Published: 10 November 2025

Citation in BibTeX format

Koli Calling '25: 25th Koli Calling International Conference on Computing Education Research
November 11 - 16, 2025
Koli, Finland

RESEARCH-ARTICLE

From Marco to Maria: Ten Years of the Computing Education Practice Conference

STEVEN BRADLEY, Durham University, Durham, County Durham, U.K.

ROSANNE ENGLISH, University of Strathclyde, Glasgow, Scotland, U.K.

SALLY FINCHER, University of Kent, Canterbury, Kent, U.K.

DUNCAN HULL, The University of Manchester, Manchester, Greater Manchester, U.K.

MARK ZARB, Robert Gordon University, Aberdeen, Scotland, U.K.

Open Access Support provided by:

Durham University

The University of Manchester

University of Kent

University of Strathclyde

Robert Gordon University

Koli Calling '25: Proceedings of the 25th Koli Calling International Conference on Computing Education Research (November 2025)
https://doi.org/10.1145/3769994.3770003
ISBN: 9798400715990

From Marco to Maria: Ten Years of the Computing Education Practice Conference

Steven Bradley
Durham University
Durham, United Kingdom
s.p.bradley@durham.ac.uk

Rosanne English
University of Strathclyde
Glasgow, Scotland Uk
rosanne.english@strath.ac.uk

Sally Fincher
University of Kent
Kent, United Kingdom
S.A.Fincher@kent.ac.uk

Duncan Hull
University of Manchester
Manchester, United Kingdom
Duncan.Hull@manchester.ac.uk

Mark Zarb
Robert Gordon University
Aberdeen, Scotland Uk
m.zarb@rgu.ac.uk

Abstract

There has been a relentless march towards research-focussed papers in Computing Education conferences, influenced by Valentine's pejorative "Marco Polo" characterisation of practice papers. We argue that there is an important place for practice-focussed papers when presented in a format which provides careful contextual description, as exemplified by the work of seventeenth century ecologist Maria Sibyella Merian.

Computing Education Practice (CEP) is a UK practice-based conference that has its tenth year in 2026. In this paper we discuss and describe the CEP reporting format, and review the character and contribution of the CEP series.

We explore a) the structure and content of CEP papers in relation to papers at comparable conferences (Koli, ITiCSE and ICER) using Simon's classification system; b) what kinds of institution are represented at CEP; and c) CEP's influence on participants' teaching, through a (mostly) qualitative survey. We find that a) CEP papers are similar in scope, theme and context to other computing education conferences despite the nature of papers showing an expected and distinctive preponderance of reports; b) CEP is more representative of the full range of UK computing departments than comparator conferences, when considering average student entry standards; and c) participants identify value in participation and discussions as well as some specific contributions of the publications.

We champion the value of carefully constructed reports of practice in computing education.

CCS Concepts

• **Social and professional topics → Computing education**; • **General and reference → Empirical studies**.

Keywords

Computing, Education, Practice, Research, Representation

ACM Reference Format:
Steven Bradley, Rosanne English, Sally Fincher, Duncan Hull, and Mark Zarb. 2025. From Marco to Maria: Ten Years of the Computing Education Practice Conference. In *25th Koli Calling International Conference on Computing Education Research (Koli Calling '25), November 11–16, 2025, Koli, Finland.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3769994.3770003

Koli Calling '25, Koli, Finland

ACM ISBN 979-8-4007-1599-0/25/11
https://doi.org/10.1145/3769994.3770003

1 Introduction

1.1 Why start a practice conference?

In 2017, when the Computing Education Practice (CEP) conference was founded, there were already several conferences dedicated to Computing Education — the SIGCSE Technical Symposium (TS), ITiCSE, Koli Calling, ICER and (to a lesser extent) FiE. All of these have a "research" focus, some explicitly (Koli) and some exclusively (ICER) while for others the research component is only one part of the conference, but the overall aim is still generalised knowledge: "problems common among educators working to develop, implement and/or evaluate computing programs, curricula, and courses" (SIGCSE TS)[31]. There seemed no pressing need to introduce another research conference. At the same time, however, there remained unmet needs. Firstly the knowledge and skills required to be a good teacher of computing are not the same as those of a good researcher of computing education (although both may be present in a given individual). Secondly, many colleagues (often teaching-focussed) simply did not attend these conferences or read their output.

This phenomenon, of teachers not engaging with educational research (even if conducted specifically to assist them) is not unique to Computing, and is widely commented on. Several reasons have been suggested as to why this should be: lack of time; lack of relevance; and lack of accessibility are all cited as possible causes [18, 43]. Some ways to address the disparity have included "make research more readable" and to have "research translators" that interpret a paper so that the results are directly implementable by teachers [22], neither of which seem practical. However, alongside this situation, we knew from previous projects that teachers value and benefit from structured ways to share practice with other teachers, and in turn learn from them [13, 15].

Consequently, we chose to address the education research → practice problem more directly by establishing a conference where the focus was centrally on scaffolded description of contextual practice. In doing this we hoped that a) teachers would recognise its value (and so read and write CEP papers) forming a body of work richly detailing UK computing teaching in a comparable way; b) teachers would be able to cite the source of a new idea (so providing evidence of wider contribution and accruing value to the originators); and thus c) a community of discussion and exchange would form around the principled examination of practice.

1.2 What is practice and why it isn't research

Focussing on the practice of teaching had implications both for the content of the CEP conference and for its presentation. Descriptions of practice are fundamentally different from reports of research. Research necessitates pre-determined questions, objective measures, generalisations and abstraction across contexts. The claim of all research, in effect, is to say "if you do the same things that I did, in the way that I describe, you will get the same result". This is not true for teaching, where replication is not guaranteed by scientific method. As Shulman observes "most teachers are better at some things than at others; they teach some topics better than others. They do a better job on some days than others." [6] And if a teacher takes on a practice from another institution, the problem is further compounded because practice is intimately connected to its originating context, and context, as has been previously described, is *evolutionary*: "By this we do not mean that it develops over time in a series of very small incremental changes (although that, too, is true) but that the final product (which is a piece of practice) is uniquely designed in synergy with the contextual niche in which it was created. Equally, evolutionary products cannot be reverse-engineered. You cannot look at a thing and say 'Oh, OK, we'll grow one just like that too' and then re-create the particular combination of environment and forces which produced it." [13]

Taking the notion of an "evolutionary context" seriously for a moment, it becomes obvious that if you want to "transfer" practice from one teacher to another, from one institution to another, then there is much to take into account *in addition to* the details of the practice itself. A better metaphor than "transfer" might be "transplant". A gardener grows some wonderful lupins. Another gardener may genuinely admire them *and at the same time* know that they will not grow in their own soil (because they know it to be too acid, too hot, too stony or otherwise inimical). Teachers make these kinds of decisions on practice very quickly, a behaviour we have previously characterised as "deal-breaking": "The reaction to material presented – *that wouldn't work for me* - or its recto face - *I'd like to try that* – is often immediate, much too fast to represent an analysis of the background and content of the proposed practice, or to be the response of an extended consideration." [12]. So, if refusal of a practice is not associated with details of the practice itself, it follows that the reason must be located in a deep understanding of the local environment, its particularity and idiosyncrasies (including staff). The reaction comes first and the meaning, the rationale, (if any is required) comes after.

