Becker and Quille [6]. The review focused on full papers about introductory programming courses at the university level published at the symposium in the 49 years from 1970 to 2018. Using a systematic search and selection the authors built a list of 481 papers. The analysis examined the focus of each paper, using a framework of 54 sub-categories in eight top-level categories: first languages and paradigms; CS1 design, structure and approach; CS1 content; tools; collaborative approaches; teaching; learning and assessment; and students. Trend analysis showed an increasing focus on students and learning and assessment and a decreasing focus on CS1 design, structure, and approach, and on first language and paradigms.

While these reviews use different approaches to examine the literature of different venues over different time frames, their findings do tend to show some common features:

- within computing education, programming education garners far more attention than education in any other topic area;
- there is broad evidence of a move away from experience reports toward generally empirical research;
- there is evidence of diminishing interest in some topics (such as what programming languages should be used) and increasing interest in others (such as how students learn).

The set of reviews of these key publication venues for computing education research has deepened our knowledge of the literature profiles of the different venues and our understanding of how research in this field has evolved.

7 Discussion

Overall, our analysis of various meta-analysis papers in CER builds a picture of a discipline that is growing in maturity and independence. Our work focuses strongly on the development during the past 20 years, because we found very few relevant meta-analysis papers addressing the field prior to 2000. Moreover, from our perspective, the last two decades are the most relevant years for the maturing field, as the numbers of published papers and submissions have increased substantially during this time. This is well reflected in the ICER conference. In its early years the acceptance rate of papers was around 40% and sometimes more, and during the past few years it has dropped to about 20%. At the same time, the number of participants has increased from around 50 to more than 200. The field is growing rapidly.

We have seen the proportion of research papers increasing in different venues, as discussed in Sect. 6. The growing interest in improving research quality has been clearly visible in the past few years, with the publication of more specific methodological reviews [12, 32, 42, 45, 63], as well as surveys of the use and devel-

opment of theoretical frameworks [37, 38, 40, 74]. These works are complemented by several chapters of the recent Cambridge Handbook of Computing Education Research [19].

The evidence collected in these reviews identifies increased and broadening use of theoretical frameworks from other disciplines, especially from education and psychology, as well as growing interest in developing domain-specific theoretical constructs. Moreover, there has been a clear transition from anecdotal data towards more versatile data collection and analysis. However, there is much room for improvement in methodological rigour. Considering these findings from the point of view of Fensham's criteria (those that are relevant to this chapter), we find that they support the claim that CER fulfills the criteria of "conceptual and theoretical development" and "research methodologies". Moreover, the findings also support partially fulfilling the criterion "progression", as methodological rigour forms a base for any replication studies and theoretical developments support the building of more complex theories and models. However, replication studies are still rare in the field and difficult to carry out [1]. Moreover, the examples of research where some previous theoretical construct is being extended or used to inform new research are still scarce, except for research that applies or further develops theory-informed instruments [39]. Validated instruments are increasingly used in the field, and new instruments are being developed and validated [37–39, 42].

Despite the general interest in research quality, there is still much scope for improving the rigour of research and how it is reported, as pointed out in Sect. 5. One reason for some of the commonly-noted shortcomings in reporting may be the limited space in papers. It is widely acknowledged that reporting qualitative research generally requires more space than reporting quantitative research. While this may be a factor in some cases, rigour is not related solely to the length of the publication. The challenges include how well the research design, data collection, and analysis methods have been documented in the paper, and how well the arguments for these choices have been reported. As discussed above, there are shortcomings in reporting why specific tests were used and whether their applicability for the specific cases was checked. There are shortcomings in describing the demography of human participants, which may be significant for understanding the specifics of the research context. In qualitative research, the process of how categories were formed from the data is sometimes described only vaguely. In content analyses including quantitative results, checking inter-rater reliability or other ways of confirming the classification of data items might not be reported and might not even have been conducted.

Overall, these shortcomings may undermine future research in several ways. While some results might not hold up under closer scrutiny, other researchers might nevertheless build on them in their own work. Moreover, while replication is a powerful tool for confirming and generalising findings, shortcomings in reporting research settings and methods can make replication difficult or even impossible. Furthermore, the results of meta-analyses that summarise findings from original research in are weakened, as it is more difficult to judge the reliability of the individual studies on which they are building.

