# Collaborative Culture: Analyzing Global Trends in Computing Education

Maíra Marques Samary*[1], Stephanie Lunn*[2], and Alan Peterfreund[3]

[1] Boston College, Chestnut Hill, MA, USA, marquemo@bc.edu,
[2] Florida International University, Miami, FL, USA, slunn002@fiu.edu
[3] SageFox Consulting Group, Amherst, MA, USA, apeterfreund@sagefoxgroup.com

**Abstract.** Computing education (CEd), or computer science education (CSEd), research has the potential to affect not only what and how we teach, but also who is taught and where. While CEd has grown as a discipline over the past two decades, many institutions still lack formal departments or programs. Given that it is a specialized and interdisciplinary area of research, we wanted to assess the values of collaboration and access. To develop a better understanding of the researchers and institutions working in CEd, we manually collected publication data from the Innovation and Technology in Computer Science Education (ITiCSE) and the ACM International Computing Education Research (ICER) conferences, and the ACM Transactions on Computing Education journal, between 2015 and 2020. Using a collective total of 1099 publications, we analyzed affiliation information about the authors and their institutions. Although we hoped to uncover a global presence and collaborative relationships demonstrating a "CSEd for all" mindset, instead we found that North America and Europe were over-represented relative to other continents. Additionally, collaborations remained a national or regional affair, for the most part. While many factors may contribute, from language barriers to financial obstacles, communication across country lines needs to improve to truly develop a more equitable international presence in the field. Through this research, we hope to raise awareness of where CEd research is being conducted and what level of collaboration occurs between institutions and countries. Moreover, we want to encourage researchers to seek alternative perspectives and to expand their collaborations to ensure CEd work truly encompasses a broader worldview.

**Keywords:** Computing education, computer science education, collaboration, bibliometrics

## 1 Introduction

Although computer science education (CSEd), or computing education (CEd), is currently of indisputable relevance to emerging topics, the field is still in its infancy and trying to find its place as a discipline [7, 11]. The projected need for graduates in computing, including computer science (CS) and information technology (IT), is on the rise, along with demand for workers and instructors [16].

* Both authors contributed equally to this research.

As such, the need for CEd will only continue to gain momentum. In the United States of America (USA), a movement arose in the late 1990s to encourage the spread of computer science to more students, and to aid in professional development for teachers as part of a campaign called "CSForAll" [5]. Meanwhile, the Eurydice Report in Europe [6] highlighted educational contributions made in individual countries and also articulated the minimum expectation to be a constituent of a country. These are just two examples of initiatives to expand CEd to a larger population.

However, to improve CEd participation in schools from kindergarten to $12^{th}$ grade (referring to the system of primary and secondary education in the United States) and at colleges/universities, it is important to consider a global effort to promote CEd for all. Before examining the factors that contribute to a global community, we need to know who is currently doing research in the field, and which countries contribute to developing the CEd knowledge base. Achieving this may take time, but as a first step towards developing that landscape we wanted to profile institutions currently publishing in the field, and the collaborations that are already in place. We also explored where research occurs, and tried to understand the values that may shape participation. We sought to explore: 1) Institutional collaborations; and 2) Whether collaborations are typically national (within a country) or global (between countries).

It is important for researchers, faculty, and students to work together in CEd, since often they face similar problems, and can help each other uncover solutions and lend support where their home institution may be unfamiliar with the discipline. Yet, if someone wants to find a partner to seek out a collaboration, or if a student wants to do research or pursue a PhD in CEd in a different place, where can he/she find information about the institutions that conduct research in this area? Towards this goal, Dr. Amy J. Ko has made strides towards collecting and presenting this kind of information[4].

Previously, Malmi [10] evaluated computing education research (CER) venues (conferences and journals), and he created a pool of them that he defined as "central forums." Using his list as a starting point, we further narrowed and selected three venues to begin our inquiry: Proceedings of the Innovation and Technology in Computer Science Education (ITiCSE), ACM International Computing Education Research (ICER), and ACM Transaction on Computing Education (TOCE). In particular, we considered work published over the last six years (2015 to 2020).

In this study, we wanted to develop a better understanding of the landscape of institutions, countries, and collaborations that exist in CER. We sought to answer the following research questions: **RQ1:** *What countries are represented in CER at ICER, ITiCSE, and TOCE in particular?* **RQ2:** *How much collaboration exists in CEd, nationally and internationally?*

The rest of this article is structured as follows: In Section 2 we discuss related works and how our approach is novel. In Section 3 we describe how data was gathered, cleaned, and analyzed. Section 4 then provides an overview of

[4] http://faculty.washington.edu/ajko/cer

the data, then analyzes participation by country before assessing international collaborations. Section 5 includes a general discussion of what we found, and we finish with threats to validity (Section 6) and conclusions (Section 7).