So, for practice to be understandable (and thus accessible to others) it needs to be presented in a way that exposes the salient aspects. This is not well served by the traditional paper format of Literature, Research Questions, Methodology, Results, Discussion, Further Work. As Brown and Duguid note: "Abstractions detached from practice distort or obscure intricacies of that practice. Without a clear understanding of those intricacies and the role they play, the practice itself cannot be well understood, engendered (through training), or enhanced (through innovation)" [8].

1.3 Success in teaching

Success in teaching is not driven by the scientific method, rather by the acquisition and development of Pedagogical Content Knowledge (PCK), the combination of content knowledge and pedagogic skill that is the preserve of the teacher. PCK is a well-established concept which has well-established characteristics. See Gess-Newsome [17] and Berry [5] for extensive research explorations of the concept, Bullough [9] for the history and evolution of the idea, and Robins [14] for an overview in the context of computing education. PCK is, in Shulman's initial definition [30], embodied in teacher-knowledge, in the "wisdom of practice". It is also contingent, in that it only emerges in the relationship of a particular teacher to her students, and is different for different students, for different levels and for different years. A teacher may know eleven, twelve or thirteen ways to teach recursion, but she will only utilise them when the first two, three or four explanations fail. In this way PCK is individual and non-transferable (you can augment your own, of course). It is often also tacit: 'That's the way we do things here' or 'I do it like that because that's what works'. Most importantly, it is always intricately situated in subject matter: teaching for-loops is different to teaching irregular French verbs. It is this subject-specific quality that lends power. As Shulman says "If I'm having a devil of a time teaching introduction to fractions, I don't want to read all the case literature of teaching. I want some good cases of fraction teaching" which is obvious enough. He continues "And I'll tell you something. If I'm having trouble teaching fractions and you give me a set of cases of fraction teaching, you won't have to motivate me to look at them, because my self-respect is on the line tomorrow." [6]

1.4 The Race to Research

The "race to research" in Computing Education arguably started in 2004, not because that was the first time Computing Education research was done, or published, but because David Valentine published "CS Educational Research: A Meta-Analysis of SIGCSE Technical Symposium Proceedings." [42] In this he put papers from the SIGCSE Symposium into six categories, the most damning, and the one that stuck most fast, being "Marco Polo" papers, so-called because they have the same character as Marco Polo's chronicles - "I went there, I saw that, I did this" - with no further evidence adduced, and no concern as to whether others might have seen the same things, and done the same things (or seen the same things and done something different). This is troublesome for a conference focussed on practice, but one of the significant deficiencies of papers characterised in this way is that they are widely diffuse, in form as well as subject. Their authors have no common expectations to adhere to, no exemplars to follow.

In an attempt to redress this, we sought another model and propose Maria Sibylla Merian as a positive eponymous category for practice papers. Maria, a seventeenth century naturalist, was famous for close observation of, and documentation of, the natural world. Not only did she study and illustrate insects and plants but was careful to record them in their ecological context, "While other contemporaries worked with dried specimens, Merian's depictions of metamorphosis were based on her own meticulous observations of living creatures and the plants that sustained them." [26] She was not herself a theorist, but the existence of her careful studies have had a lasting impact on the field, and especially influenced the work of Linneaus and his work on taxonomies, some 50 years later. We believe that this is a role model that more accurately and more usefully captures the importance of practice papers.

1.5 The CEP format

Against this background, we sought a format that would foreground PCK, while preserving sufficient detail to make the contextual visible and the embedded apprehensible. In the spirit of Maria's careful observation and reporting, we framed the structure of CEP papers to both scaffold presentation, and allow comparison between reported practices, using the following guidelines:

What is it?	A short description of the practice you're presenting
Why are you doing it?	What happened before? What is it changing / replacing / improving? What gap is it filling?
Where does it fit?	A short description of your teaching context. You may, for instance, include a description of intake, class size, curriculum sequence; anything that's necessary for others to understand your situation. How do things work at your institution.
Does it work?	How do you know? Give some evidence of effectiveness in context.
Who else has done this?	Where did you get the idea from? (If from published reports, please include references). How did you find out about it? Was it easy/hard to adopt? What did you change?
What will you do next?	Will you vary this, or develop it further?
Why are you telling us this?	What is interesting, or useful, about this to someone else?

1.6 Research Questions

The purpose of this work is to examine the focus and range of contributions to papers from CEP compared with other CS education conferences, as well as to understand how participants perceive the value of CEP. Consequently, a mixed methods approach is used to combine quantitative analysis of conference data with qualitative insights from CEP corpus categorisation analysis using Simon's classification system (see below for more details) and participant experiences. This enables a fuller understanding of the content, context and perceived contribution of the conference.

RQ1 What are CEP papers about, and are these any different to comparable conferences?
RQ2 Which institutions have contributed to CEP, and are these any different to comparable conferences?
RQ3 What is the perceived value of CEP for participants?

1.7 RQ1 rationale

In 2007, Simon proposed a system for classifying Computing Education publications which had four dimensions: Context, Nature, Scope and Topic. From 2009 to 2019 this system was used extensively by Simon himself, and by varying groups of Finnish researchers (often including Simon's PhD supervisor, Lauri Malmi) to classify papers in a variety of Computing Education conferences and journals. In 2015, Simon revised the system as part of his PhD thesis "Emergence of computing education as a research discipline". In this he defined four dimensions of classification: Nature, Scope, Context and Theme. Two of these, Nature and Scope are "fixed", that is, they may only contain the values prescribed.

- The *Nature* of a paper may be: experiment, study, analysis, report, position/proposal.
- The *Scope* of a paper may be: subject (within a single subject/course/unit/paper at a single institution), program/ department, institution, many institutions or not applicable

The two other dimensions, Context and Theme are "dynamic", that is they have suggested contents, but may be expanded as other areas become of interest to authors.

- The *Context* of a paper is the academic setting (often the subject of a single course, or module)
- The *Theme* of a paper is the topic of interest, what the paper is about

We used this system to categorise the CEP corpus. We hoped that using the Context and Theme dimensions (and their suggested categories) would provide a useful point of comparison with reports from other studies which have used the same classification to categorise computing education research (CER) papers, allowing us to ask whether authors of research papers are interested in the same subject matter as authors of practice papers. We anticipated that the Nature and Scope dimensions would be limited for our purposes in characterising practice, offering only "report" and "subject" as likely categories.

1.8 RQ2 rationale

Authors of CER papers commonly make use of "convenience sampling", that is, they study the students they are teaching. Generalisation from such studies rests on the assumption that "first year" students (or whatever group is under study) are the same, no matter the institution they attend. However, different institutions attract different kinds of student, with different backgrounds and abilities, and this approach can ultimately lead to showcasing an unrepresentative sample of institutions. In the UK, average departmental student entry standards (exam grades) are strongly correlated with departmental research quality, as can be seen in Figure 1[1]. So institutions with a stronger culture of research publication may be dealing with different kinds of students. Conversely it may be that, given the pressure to publish, teaching-focussed institutions may publish more about things "they are doing anyway" i.e. teaching. The representation of different types of institutions is important not only because of bias in sampling when looking at the impact of different pedagogical approaches or innovations from a research perspective, but also because of the types of interventions or innovations that are reported. Appropriate pedagogical strategies for teaching

[1]We can mainly use "institution" and Computer Science Department (CS dept) interchangeably in this paper. Nearly all of the institutions are universities, and it is very rare for a university to have more than one department offering Computer Science programmes

students who have not excelled in the education system to date, for whatever reason, are likely to be different to those who have a solid foundational understanding, good motivation, and fewer barriers to learning. Comparing institutional representation in CEP with other conferences helps us understand how representative they are of different types of teachers and students.