Another source for these challenges in reporting research is shortcomings in researchers' own competences. Statistics is a broad area with numerous methods, and many computing education researchers are not well versed in it. Statistics is not widely needed in many areas of computing, so experts in those areas who undertake computing education research may have a lot to learn when using statistical analysis methods. Learning analytics is also increasingly used in the field, using data mining and machine learning methods and broadening the scope of quantitative methodological approaches even more.

A similar situation concerns qualitative methods, which can present even more challenges. They are used only in some subareas of computing, such as human-computer interaction and empirical software engineering. However, computing degree programs may have no compulsory courses on these methods. Most researchers naturally extend their methodological competences during their doctoral studies, but they are likely to focus only on the skills that are most relevant for their particular research.

There is also a wider context to the development of CER. It is still the case that, in some departments, computing education research is not accorded the same status as "real" computing research. In such an environment, academic staff may not receive recognition for work in CER and indeed may have to undertake such work outside their full academic workload. Similarly, in many countries, funding for CER projects and PhD studentships is very difficult to come by. These factors do not provide an ideal climate for a research area to flourish and it is to the credit of many academic staff that their commitment to developing the teaching of their subjects leads them to devote their own time to researching and publishing in the area.

Finally, theoretical frameworks in educational sciences and psychology are very rarely addressed in computing research proper, and thus people entering CER with a background in computing have much to learn. This is not limited to learning some specific theory; rather it concerns learning the whole research paradigm of the social sciences, where theoretical frameworks have a much stronger role than in computing. The whole concept of theory is different in the social sciences, when compared with theoretical computer science which builds on a mathematical research tradition with its emphasis on the proof of theorems.

There are several ways in which these challenges might be addressed to improve research quality. First, the organising bodies of conferences and journals have an important role to play in motivating progress. Calls for papers in conferences and instructions for authors in journals give researchers important guidance on how they can improve their work and their prospects of having their papers accepted. However, this is only one perspective. Another important perspective is the organisation of the review process and the competence of reviewers in giving adequate feedback and judgment of submitted papers.

In the past few years there has been a significant increase in both the number of submissions to CER venues and the number of participants in conferences. One factor behind this is the growth in publications addressing computing in schools. This has caused pressure to add more reviewers to the field, and the need to monitor

the quality of reviews. Leading conferences have revised their review processes by introducing senior program committee members or associate program chairs, whose main task is to encourage reviewers, after submitting their reviews, to discuss them with one another to resolve possible conflicting views or to clarify their arguments. The senior members will subsequently write metareviews, which not only summarise the main issues and arguments raised in the reviews to support the proposed acceptance or rejection, but also judge the significance of the issues based on the quality of reviewers' arguments and the metareviewer's own judgement as a senior researcher in the field. This helps program chairs to make their final decisions on acceptance, as well as improving the overall quality of reviews. These roles thus mirror the role of editorial board members in journals, who coordinate reviews of papers assigned to them by the editor-in-chief.

Another development is the clarification of review criteria based on feedback from reviewers and observations made by program chairs. It is vital for review quality that papers are assigned reviewers who know the topic area well and are familiar with the methods applied. Some conferences help in this regard by using a bidding phase, where reviewers can give their preferences for papers they are willing and competent to review. Moreover, conferences frequently tune their review criteria to clarify their interpretation both for reviewers and authors: this also helps to improve the quality of future submissions. An in-depth discussion of review practices in CER is presented by Petre et al. [55].

There are many other ways in which the research community can support its members to build their competence. One traditional practice is organising doctoral consortia for PhD students, which has been a regular practice in ICER, is somewhat less frequent in Koli Calling, and has recently been adopted by ITiCSE. In Lithuania, Vilnius University conducts an extended doctoral consortium annually in Druskininkai, which is targeted to STEM education and educational technology PhD students and has frequently had CER PhD students and senior CER researchers participating. In Europe, the annual SEFI engineering education conference organises doctoral consortia which many CER PhD students have attended.