## 2 Related Research

Prior assessments of CER publications have often had very different goals. Valentine examined 20 years of technical symposiums from SIGCSE in 2004 [20]. The main goal was to classify works being published according to the content that the work presented. He categorized the presentations using informal labels like "John Henry" to denote a paper discussing an absurdly challenging course. Another example was his "Philosophy" label, used to describe when an author used his/her publication to take a philosophical stand on a topic.

In 2012, Miró Julià *et al.* [15] categorized and separated conferences in CS research from those in education, based on the main focus of the conference. They considered publications from 1996 to 2011. Among those evaluated, the following conferences were designated as CSEd: ICER, ITiCSE, the Australasian Computing Education (ACE) Conference, Frontiers in Education (FIE), Jornadas de Enseñanza Universitaria de la Informatica (Jenui; Spain's CSEd Conference), the Conference on Software Engineering Education and Training (CSEET), and the Technical Symposium on Computer Science Education (SIGCSE). They made a case for separating out those authors publishing for the first time (newcomers) and veterans (old-timers), and observed that in CSEd, authors are more likely to collaborate with those with whom they had a prior connection. Their goal was to clearly and uniquely identify each author, to be able to show clusters of authors collaboration and the differences in CS research and education, and to demonstrate value of of cluster analysis.

Another example of bibliometrics undertaken to examine collaborations, was a study conducted on ITiCSE working groups (WGs) by McCartney and Sanders [13]. In their study, they used data from 1996 up to 2016. Their work was built on the previous work done by Miró Julià, and considered authors' connections. Focusing only on one type of publication, the WGs, they stated that the numbers of collaboration for ITiCSE were a bit higher from all other conferences and from previous analyses conducted. This reflects the essence of the WG as a place for community research and collaboration.

In 2016, Simon conducted an analysis of ICER community [18], examining the authors that published there. His work analyzed prolific authors, sociable authors (authors with many different co-authors), Lotka's law, collaboration clusters, and authors' country of origin. However, he focused primarily on collaboration clusters, considering only the novelty of authors names appearing at ICER. Based on his country analysis, the United States had the majority of papers from ICER, representing 57% of the papers published. While these efforts provided valuable insight, this analysis was focused only on authors in the ICER community.

Then, in 2020, Simon and Sheard analyzed ITiCSE publications between 1996 and 2019 [19]. Again, their examination focused more on the content presented, although also considered authors' affiliation as a measure of their location to consider collaborations. They demonstrated that typically full paper collabora-

tions involve two or more countries, and that for WGs the majority come from two or more authors at the same institution. However, they did not fully examine, or mention, which countries in particular were involved, nor did they give additional details beyond a broad categorization as "single" versus "2 or more."

Our research presents an overview of work in the CEd over the last six years. We analyzed two conferences that are dedicated to computing education, and also a CEd journal. Ultimately, our focus was on where research comes from in terms of country and institutions, and the values that shape collaborations. A partial analysis of the data collected in this project was presented at SIGCSE 2021, which focused only on where research occurs, rather than collaborations and global representation [2]. Instead of limiting this analysis only to novel authors, we explore all contributing authors' affiliations.

## 3 Methods

### 3.1 Data Collection

ITiCSE and ICER were chosen since these represent two major global conferences in CEd, and TOCE was selected since it is a key source for CEd publications [9, 12]. Although authors may also publish in other venues, we kept our scope limited to evaluating and presenting this data. Authors and their affiliation information were gathered using these sources.

The information about conference publications (papers' name and its authors) is open-access, but it is not readily available for analysis since it is scattered on the papers' pdfs and within digital library click boxes. As such, this data had to be compiled manually since the digital library prevented the use of crawlers to gather the information. In our case, the main sources of data were ACM Digital Library ITiCSE[5], ICER[6] and TOCE[7]

We created a common spreadsheet and gathered information from each publication in the time frame of 2015 up to 2020, from these venues. However, for TOCE we only considered publications available up to July 2020, when this collection was conducted. The information assessed included: publication title, type of work accepted (in which track it was accepted), amount of authors, author(s) name(s), author(s) email(s), author(s) affiliation(s). All of the data aggregation was completed by the first two authors, who checked each other's work. We obtained a total of 1,099 publications. Hereafter, we use the term "publication" to reference all works accepted and included at ICER, ITiCSE, and TOCE. Comparatively, we when we use the term "paper," we are limiting this larger categorization solely to research papers. Since the data collected from conferences and journals are not identical, we ultimately created two types of spreadsheets.