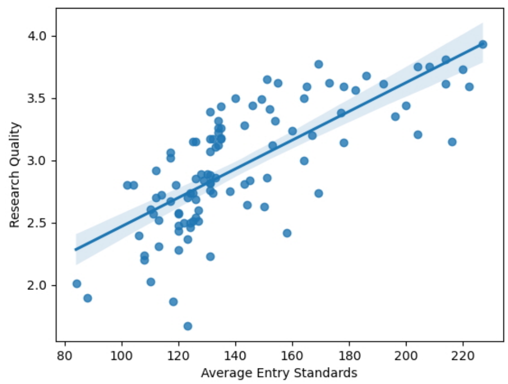

Figure 1: Undergraduate average entry standards (UCAS 2021/22) vs research quality (REF 2021) for UK Computer Science departments. Data from Complete University Guide Subject Tables 2025 [10]

1.9 RQ3 rationale

The Computing Education Practice conference was designed with improving sharing of computing teaching practice in mind, however thus far there has only been anecdotal evidence of the value of the conference. RQ3 aims to formalise this, gathering insight on perceived value through increased connections and re-use of practice as self-reported by participants.

2 Previous Studies

Many earlier studies have reviewed the contributions to computing education conferences — see Table 1. We are particularly interested in those that inform the methodologies for RQ1 and RQ2. These studies are a combination of qualitative analysis when looking at the content of the papers and quantitative when exploring the individuals, institutions and countries of those contributing to the papers. While some progress has been made with automated qualitative analysis using techniques such as topic modelling [3] and large language models [28], this is still usually carried out by reading the relevant papers. To allow our results for RQ1 to be compared with previous work, we wanted a classification system that had been used as widely as possible, so we opted for Simon's system [35] rather than classifications used in only one paper (labelled "Bespoke" in Table 1) or those based on keywords arrived at inductively, which necessarily vary from paper to paper. Some studies included bibliometric analysis of authors and their countries, but we extracted data direct from bibliographic databases to compare with our results for RQ2. In the next section we outline our research methods, including our choice of comparator conferences and data sources.

3 Overview and Integration of Methods

The first CEP conference was held in 2017, but the first publication with ACM DL was in 2019. For the first conference, in 2017, we only have access to titles and affiliations for talks; for 2018 we have titles, abstracts and affiliations. Where comparisons are made with other publication venues, we consider other Computing Education conferences with relevant national, European or international focus: UKICER, Koli Calling, ITiCSE, ICER. We did not include other conferences with a non-European focus, such as SIGCSE TS, CompEd and ACE.

RQ1 (paper classification) is a qualitative analysis of CEP papers from 2018 onwards, which we then relate to published analyses of the comparator conferences (where available — not usually over the same time period). RQ2 (institutional representation) quantitatively analyses affiliations of CEP papers from 2018 alongside papers from comparator conferences over the same period. Finally RQ3 uses a qualitative analysis of responses to a questionnaire sent to CEP authors from 2018 onwards. So over the three research questions we sample the same papers as far as is possible, but not entirely. The papers core to our analysis are those presented at CEP: 20 from 2018 (unpublished) and 83 from 2019 onwards, published with ACM.[2]

3.1 RQ1 Methodology: Paper classification

We looked at the titles, abstracts and full pdfs (2019 onwards) of CEP papers in a deductive qualitative analysis [11], following Simon's revised classification system [35]: the categories are pre-defined, allowing comparison with other meta-analytic studies. The "Context" and "Theme" dimensions of the system are designed to be dynamic, allowing coders to add new codes to fit the situation.

We divided the papers in the CEP corpus between the five authors. Each paper was independently coded by two people, allowing coders an "other" category for both Theme and Context. Every paper and its coding was then reviewed by a single pair of authors (hereafter called the *reviewing team*). We describe our process of coding, the application of the system to the CEP corpus, and make some observations concerning the system itself.

Our paper classification results are presented alongside previously reported results for comparator conferences, with data sources identified in Table 1. This is convenience sampling: we use the data that are available for ICER [39], ITiCSE [40] and Koli [33]. The publication years considered by these studies are different to the range of CEP, but they often include the formative years of the relevant conferences, making them comparable with our analysis of CEP papers.

There have been no comparable classifications of UKICER papers, so it is excluded from RQ1. The classification of papers at ITiCSE [40] separates the working group reports, but there is less detail provided on the results for ITiCSE WG papers so they are also excluded here.

[2]Source data and code for the analyses used in the development of this paper are available at https://github.com/stevenaeola/CEP_Retro

Ref	PubYear	Conference	Years	Num Papers	Analysis	Restriction
[42]	2004	SICGSE TS	1984-2003	444	Bespoke	CS1/CS2
[32]	2007	ACE, NACCQ	2004-2007	129 (ACE) 46 (NACCQ)	Simon	
[33]*	2007	Koli	2001-2006	102	Simon	
[39]*	2008	ICER	2005-2007	43	Simon	
[29]	2009	ICER, SIGCSE TS, ITiCSE, ACE, Koli, NACCQ	2005-2008	164	Simon	programming research excluding practice
[34]	2009	ACE	1996-2008	328	Simon,bibliometric	
[20]	2010	ICER	2005-2009	67	Bespoke	
[23]	2010	ICER	2005-2009	72	Bespoke	
[37]	2016	ICER	2005-2015	187	bibliometric	
[36]	2016	Koli	2001-2015	361	bibliometric	
[24]	2018	ITiCSE WG	1996-2016	103	bibliometric	
[40]*	2020	ITiCSE	1996-2019	1295 (regular) 129 (WG)	Simon	
[38]	2020	ACE	1996-2020	541	Simon,bibliometric	
[27]	2020	ICER, ITiCSE	2005-2019	1274	keywords	
[44]	2021	SIGCSE TS, ICER, ITiCSE	1970-2017	4548	bibliometric	
[1]	2022	ACE	1996-2020	541	keywords,bibliometric	
[2]	2023	any	1970-2021	16863	keywords,bibliometric	
[4]	2023	any	1970-2021	1301	keywords,bibliometric	UK

Table 1: Previous Reviews/Classifications of Computing Education Papers, ordered by publication year. Starred papers are used as source data for comparison

To evaluate the consistency of coding across multiple coders, an inter-rater reliability (IRR) analysis was carried out using Cohen's kappa κ on each of the four different dimensions, prior to the work of the reviewing team, and excluding any use of the "other" category. The figures given in Table 2 may be interpreted as "fair" (0.21-0.4) and "substantial" (0.61-0.8) agreement. Cohen's Kappa is known to give a low score on high agreement with an imbalanced data set [25], as is obvious in the "Nature" dimension, so we have reported percentage agreement as well.