Another community activity is organising narrow workshops or tutorials around selected methodological themes. An early example was the PhiCER (phenomenography in computing education research) workshop, which was organised twice in 2006–2007 and helped many researchers to learn this method [8]. ICER has frequently supported pre- or post-conference events on various themes. In addition, its work-in-progress workshops allow researchers to present ongoing research and get feedback and support from others. Recently significant initiatives have been the establishment of CSEdResearch.org, which especially supports K-12 level computing education, and the CSEdGrad.org project for supporting PhD students in the field. The former, for example, includes among its resources more than 200 instruments that can be used in research.

While all of the elements above are laudable initiatives helping to attain the goals of improving research quality, there remains room for critique. There is some tension between the requirements for rigorous research and the presentation

of innovations in the field, as pointed out by Nelson and Ko [49]. Theories are obviously strong assets for guiding research, but there is a risk that undue emphasis on theory may restrict the design of new educational innovations, as not all development and research in computing education is guided by existing theories. New domain-specific theories may emerge from the analysis of existing settings and collected data with an open mind rather than relying solely on existing theoretical lenses. These opportunities should not be ignored. Moreover, if all papers must include a rigorous empirical evaluation of new pedagogical innovations, learning tools, or learning resources, this may inhibit the presentation of innovations to the computing education audience. Overall, the work carried out in the field is targeted at improving computing education practice, and there should be space for early presentation of ongoing work. For example, the Koli Calling conference currently has two tracks for papers: research papers can be long, and have strict requirements; while short papers, often called discussion papers, provide an opportunity to report new developments with only preliminary results, as well as opinion pieces that discuss relevant themes supported only by argumentation. The CER field is rich and it is worthwhile to make this richness visible.

8 Recommendations

We conclude by briefly summarising the main recommendations that we have collected from the meta-analyses addressed in previous sections.

- There are as yet few theories targeted at designing new educational activities [39]. More work, as exemplified by Nikula et al. [50] and Xie et al. [87], is needed.
- When theoretical frameworks are cited in papers, their actual role in research design and in analysis and interpretation of results must be made clear. There is currently a reporting problem in that theoretical frameworks and how they are applied in the research are not clearly identified with citations in papers. This follows in part from the practice of considering theories only as related work.
- Research questions and/or hypotheses should be properly reported, as should the goals or purpose of the study itself.
- There are many shortcomings in reporting research, which should be addressed. Shortcomings in the following make replication of studies difficult or impossible and undermine interpretation of the results.
 - Contextual information—where the study was carried out—should where appropriate provide relevant information such as course syllabus, required prerequisite information, course requirements and schedule, and grading principles.
 - Participant demographics, their background, and recruitment practice or incentives should be reported accurately.
 - When using questionnaires, preference should be given to validated instruments. When new questionnaires are designed, some evidence of their

reliability and validity should be reported. The survey questions should ideally be made available for readers.

- Reasons should be given for selecting the analysis methods, such as statistical tests, their variations, and their applicability in the setting.
 - In qualitative settings, the methods used for building categories should be explained.
- Publication venues have an important role in supporting research quality in terms of the instructions they provide for authors, the page limits, and improving the quality of the review process and the competency of reviewers.
 - Various research training activities, such as pre- or post-conference workshops and doctoral consortia, can play an important role in supporting and further developing the competencies of community members.