	All years		2018 Only		2018 Excluded	
	κ	%Agree	κ	%Agree	κ	%Agree
Nature	0.249	72.8	0.191	45.0	0.273	79.5
Scope	0.280	52.4	-0.046	20.0	0.389	68.2
Context	0.637	69.1	0.182	26.3	0.785	82.3
Theme	0.327	38.6	0.282	33.3	0.329	49.0

Table 2: Inter-rater reliability on our coding of CEP papers on different dimensions of Simon's system [35]

3.1.1 Coding Reconciliation. Whilst there was reasonable consistency between coders, there were also differences.

- We found some *coder-confusion* errors where Context was mistaken for Theme (and vice versa). The low Theme-level reliability we found appears to stem less from semantic ambiguity than category overlap. Simon himself noted this tendency in a previous iteration of the system [35] and he renamed "Topic" to "Theme" as a fix. It seems that the potential for confusion may run deeper than that. The reviewing team corrected these errors.
- We found that some pairs made *coder-consistent* errors, that is, for example, if one picked *assessment tool*, uniformly the other picked *teaching/learning tool*. In these cases we chose to keep the one the reviewing team thought was the best fit.
- Sometimes pairs chose different categories which were a *distinction without a difference*, for example "subject" or "program/department" for postgraduate projects. In these cases, the reviewing team chose one.
- Where two coders picked different Themes, and the reviewing team considered both to be appropriate, we marked "keep both" because we wanted the advantage of greater descriptive power. Our primary goal in using this system was not to sort the CEP paper into boxes, like apples and oranges, but to characterise their content.
- Sometimes coders used the "other" category to suggest something new. On other occasions, it was used to comment, perhaps "it could be this or that" or "I could use programming, but feel that would be misleading" Where we found these sorts of comment, and one of the suggestions included that which was picked by the other coder, the reviewing team changed it to agree.

- Finally, the reviewing team considered the "teaching/learning theories & models" category to be inapplicable to CEP papers, in that whilst they might use theories and models they do not contribute to their creation. Where a coder had chosen this option, the reviewing team universally changed it to "teaching/learning techniques".

The outcome of the reviewing team was that each paper had the same coding for each dimension, except for the cases where "keep both" categories was applied.

Because we only had titles and abstracts for CEP 2018 papers it was much more difficult to come to conclusions about classification, even after discussion by the review team. This is borne out by the inter-rater reliability scores, particularly for Scope, so the 2018 papers were dropped from the results of RQ1. The remaining years (2019-2025) are reported in Section 4.1, alongside classification data from previous studies of comparator conferences.

3.2 RQ2 Methodology: Institutional Representation

We explored UK institutional representation by extracting affiliation details through the SCOPUS API, for CEP and comparator conferences, only including papers published since 2018.

Paper affiliations are weighted to authors' institutions based on the total numbers of authors involved in a paper (known as *complete-normalised* [21]), whether or not the authors were from UK institutions (although non-UK affiliations were then removed). We refer to this combination of conference, institution and weighting as the *weighted paper affiliation.*

We further refine the ITiCSE papers to identify separately the Working Group papers (like [40]), as these may have different characteristics to regular ITiCSE papers. The separation is based on publication date, as ITiCSE working group reports are published later than papers at the main conference. We also remove the corresponding ITiCSE Working Group methodology papers presented at the main conference, as this would essentially lead to double-counting.

To compare the types of institutions, we can join the affiliation data with data from the Complete University Guide (CUG) 2025 Subject table for Computer Science [10]. Their definition of Computer Science is fairly broad, as it includes but is not limited to: Computer Networks; Software Engineering; Machine Learning; Computer Games; Business Computing. For each institution which offers undergraduate programmes in Computer Science, data is presented from a range of sources on Average Entry Standards, Student Satisfaction, Research Quality and Graduate Prospects, for the programmes included. It also provides an 'Overall Score' which combines these measures to form the basis of a league table. Weighted paper affiliation data was matched with CUG data based on the name of the institution, with alternative names provided where missing data was identified e.g. 'University of NNN' and 'NNN University'. Similarly, we join the affiliation data with student enrolment numbers from the UK Higher Education Statistics Agency (HESA) [19] as another basis for comparison, using the subject coding "121 IT, systems sciences & computer software engineering".

Not all of the UK institutions with identified publication affiliations have available data from HESA or CUG — see Section 4.2 for a discussion of these and presentation of the results relating to RQ2.

3.3 RQ3 Methodology: Perceived Value of CEP

We designed an on-line qualitative survey to explore whether attendees had implemented or planned to implement any practices they learned at CEP to their teaching, and whether they made any connections within the community as a consequence of the conference. The survey was developed by one author and reviewed by the co-authors for face and content validity before approval by the Strathclyde Computer and Information Sciences ethics committee. Email addresses of 166 unique CEP authors were collected from publications, and the survey was distributed to them with instructions to respond only if they had attended CEP.

The only demographic data requested was the participant's current department and institution. The first main question was 'Have you used any ideas, tools, or practices you encountered at a previous CEP conference?", noting that participants should not report practices they were involved in creating. Possible responses were; "yes", "no", and "not yet but I plan to".

The second question explored the origin of the idea, tool or practice, asking "What was the origin of the idea e.g. paper, workshop, discussion. If you can recall specifics, please provide them e.g. a reference to the specific paper, if not, please complete to the best of your recollection." Participants were encouraged to answer this for as many examples as they were able to, with a free-text response.

The final question explored cross institution engagement as a result of CEP by asking participants to note whether they had worked with colleagues at other institutions as a result of attendance at CEP, again with a free-text response.

4 Results and Analysis

4.1 RQ1: Paper classification

4.1.1 Nature. The breakdown of the Nature dimension is shown in Figure 2. CEP papers fall almost exclusively in the Report category, as would be expected of a practice-focussed conference, while other conferences have proportionally fewer Reports.

4.1.2 Scope. Figure 2 also shows the "Scope" dimension. For all conferences apart from ICER, the majority of the papers are either "Subject" or "Program/department", i.e. single-institutional Scope. ICER has the largest number of multi-institutional Scope papers — but notably ITiCSE Working Groups, which are designed to be multi-institutional, are not included here.

4.1.3 Context. In Simon's system both Context and Theme are designed to be dynamic, with extra categories added if necessary. This complicates direct comparisons between studies, which is already difficult (statistically and visually) because the different categories are imbalanced, some much more common than others. It is no surprise, then, that many of the papers reporting on Simon's system do not include a full summary of these data, which is the final nail in the coffin for comparison between studies. We follow earlier studies in highlighting the most frequent. The Contexts representing more than 5% of papers at CEP were programming (28%), broad-based (19%), school/outreach (8%), security (8%), groupwork (8%) and software engineering (7%). These are generally in line with the comparator conferences except that the spread of Contexts is generally broader, the proportion of programming papers is lower (although still the highest) and the number of security papers is uniquely high.

In the analysis we also found some Contexts previously unidentified: service teaching; cloud computing; and data science. We broadened the Scope of "schools/outreach" to include outreach in other non-university settings e.g. with employers.

4.1.4 Theme. The issues of reporting that affect Context also have an impact on Nature, so we report in the same way. The Themes with at least 5% representation at CEP were teaching/learning techniques (24%), curriculum (10%), assessment techniques (10%), ability/aptitude/understanding (8%) and assessment tools (7%). Again, these are in line with comparator conferences, but with a slightly broader spread, and less emphasis on "teaching/learning tools", and "teaching/ learning theories & models". New Themes that we found were: generative AI; peer learning/ evaluation/ feedback; and accreditation. We also broadened the Scope of "gender issues" to "inclusion" more generally.

4.2 RQ2: Institutional representation

Of the 103 CEP papers we analysed we found 47 distinct affiliated UK institutions. There are three UK universities that have affiliations on papers under study but are not listed in CUG. Two have very few students and total weighted paper affiliations of 0.5 or less. The third is the Open University (OU), which is very distinctive, being focussed on distance learning and having the largest number of students of any UK university, who predominantly study part-time and mostly do not start straight from school. OU does not accept entries through the standard Universities and Colleges Admissions Service (UCAS) so these data are not available, and in any case are not comparable due to the typical non-standard entry routes. The OU has produced a substantial number of computing education papers in the period of study, putting it fourth in the list of weighted paper affiliations. The universities with more than 5% of the total weighted affiliations at CEP are Glasgow (12%), Strathclyde (7%), Swansea (7%), Open (7%), Newcastle (6%), Durham (6%) and Coventry (5%).

A further 207 papers were identified at comparator conferences: Figure 3 visualises the number and weight of affiliations and the average entry standards of the affiliated computing department. The distribution of average entry standards for all CS departments is included at the bottom for comparison. The distribution of average entry standards is notably not normal (as confirmed by a Shapiro-Wilk test for normality $p = 0.000$), with a definite skew to the left of centre, a bulge from 110-140, and a long right hand tail. If the affiliations of papers were representative of all institutions then the density of the affiliations would match density of all CS departments. We can see that the CEP papers are skewed to the left of the distribution, and hence closer to the distribution for all institutions, but it is difficult to assess by eye whether this is more or less representative than the other conferences. We also note that, as expected, the weights of affiliations for ITiCSE working groups are generally a lot smaller, reflecting the large and diverse set of authors who typically contribute to each working group.

A standard statistical analysis to explore representativeness is variability of the mean [41], where the distributions of the means of various features of the population are compared between the whole population and the sample. For example, in a human population you might find the average height and income of the whole population and compare these to the average height and income of the sample. The closer the means of the sample and the whole population, the more likely the sample is representative. To compare how similar two groups are under different measures, meta-analyses usually calculate the *Standardized Mean Difference* [16] (also known as Cohen's d) by subtracting the whole population mean from the sample mean and dividing by the population standard deviation. We want to compare a range of samples (i.e. conferences) and compare them over different measures (we chose entry standards and student numbers as the most likely to affect teaching methods) with the whole population of computer science departments in the UK. For each conference we find the associated weighted paper affiliations and calculate the average of corresponding institutional measures, weighted according to author contribution. These weighted means of average entry standards and undergraduate student numbers are shown in Table 3, along with the corresponding standardized mean differences (SMD) with the whole population of CS departments, as listed in the CUG.

From Table 3 we can see that all of the conferences on average over-represent departments with higher average entry standards and larger numbers of students, making all of the SMDs positive. Weighted affiliations from CEP are the closest to being representative with respect to entry standards, with a SMD of 0.351, which would be interpreted as a small to medium (0.2-0.5) effect size if it were the result of an intervention rather than an unrepresentative sample. UKICER, ITiCSE and ITiCSE WGs have a larger SMD, very close to medium effect size, but Koli and particularly ICER have a larger SMD, which would correspond to a large effect size ($SMD > 0.8$). Looking at the representation of student numbers, affiliations from Koli are very close to the whole population average, with CEP further away and ICER the furthest, but the size of the variation is overall much smaller than that of average entry standards, with the SMD for CEP still corresponding to a small effect size.

Overall we see that, while all of the conferences have issues over representation, the effects are small (Koli) to medium (ICER) for student numbers but range from small/medium (CEP) to large/very large (Koli and ICER) for average entry standards.

4.3 RQ3: Perceived Value

A total of 15 responses were received from a possible total of 166. Of these, 10 indicated they were either already adopting practices discussed at the conference or were planning to. Whilst limited conclusions can be drawn due to the limited responses, this provides some evidence of teaching practice being shared and embedded across contexts as a result of the conference.

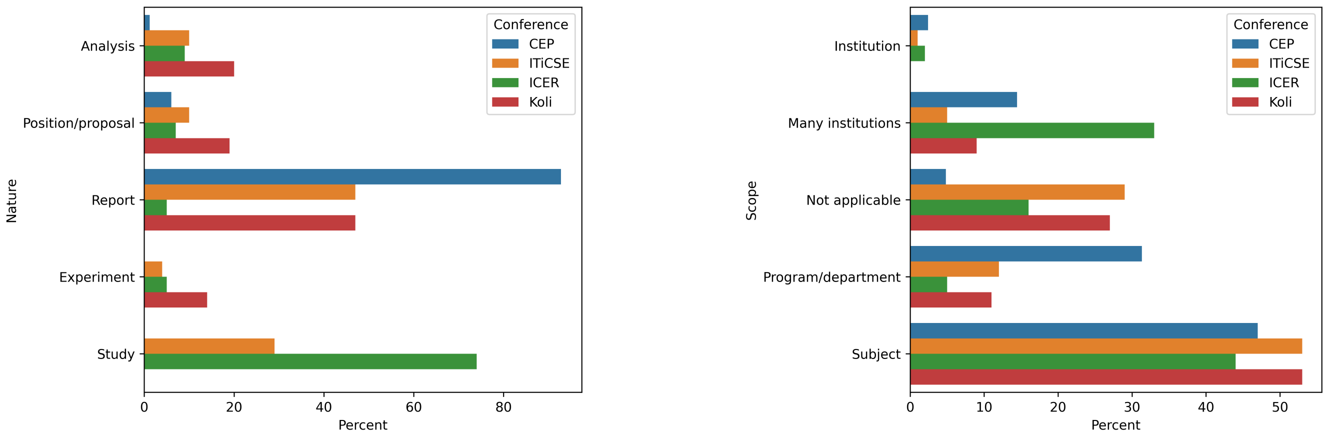

Figure 2: Nature and Scope of papers from CEP (our study) and other conferences (previous studies) based on Simon's system

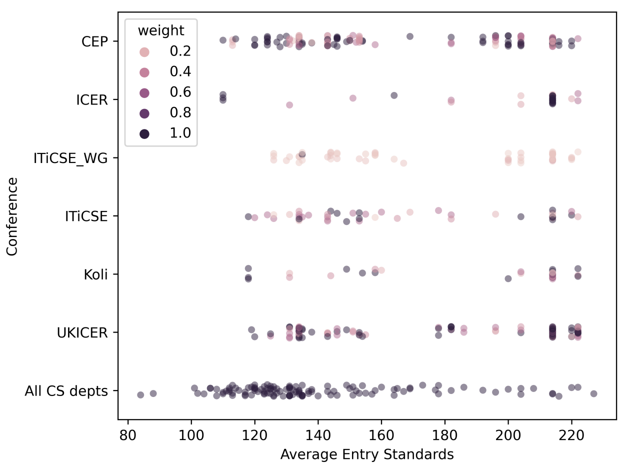

Figure 3: Distribution of average entry standards (UCAS) for UK computer science departments with weighted paper affiliations, by conference, since 2018. Bottom line shows distribution of average entry standards for all CS departments

	Entry Standards		Student Numbers	
Population	Mean	SMD	Mean	SMD
All CS depts	140	-	1060	-
CEP	151	0.351	1240	0.293
ICER	180	1.32	1370	0.495
ITiCSE_WG	155	0.494	1100	0.069
ITiCSE	154	0.454	1150	0.155
Koli	169	0.959	1090	0.049
UKICER	158	0.577	1180	0.201

Table 3: Means and Standardized Mean Differences (SMD) for average entry standards and undergraduate student numbers apportioned by weighted conference affiliations.