References

1. Ahadi, A., Hellas, A., Ihanola, P., Korhonen, A., Petersen, A.: Replication in computing education research: researcher attitudes and experiences. In: Proceedings of the 16th Koli calling international conference on computing education research, pp. 2–11 (2016)
2. Alaqsam, A., Ghabban, F., Ameerbakhsh, O., Alfadli, I., Fayez, A.: Current trends in online programming languages learning tools: a systematic literature review. *Journal of Software Engineering and Applications* **14**(7), 277–297 (2021)
3. Atchison, W.F., Conte, S.D., Hamblen, J.W., Hull, T.E., Keenan, T.A., Kehl, W.B., McCluskey, E.J., Navarro, S.O., Rheinboldt, W.C., Schweppe, E.J., Viavant, W., Young, D.M.: Curriculum 68: recommendations for academic programs in computer science: a report of the ACM curriculum committee on computer science. *Communications of the ACM* **11**(3), 151–197 (1968)
4. Austing, R.H., Barnes, B.H., Bonnette, D.T., Engel, G.L., Stokes, G.: Curriculum recommendations for the undergraduate program in computer science: a working report of the ACM committee on curriculum in computer sciences. *ACM SIGCSE Bulletin* **9**(2), 1–16 (1977)
5. Bandura, A.: Self-efficacy mechanism in human agency. *American Psychologist* **37**(2), 122 (1982)
6. Becker, B.A., Quille, K.: 50 years of CS1 at SIGCSE: a review of the evolution of introductory programming education research. In: 50th Technical Symposium on Computer Science Education, pp. 338–344 (2019)
7. Ben-Ari, M., Bednarik, R., Levy, R.B.B., Ebel, G., Moreno, A., Myller, N., Sutinen, E.: A decade of research and development on program animation: the Jeliot experience. *Journal of Visual Languages & Computing* **22**(5), 375–384 (2011)
8. Berglund, A., Box, I., Eckerdal, A., Lister, R., Pears, A.: Learning educational research methods through collaborative research: the PhICER initiative. In: Tenth Australasian Computing Education Conference, pp. 35–42 (2008)
9. Boustedt, J.: Students’ understanding of the concept of interface in a situated context. *Computer Science Education* **19**(1), 15–36 (2009)
10. Brown, N.C., Altadmri, A., Sentance, S., Kölling, M.: Blackbox, five years on: an evaluation of a large-scale programming data collection project. In: 14th International Computing Education Research Conference, pp. 196–204 (2018)
11. Chahal, K.K., Kaur, A., Saini, M.: Empirical studies on using pair programming as a pedagogical tool in higher education courses: a systematic literature review. *Research and Evidence in Software Engineering*, pp. 251–286 (2021)

12. Decker, A., McGill, M.M.: A topical review of evaluation instruments for computing education. In: 50th Technical Symposium on Computer Science Education, pp. 558–564 (2019)
13. Denning, P.J., Tedre, M.: *Computational Thinking*. MIT Press (2019)
14. Dorn, B., Elliott Tew, A.: Empirical validation and application of the computing attitudes survey. *Computer Science Education* **25**(1), 1–36 (2015)
15. Du, Y., Luxton-Reilly, A., Denny, P.: A review of research on parsons problems. In: *Proceedings of the Twenty-Second Australasian Computing Education Conference*, pp. 195–202 (2020)
16. Dweck, C.S.: *Self-theories: their role in motivation, personality, and development*. Psychology Press (2013)
17. Fensham, P.J.: *Defining an Identity: The Evolution of Science Education as a Field of Research*. Springer Science & Business Media (2004)
18. Fincher, S., Petre, M.: *Computer Science Education Research*. CRC Press (2004)
19. Fincher, S.A., Robins, A.V.: *The Cambridge Handbook of Computing Education Research*. Cambridge University Press (2019)
20. Gregor, S.: The nature of theory in information systems. *MIS Quarterly*, pp. 611–642 (2006)
21. Guzdial, M.: Exploring hypotheses about media computation. In: *Ninth International Computing Education Research Conference*, pp. 19–26 (2013)
22. Guzdial, M., du Boulay, B.: The history of computing. *The Cambridge Handbook of Computing Education Research* (2019) **11** (2019)
23. Heckman, S., Carver, J.C., Sherriff, M., Al-Zubidy, A.: A systematic literature review of empiricism and norms of reporting in computing education research literature. *ACM Transactions on Computing Education* **22**(1), 1–46 (2021)
24. Hundhausen, C.D., Douglas, S.A., Stasko, J.T.: A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing* **13**(3), 259–290 (2002)
25. Kaplan, A., Maehr, M.L.: The contributions and prospects of goal orientation theory. *Educational Psychology Review* **19**(2), 141–184 (2007)
26. Kelleher, C., Pausch, R.: Lowering the barriers to programming: a taxonomy of programming environments and languages for novice programmers. *Computing Surveys (CSUR)* **37**(2), 83–137 (2005)
27. Kinnunen, P., Meisalo, V., Malmi, L.: Have we missed something? Identifying missing types of research in computing education. In: *Sixth International Computing Education Research Workshop*, pp. 13–22 (2010)
28. Kinnunen, P., Simon, B.: My program is ok – am I? Computing freshmen’s experiences of doing programming assignments. *Computer Science Education* **22**(1), 1–28 (2012)
29. Kölling, M., Quig, B., Patterson, A., Rosenberg, J.: The BlueJ system and its pedagogy. *Computer Science Education* **13**(4), 249–268 (2003)
30. Kong, S.C., Chiu, M.M., Lai, M.: A study of primary school students’ interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education* **127**, 178–189 (2018)
31. Lewis, C.M.: Exploring variation in students’ correct traces of linear recursion. In: *Tenth International Computing Education Research Conference*, pp. 67–74 (2014)
32. Lishinski, A., Good, J., Sands, P., Yadav, A.: Methodological rigor and theoretical foundations of CS education research. In: *12th International Computing Education Research Conference*, pp. 161–169 (2016)
33. Lister, R.: The Randolph thesis: CSEd research at the crossroads. *SIGCSE Bulletin* **39**(4), 16–18 (2007)
34. Lopez, M., Whalley, J., Robbins, P., Lister, R.: Relationships between reading, tracing and writing skills in introductory programming. In: *Proceedings of the Fourth International Workshop on Computing Education Research*, pp. 101–112 (2008)
35. Lukkariinen, A., Malmi, L., Haaranen, L.: Event-driven programming in programming education: a mapping review. *ACM Transactions on Computing Education* **21**(1), 1–31 (2021)
36. Luxton-Reilly, A., Simon, Albluwi, I., Becker, B.A., Giannakos, M., Kumar, A.N., Ott, L., Paterson, J., Scott, M.J., Sheard, J., Szabo, C.: *Introductory programming: a systematic literature review*. In: *ITiCSE 2018 Working Group Reports*, pp. 55–106 (2018)