Nine respondents had made connections with colleagues as a result of the conference, again evidencing value for most respondents as they have increased their network in the teaching practice community.

An inductive thematic analysis on the free-text responses was completed which followed a reflexive thematic analysis approach [7] by one author. From this, the following themes were identified. These themes should be interpreted with caution due to the limited number and representation of the responses received.

4.3.1 Re-use of Presented Practice. Several participants were able to highlight specific practices presented in the form of paper presentations which resulted in contextualising the practice for their own institution. This included common group work strategies, programming pedagogy, and peer assessment.

One participant noted that 'I recall this paper in 2020 caused me to rethink how I taught Python in my class, with a lot more emphasis on the mutable vs immutable type behaviours" demonstrating how exposure to examples of teaching practice can lead to rethinking instruction in a different context.

4.3.2 Value of sharing small practice. The value of sharing small examples of teaching practice was evidenced most clearly in this comment from a participant "I learn a lot from conversations rather than just the papers being presented to be honest."

This highlights a well-recognised but often under-documented aspect of teaching practice, that the informal, peer-to-peer exchanges support professional development of educators. Unlike formal research outputs, these conversational insights are difficult to attribute, cite, or trace, yet they often have a significant and immediate value for teaching practice.

4.3.3 Contextual Importance. Several responses indicated the importance of adapting practice to their own context (whether presented or discussed in side conversations) and, at times, the difficulty of this.

One participant noted that a particular practice would "work well especially with a certain demographic in our institution that have been identified as consistently under performing." This shows that there is a perceived importance of understanding the institutional context in adopting practice.

Other responses were more general, such as the participant who noted "I have extended or contextualised an idea about how to engage students...". Additionally, there was also reference to the possible challenges in adopting practice into their context, e.g. "...it's been on my mind but I'm not sure how to transfer from the written report of the experience into practice." This highlights a potential gap between the dissemination of practice and implementation in context, as discussed in Section 1.1.

5 Discussion

When we started exploring RQ1 we expected Simon's framework to be impoverished in the Nature and Scope dimensions, and this was indeed the case. However, we failed to anticipate the extent to which these fixed categories provided comparability across contexts. We had hoped for richer and more nuanced results under Context and Theme but, in reality, the Context and Theme data proved hard to compare, and largely agreed with comparator conferences where data was available.

The results on Nature in Figure 2 give us a helpful characterisation of how much focus each conference has on practice (characterised by "Report") and research (characterised by "Analysis", "Study" and "Experiment"). Ranking the conferences on this basis from most practice-focussed to least practice-focussed we have CEP, (ITiCSE and Koli) and ICER. The classification data for Koli [33] is relatively old compared with the data for ITiCSE [40], and there has been a trend over time towards research-focussed papers. Looking at the current Calls for Papers for Koli and ITiCSE it is notable that ITiCSE includes a whole track for "Experience reports and tools", whereas practice papers at Koli are only in scope for "Discussion papers", which can also be research-focussed. Based on these two points, we can break the tie between Koli and ITiCSE to suggest that ITiCSE is more practice-focussed than Koli, giving us the practice-focussed ranking: CEP; ITiCSE; Koli; ICER.

Combining these qualitative results with the quantitative results for RQ2 makes for an illuminating comparison. There is a fairly narrow range and overall small size of standardized mean difference when looking at student numbers. This reflects the underlying data (not shown here, for space reasons) that the universities with the highest and lowest entry standards have relatively low student numbers, and the largest student numbers are at universities that are near to the median of entry standards. This is good news for studies or practices addressing large class sizes, but when looking at average entry standards there is a lot of difference between the conferences, and some very large differences with the average over all CS departments. Ranking the conferences from most representative to least representative on average entry standards we have: CEP; ITiCSE; ITiCSE WG; UKICER; Koli; ICER. It is notable that this ordering is, where applicable, the same as the practice-focussed ranking. UKICER (UK and Ireland Computing Education Research) is, by description, less practice-focussed than ITiCSE — it has no practice or experience track — so the fact that ITiCSE is more representative than UKICER follows the trend that more practice-focussed conferences tend to be more representative with respect to average entry standards.

But why should we care about representativeness of publishing institutions? From a research perspective, where local students are studied through convenience sampling, the data set is biased/unrepresentative. The results for RQ1 show that for all conferences the majority of papers are scoped to a single subject, program or department. Any attempted generalisation to the student population as a whole is therefore problematic. While most research papers will identify this as a threat to validity, we can see how real this problem is — and so some interventions or practices that appear to work well may in fact only work for students who have tended to perform well. In RQ3 we found evidence that particular demographics within cohorts could under-perform, so the mix of students is seen as important here too.

Coming from the practice angle, as well as having biased results on practices that are proposed, there is a challenge that new practices which are developed in contexts with lower average entry standards are not being reported as much. If different practices are needed for different students — particularly those that have not thrived in school education for whatever reason — then we may not hear about them in the literature. This could include marginalised students and/or those with specific learning needs or difficulties. Anybody coming to the corpus of computing education papers as a whole with questions about how best to support learning and teaching may not find practices that work in their context, or data that generalises to their students.

Facilitating the transplantation of practice in computing education is important both for students and for staff. Times change, and they change particularly quickly in computing, so keeping content and pedagogy relevant and appropriate for students is a challenge. Issues that have faced the computing education community include: volatility in student numbers; pandemic; rapid shifts in political/cultural climate; challenges of retaining reliable of assessment in the face of generative AI; adapting our curriculum so that it underpins the future needs of students and society more broadly. The initial evidence from RQ3 is that participants were able to use the contextual information provided through CEP papers and discussions to judge whether and how to adopt and adapt presented practices, although further work is required to address this research question fully.

6 Threats to validity

RQ1
- Qualitative analysis always has an element of subjectivity, so more work could be done to achieve classification agreement, by using more coders or adopting a Delphi approach [39]
- Some of the papers relating to comparator conferences are quite old and limited in range of years and total papers
- Researchers completing paper classifications may have had different interpretations of the classification system

RQ2
- Relatively few papers were included from ICER and Koli, because of low publication rates. More years of affiliation data could have been included, but that would weaken the comparison with the CEP papers
- Affiliation contributions may be unequal, in particular the leaders of ITiCSE WGs are much more influential than

other members in identifying the research questions of those studies
- We assume that institutional affiliation corresponds with investment in the innovations and/or representation in any data used for evaluation or detailed study. In particular, papers relating to schools are included, weakening the relationship between the author institution and average entry standard measures. Papers could be screened for school context, either by hand or using a LLM

RQ3
- The number and diversity of responses was limited in number
- Not all participants could be reached, and not all of those reached did respond, so some issues may have been missed or misrepresented
- There is potential for selection bias — people who did not find the conference useful might be less likely to respond

Synthesis
- Samples are not always aligned between the different methodological approaches, e.g. use of other study data for RQ1 which did not align with date selection of papers for RQ2
- Only UK affiliations were used, so overall inferences about the practice vs research papers may not generalise

7 Conclusions

Many other studies have taken one conference as focus, sometimes looking at trends over time. We have looked at multiple conferences and trends over the type of conference, relating the type of papers published at the conference to the type of institutions publishing there, from a UK perspective. The combination of qualitative and quantitative components form a consistent picture that practice-based conferences tend to have a more representative set of institutions presenting at them. This diversity has the potential to produce a broader and more applicable set of innovations in practice, with a less biased evidence base. The contextualisation and practical detail offered in the practice papers at CEP supports adoption at other institutions, along with more recognition for those staff, often on teaching-only contracts, who have developed the practices.

Computing Education Research has its place and is widely reported. But CER papers rarely pay attention to specific teaching practice and almost universally ignore factors that impact transplantation: they rarely have an effect on teachers or what they teach. At the same time, the importance of Computing Education Practice is overlooked and papers that report it are derided — "Marco Polo" is meant as a disparaging term. To be taken seriously, it needs to be considered as a serious endeavour. We admire the careful recording and ecological contextualisation in the writing and drawing of Maria Sibylla Merian, and offer her as a more positive model for the description of practice. There is more work to be done to aid colleagues in sharing and adopting computing education practice effectively.

In summary, we believe:

- Practice papers in computing education should not be discounted but valued, with clear opportunities for publication: either at practice-specific venues, or in specific tracks at existing conferences.
- Specified formats for practice-focussed papers should require enough contextual detail to allow for transplantation, enabling practitioners to decide whether the approach is applicable within their own environment. This is in contrast to generalisation through abstraction, which is the aim of much research.
- Convenience sampling of individual courses at research intensive universities leads to unrepresentative results across the discipline. Researchers should give careful consideration to the characteristics of the classes that are represented in studies, particularly through the use of department-level metrics such as average entry standards and class sizes.

7.1 Further Work

As part of our future work, we plan to review promotion criteria for teaching focused staff across UK higher education institutions to identify common requirements and characterise the types of evidence typically included, with the aim of mapping these onto computing-specific contexts. Building on this, we will run a series of workshops for computing academics and those supporting promotion, helping participants document examples of their teaching practice aligned to their institution's criteria using the Computing Education Practice (CEP) format. These documented examples aim to support promotion cases, encourage submissions to CEP for peer review and publication, and create a collection of exemplars demonstrating how computing education practice can meet general T&L criteria while maintaining a strong disciplinary focus.

More could be done with bibliometric and scientometric studies of practice papers. Whilst the SCOPUS API allowed us to extract some details, including the number of times a paper was cited, it is not able to give sufficient data for us to analyse this in greater detail. Future work could involve manually examining these citations with greater granularity, for example, identifying self-citations, citations within CEP, and citations in other contexts. This will help us better understand the audience and reach of the conference. Similarly we could explore the publication histories of people who write computing education practice papers, to identify how they start and how they develop, both as individual authors and within their collaboration networks.

Initial qualitative results on the contribution of the conference are limited in scope and generalisability due to the small and potentially unrepresentative sample. A broader qualitative study of how published work contributes to practice within computing education could provide richer and more robust results.

Lastly, although we find our results compelling within the UK context, it would be good to explore whether our findings apply outside the UK, with its distinctive marketisation of Higher Education and staffing structures.

References

[1] Alireza Ahadi and Luke Mathieson. 2022. A Bibliometrics Analysis of Australasian Computing Education Conference Proceedings. In *Proceedings of the 24th Australasian Computing Education Conference (ACE '22)*. ACM, New York, NY, USA, 1–9. doi:10.1145/3511861.3511862

[2] Mikko Apiola, Sonsoles López-Pernas, and Mohammed Saqr (Eds.). 2023. *Past, Present and Future of Computing Education Research: A Global Perspective*. Springer. doi:10.1007/978-3-031-25336-2

[3] Mikko Apiola, Mohammed Saqr, and Sonsoles López-Pernas. 2023. The Evolving Themes of Computing Education Research: Trends, Topic Models, and Emerging

Research. In *Past, Present and Future of Computing Education Research : A Global Perspective*, Mikko Apiola, Sonsoles López-Pernas, and Mohammed Saqr (Eds.). Springer, 151–169. doi:10.1007/978-3-031-25336-2_8

[4] Brett A. Becker, Steven Bradley, Joseph Maguire, Michaela Black, Tom Crick, Mohammed Saqr, Sue Sentance, and Keith Quille. 2023. Computing Education Research in the UK & Ireland. In *Past, Present and Future of Computing Education Research : A Global Perspective*, Mikko Apiola, Sonsoles López-Pernas, and Mohammed Saqr (Eds.). Springer, 421–479. doi:10.1007/978-3-031-25336-2_19

[5] Amanda Berry, Patricia Friedrichsen, and John Loughran (Eds.). 2015. *Re-examining Pedagogical Content Knowledge in Science Education*. Routledge, New York. doi:10.4324/9781315735665

[6] Ron Brandt. 1992. On Research on Teaching: A Conversation with Lee Shulman. *Educational Leadership* 49, 7 (1992), 14–19. https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_199204_brandt2.pdf

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[8] John Seely Brown and Paul Duguid. 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science* 2, 1 (Feb. 1991), 40–57. doi:10.1287/orsc.2.1.40

[9] Robert V. Bullough. 2001. Pedagogical content knowledge circa 1907 and 1987: a study in the history of an idea. *Teaching and Teacher Education* 17, 6 (Aug. 2001), 655–666. doi:10.1016/S0742-051X(01)00022-1

[10] CUG. 2025. Computer Science Subject League Table 2025. https://www.thecompleteuniversityguide.co.uk/league-tables/rankings/computer-science

[11] Stephen T. Fife and Jacob D. Gossner. 2024. Deductive Qualitative Analysis: Evaluating, Expanding, and Refining Theory. *International Journal of Qualitative Methods* 23 (Nov. 2024). doi:10.1177/16094069241244856 Publisher: SAGE Publications Inc.