37. Malmi, L., Sheard, J., Kinnunen, P., Simon, Sinclair, J.: Computing education theories: what are they and how are they used? In: 15th International Computing Education Research Conference, pp. 187–197 (2019)
38. Malmi, L., Sheard, J., Kinnunen, P., Simon, Sinclair, J.: Theories and models of emotions, attitudes, and self-efficacy in the context of programming education. In: 16th International Computing Education Research Conference, p. 36–47 (2020)
39. Malmi, L., Sheard, J., Kinnunen, P., Simon, Sinclair, J.: Development and use of domain-specific learning theories, models and instruments in computing education. *ACM Transactions on Computing Education*, **23**(1), Article 6, pp. 1–48 (2023)
40. Malmi, L., Sheard, J., Simon, Bednarik, R., Helminen, J., Kinnunen, P., Korhonen, A., Myller, N., Sorva, J., Taherkhani, A.: Theoretical underpinnings of computing education research: what is the evidence? In: Tenth International Computing Education Research Conference, pp. 27–34 (2014)
41. Malmi, L., Sheard, J., Simon, Bednarik, R., Helminen, J., Korhonen, A., Myller, N., Sorva, J., Taherkhani, A.: Characterizing research in computing education: a preliminary analysis of the literature. In: Sixth International Computing Education Research Workshop, pp. 3–12 (2010)
42. Margulieux, L., Ketenci, T.A., Decker, A.: Review of measurements used in computing education research and suggestions for increasing standardization. *Computer Science Education* **29**(1), 49–78 (2019)
43. McCauley, R., Grissom, S., Fitzgerald, S., Murphy, L.: Teaching and learning recursive programming: a review of the research literature. *Computer Science Education* **25**(1), 37–66 (2015)
44. McGill, M.M., Decker, A.: Construction of a taxonomy for tools, languages, and environments across computing education. In: 16th International Computing Education Research Conference, pp. 124–135 (2020)
45. McGill, M.M., Decker, A.: A gap analysis of statistical data reporting in K-12 computing education research: recommendations for improvement. In: 51st Technical Symposium on Computer Science Education, pp. 591–597 (2020)
46. McGill, M.M., Decker, A.: Tools, languages, and environments used in primary and secondary computing education. In: 25th Conference on Innovation and Technology in Computer Science Education, pp. 103–109 (2020)
47. Morrison, K., van der Werf, G.: Editorial. *Educational Research and Evaluation* **18**(5), 399–401 (2012)
48. Naps, T.L., Rößling, G., Almstrum, V., Dann, W., Fleischer, R., Hundhausen, C., Korhonen, A., Malmi, L., McNally, M., Rodger, S., Velásquez-Iturbide, J.Á.: Exploring the role of visualization and engagement in computer science education. In: ITiCSE 2002 Working Group Reports, pp. 131–152 (2002)
49. Nelson, G.L., Ko, A.J.: On use of theory in computing education research. In: 14th International Computing Education Research Conference, pp. 31–39 (2018)
50. Nikula, U., Gotel, O., Kasurinen, J.: A motivation guided holistic rehabilitation of the first programming course. *ACM Transactions on Computing Education* **11**(4), 1–38 (2011)
51. Papert, S.A.: *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books (2020)
52. Park, T.H., Saxena, A., Jagannath, S., Wiedenbeck, S., Forte, A.: Towards a taxonomy of errors in HTML and CSS. In: Ninth International Computing Education Research Conference, pp. 75–82 (2013)
53. Parsons, D., Haden, P.: Parson’s programming puzzles: a fun and effective learning tool for first programming courses. In: Eighth Australasian Computing Education Conference, pp. 157–163 (2006)
54. Peng, J., Yuan, B., Spector, J.M., Wang, M.: Integrating technology in programming learning and instruction: a critical review. *International Journal of Smart Technology and Learning* **1**(4), 323–343 (2019)
55. Petre, M., Sanders, K., McCartney, R., Ahmadzadeh, M., Connolly, C., Hamouda, S., Harrington, B., Lumbroso, J., Maguire, J., Malmi, L., McGill, M.M., Vahrenhold, J.: Mapping the landscape of peer review in computing education research. In: ITiCSE 2020 Working Group Reports, pp. 173–209 (2020)

56. Randolph, J., Bednarik, R., Silander, P., Gonzalez, J., Myller, N., Sutinen, E.: A critical analysis of the research methodologies reported in the full papers of the proceedings of ICALT 2004. In: Fifth International Conference on Advanced Learning Technologies, pp. 10–14 (2005)
57. Randolph, J.J.: Computer science education research at the crossroads: a methodological review of computer science education research, 2000–2005. Utah State University (2007)
58. Randolph, J.J.: A methodological review of the program evaluations in K-12 computer science education. *Informatics in Education* **7**(2), 237–258 (2008)
59. Randolph, J.J., Bednarik, R., Myller, N.: A methodological review of the articles published in the proceedings of Koli Calling 2001-2004. In: Fifth Finnish/Baltic Sea Conference on Computer Science Education, pp. 103–109 (2005)
60. Randolph, J.J., Julnes, G., Sutinen, E., Lehman, S.: A methodological review of computer science education research. *Journal of Information Technology Education: Research* **7**(1), 135–162 (2008)
61. Reeves, S., Albert, M., Kuper, A., Hodges, B.D.: Why use theories in qualitative research? *British Medical Journal* **337** (2008)
62. Research Council of Norway: The Role of Theory in Educational Research – Report from the March Seminar 2011 (2012)
63. Sanders, K., Sheard, J., Becker, B.A., Eckerdal, A., Hamouda, S., Simon: Inferential statistics in computing education research: a methodological review. In: 15th International Computing Education Research Conference, pp. 177–185 (2019)
64. Shaffer, C.A., Cooper, M.L., Alon, A.J.D., Akbar, M., Stewart, M., Ponce, S., Edwards, S.H.: Algorithm visualization: the state of the field. *ACM Transactions on Computing Education* **10**(3), 1–22 (2010)
65. Sim, T.Y., Lau, S.L.: Online tools to support novice programming: a systematic review. In: Second Conference on e-Learning, e-Management and e-Services (IC3e), pp. 91–96 (2018)
66. Simon: A classification of recent Australasian computing education publications. *Computer Science Education* **17**(3), 155–169 (2007)
67. Simon: Koli Calling comes of age: an analysis. In: Seventh Baltic Sea Conference on Computing Education Research (Koli Calling 2007), pp. 119–126 (2008)
68. Simon: Informatics in Education and Koli Calling: a comparative analysis. *Informatics in Education* **8**(1), 101–114 (2009)
69. Simon: Ten years of the Australasian Computing Education Conference. In: 11th Australasian Computing Education Conference, p. 157–164. AUS (2009)
70. Simon, Sheard, J.: Twenty-four years of ITiCSE papers. In: 25th Conference on Innovation and Technology in Computer Science Education, p. 5–11. Association for Computing Machinery, New York, NY, USA (2020)
71. Simon, S., Sheard, J., Carbone, A., De Raadt, M., Hamilton, M., Lister, R., Thompson, E.: Eight years of computing education papers at NACCQ. National Advisory Committee on Computing Qualifications (2008)
72. Soloway, E., Ehrlich, K.: Empirical studies of programming knowledge. *IEEE Transactions on Software Engineering* (5), 595–609 (1984)
73. Sorva, J., Karavirta, V., Malmi, L.: A review of generic program visualization systems for introductory programming education. *ACM Transactions on Computing Education* **13**(4), 1–64 (2013)
74. Szabo, C., Falkner, N., Petersen, A., Bort, H., Cunningham, K., Donaldson, P., Hellas, A., Robinson, J., Sheard, J.: Review and use of learning theories within computer science education research: primer for researchers and practitioners. In: ITiCSE 2019 Working Group Reports, pp. 89–109 (2019)
75. Tedre, M., Simon, Malmi, L.: Changing aims of computing education: a historical survey. *Computer Science Education* **28**(2), 158–186 (2018)
76. Tedre, M., Sutinen, E.: Three traditions of computing: what educators should know. *Computer Science Education* **18**(3), 153–170 (2008)
77. Thomas, L., Eckerdal, A., McCartney, R., Moström, J.E., Sanders, K., Zander, C.: Graduating students’ designs: through a phenomenographic lens. In: Tenth International Computing Education Research Conference, pp. 91–98 (2014)