[12] Sally Fincher and Sebastian Dziallas. 2015. Then and Now: Past Experience Echoed in University Computing Teachers' Current Practice. Dublin, Ireland. http://www.rees2015.org/

[13] Sally Fincher, Marian Petre, and Martyn Clark (Eds.). 2001. *Computer Science Project Work*. Springer, London. doi:10.1007/978-1-4471-3700-9

[14] Sally A. Fincher, Yifat Ben-David Kolikant, and Katrina Falkner. 2019. Teacher Learning and Professional Development. In *The Cambridge Handbook of Computing Education Research*, Anthony V. Robins and Sally A. Fincher (Eds.). Cambridge University Press, Cambridge, 727–748. doi:10.1017/9781108654555.026

[15] Janet Finlay. 2013. *Representing Teaching Practice A book of bundles*. University of Kent Press. https://www.cs.kent.ac.uk/national/EPCOS/ShareProjectBookOfBundles.pdf

[16] Daniel Gallardo-Gómez, Rachel Richardson, and Kerry Dwan. 2024. Standardized mean differences in meta-analysis: A tutorial. *Cochrane Evidence Synthesis and Methods* 2, 3 (2024), e12047. doi:10.1002/cesm.12047

[17] Julie Gess-Newsome and Norman G. Lederman (Eds.). 2002. *Examining Pedagogical Content Knowledge*. Science & Technology Education Library, Vol. 6. Kluwer Academic Publishers, Dordrecht. doi:10.1007/0-306-47217-1

[18] S. Gorard, B. H. See, and N. Siddiqui. 2017. *The Trials of Evidence-based Education: The Promises, Opportunities and Problems of Trials in Education*. Routledge.

[19] HESA. 2025. Table 37 - HE student enrolment FTE by HE provider and cost centre 2014/15 to 2023/24 | HESA. https://www.hesa.ac.uk/data-and-analysis/students/table-37

[20] Päivi Kinnunen, Veijo Meisalo, and Lauri Malmi. 2010. Have we missed something? identifying missing types of research in computing education. In *Proceedings of the Sixth international workshop on Computing education research (ICER '10)*. ACM, New York, NY, USA, 13–22. doi:10.1145/1839594.1839598

[21] Peder Olesen Larsen. 2008. The state of the art in publication counting. *Scientometrics* 77, 2 (Nov. 2008), 235–251. doi:10.1007/s11192-007-1991-6

[22] Paul MacLellan. 2016. Why don't teachers use education research in teaching? https://edu.rsc.org/analysis/why-dont-teachers-use-education-research-in-teaching/2010170.article

[23] Lauri Malmi, Judy Sheard, Simon, Roman Bednarik, Juha Helminen, Ari Korhonen, Niko Myller, Juha Sorva, and Ahmad Taherkhani. 2010. Characterizing research in computing education: a preliminary analysis of the literature. In *Proceedings of the Sixth international workshop on Computing education research (ICER '10)*. ACM, New York, NY, USA, 3–12. doi:10.1145/1839594.1839597

[24] Robert McCartney and Kate Sanders. 2018. ITiCSE working groups and collaboration in the computing education community. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018)*. ACM, New York, NY, USA, 332–337. doi:10.1145/3197091.3197143

[25] Mary L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* 22, 3 (Oct. 2012), 276–282.

[26] NHM. [n. d.]. Maria Sibylla Merian: metamorphosis unmasked by art and science | Natural History Museum. https://www.nhm.ac.uk/discover/maria-sibylla-merian-metamorphosis-art-and-science.html

[27] Zacharoula Papamitsiou, Michail Giannakos, Simon, and Andrew Luxton-Reilly. 2020. Computing Education Research Landscape through an Analysis of Keywords. In *Proceedings of the 2020 ACM Conference on International Computing Education Research (ICER '20)*. ACM, New York, NY, USA, 102–112. doi:10.1145/3372782.3406276

[28] Carsten Schulte, Sue Sentance, Sören Sparmann, Rukiye Altin, Mor Friebroon-Yesharim, Martina Landman, Michael T. Rücker, Spruha Satavlekar, Angela Siegel, Matti Tedre, Laura Tubino, Henriikka Vartiainen, J. Ángel VelÁzquez-Iturbide, Jane Waite, and Zihan Wu. 2025. What We Talk About When We Talk About K-12 Computing Education. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education (ITiCSE 2024)*. ACM, New York, NY, USA, 226–257. doi:10.1145/3689187.3709612

[29] Judy Sheard, S. Simon, Margaret Hamilton, and Jan Lönnberg. 2009. Analysis of research into the teaching and learning of programming. In *Proceedings of the fifth international workshop on Computing education research workshop (ICER '09)*. ACM, New York, NY, USA, 93–104. doi:10.1145/1584322.1584334

[30] Lee S. Shulman. 1986. Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher* 15, 2 (Feb. 1986), 4–14. doi:10.3102/0013189X015002004

[31] SIGCSE. 2025. SIGCSE TS 2025. https://sigcse2025.sigcse.org/

[32] Simon. 2007. A Classification of Recent Australasian Computing Education Publications. *Computer Science Education* 17, 3 (Sept. 2007), 155–169. doi:10.1080/08993400701538021

[33] Simon. 2007. Koli Calling comes of age: an analysis. In *Proceedings of the Seventh Baltic Sea Conference on Computing Education Research - Volume 88 (Koli Calling '07)*. Australian Computer Society, Inc., AUS, 119–126.

[34] Simon. 2009. Ten years of the Australasian Computing Education Conference. In *Proceedings of the Eleventh Australasian Conference on Computing Education - Volume 95 (ACE '09)*. Australian Computer Society, Inc., AUS, 157–164.

[35] Simon. 2015. *Emergence of computing education as a research discipline*. Ph. D. Dissertation. Aalto University.

[36] Simon. 2016. The koli calling community. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research (Koli Calling '16)*. ACM, New York, NY, USA, 101–109. doi:10.1145/2999541.2999562

[37] Simon. 2016. A Picture of the Growing ICER Community. In *Proceedings of the 2016 ACM Conference on International Computing Education Research (ICER '16)*. ACM, New York, NY, USA, 153–159. doi:10.1145/2960310.2960323

[38] Simon. 2020. Twenty-Two Years of ACE. In *Proceedings of the Twenty-Second Australasian Computing Education Conference (ACE'20)*. ACM, New York, NY, USA, 203–210. doi:10.1145/3373165.3373188

[39] Simon, Angela Carbone, Michael de Raadt, Raymond Lister, Margaret Hamilton, and Judy Sheard. 2008. Classifying computing education papers: process and results. In *Proceedings of the Fourth international Workshop on Computing Education Research (ICER '08)*. ACM, New York, NY, USA, 161–172. doi:10.1145/1404520.1404536

[40] Simon and Judy Sheard. 2020. Twenty-Four Years of ITiCSE Papers. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '20)*. ACM, New York, NY, USA, 5–11. doi:10.1145/3341525.3387407

[41] Valmi D. Sousa, Jaclene A. Zauszniewski, and Carol M. Musil. 2004. How to determine whether a convenience sample represents the population 1. *Applied Nursing Research* 17, 2 (May 2004), 130–133. doi:10.1016/j.apnr.2004.03.003

[42] David W. Valentine. 2004. CS educational research: a meta-analysis of SIGCSE technical symposium proceedings. In *Proceedings of the 35th SIGCSE technical symposium on Computer science education (SIGCSE '04)*. ACM, New York, NY, USA, 255–259. doi:10.1145/971300.971391

[43] Daniel T. Willingham and David B. Daniel. 2021. Making Education Research Relevant. https://www.educationnext.org/making-education-research-relevant-how-researchers-can-give-teachers-more-choices/

[44] James Zhang, Andrew Luxton-Reilly, Paul Denny, and Jacqueline Whalley. 2021. Scientific Collaboration Network Analysis for Computing Education Conferences. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE '21)*. ACM, New York, NY, USA, 582–588. doi:10.1145/3430665.3456385