78. Thompson, E., Kinshuk: The nature of an object-oriented program: how do practitioners understand the nature of what they are creating? *Computer Science Education* **21**(3), 269–287 (2011)
79. Thota, N., Berglund, A., Clear, T.: Illustration of paradigm pluralism in computing education research. In: 14th Australasian Computing Education Conference (2012)
80. Tsai, M.J., Wang, C.Y., Hsu, P.F.: Developing the computer programming self-efficacy scale for computer literacy education. *Journal of Educational Computing Research* **56**(8), 1345–1360 (2019)
81. Umapathy, K., Ritzhaupt, A.D.: A meta-analysis of pair-programming in computer programming courses: implications for educational practice. *ACM Transactions on Computing Education* **17**(4), 1–13 (2017)
82. Urquiza-Fuentes, J., Velázquez-Iturbide, J.A.: Pedagogical effectiveness of engagement levels—a survey of successful experiences. *Electronic Notes in Theoretical Computer Science* **224**, 169–178 (2009)
83. Valentine, D.W.: CS educational research: a meta-analysis of SIGCSE Technical Symposium proceedings. *ACM SIGCSE Bulletin* **36**(1), 255–259 (2004)
84. Vessey, I., Ramesh, V., Glass, R.L.: A unified classification system for research in the computing disciplines. *Information and Software Technology* **47**(4), 245–255 (2005)
85. Weinberg, G.M.: *The Psychology of Computer Programming*. Van Nostrand Reinhold New York (1971)
86. Wigfield, A., Eccles, J.S.: Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology* **25**(1), 68–81 (2000)
87. Xie, B., Loksa, D., Nelson, G.L., Davidson, M.J., Dong, D., Kwik, H., Tan, A.H., Hwa, L., Li, M., Ko, A.J.: A theory of instruction for introductory programming skills. *Computer Science Education* **29**(2-3), 205–253 (2019)
88. Yuen, T.T., Robbins, K.A.: A qualitative study of students' computational thinking skills in a data-driven computing class. *ACM Transactions on Computing Education* **14**(4), 1–19 (2014)