### 3.2 Data Cleaning

Unfortunately, publications usually do not have mandatory standards for authors' affiliations, and the guidelines that are given are not always followed.

[5] https://dl.acm.org/conference/iticse/proceedings
[6] https://dl.acm.org/conference/icer/proceedings
[7] https://dl.acm.org/journal/toce

Such information is self-reported, and it is often inconsistent (e.g., department name, name of academic institution, organization, or company). Even authors belonging to the same institution often use different labels. When considering authors outside the USA, the problem is exacerbated by the fact that some authors keep the original names of their institutions (original language), while others use translations, e.g., Technische Universität München or Technical University of Munich.

In order to resolve this problem, and to more easily analyze the data, we put all the information into a SQL database (MariaDB [8]). To organize the information systematically, we created a parser (using Python) to read each line of the spreadsheet. Then, we separated out the salient facts by data type. To ensure homogeneity of the names of institutions despite fluctuations, we sought alternative sources for formal comparisons of the self-reported names. For institutions in the USA, we used a database of colleges from the Department of Education, which includes 9000+ colleges/universities. Meanwhile, for international universities, we used a JSON file that parsed the university names according to email domains. The file we utilized is maintained on GitHub for use and contribution[8].

Although these steps aided in data cleaning, several emails that authors used were not on the JSON file. Furthermore, authors sometimes used their personal emails (such as @gmail.com) that prevented us from automatically recognizing their affiliation. Another resource used to correctly classify the institution was applying a Python wrapper for Libpostal, available for public use on GitHub[9]. A few cases were unable to be processed automatically, so the parser had to break the affiliation up by institution name and department (department name or lab name or research group name).

After parsing all the data in database, the institutional table was manually analyzed. The first two authors independently checked all of the information to ensure fidelity and accuracy of information as much as possible. The cleaned institutional information was then returned to the database for analysis.

### 3.3 Data Analysis

Data analysis was conducted using SQL queries of MariaDB. Using separate searches for authors, publication types, institutions, locations, etc., we were able to identify trends in participation over time. We present the findings from our inquiry below.

## 4 Results

### 4.1 Overview of Conferences

Table 1 displays the general information regarding ITiCSE and ICER locations over the last six years. It includes the year, city, country and continent for the conferences. Between 2015 and 2020, ITiCSE and ICER have been held in 11 different countries and on 4 continents. It should be noted that continent is a designation which varies around the world, and there are often differences reported in terms of continents' names and quantity. For the purposes of this analysis, we

[8] https://github.com/hipo/university-domains-list
[9] https://github.com/openvenues/pypostal

divided the American continent. We considered North America as only including Canada and the United States, and all the other countries in America were classified as Latin America. Conferences are not immutable entities, and over

| Year | Conference | City | Country | Continent |
|---|---|---|---|---|
| 2015 | ITiCSE | Vilnius | Lithuania | Europe |
| | ICER | Omaha | USA | North America |
| 2016 | ITiCSE | Arequipa | Peru | Latin America |
| | ICER | Melbourne | Australia | Oceania |
| 2017 | ITiCSE | Bologna | Italy | Europe |
| | ICER | Tacoma | USA | North America |
| 2018 | ITiCSE | Larnaca | Cyprus | Europe |
| | ICER | Espoo | Finland | Europe |
| 2019 | ITiCSE | Aberdeen | Scotland | Europe |
| | ICER | Toronto | Canada | North America |
| 2020 | ITiCSE | Trondheim | Norway | Europe (Online) |
| | ICER | Dunedin | New Zealand | Oceania (Online) |

Table 1: Locations of conferences between 2015 and 2020

the last 6 years there were some changes made to both conferences, not only in terms of venue, but also to the tracks offered. Table 2 illustrates the changes and publication types that appeared over time at ITiCSE and ICER. For example, in 2015, ITiCSE only had papers, panels, posters, and working groups, but in 2016, "Tutorials" and "Tips, Techniques and Courseware" appeared. Then, the "Doctoral Consortium" was created in 2019. Also, we observed that over the years ICER kept the number of accepted papers in the same levels, with an average of 27. Although ITiCSE increased the amount of accepted papers over time, it went from 54 to 72, with an average of 60 papers annually.

| *ICER* | | | | | | |
|---|---|---|---|---|---|---|
| **Publication Type** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** |
| Paper | 25 | 26 | 29 | 28 | 28 | 27 |
| Doctoral Consortium | 20 | 17 | 20 | 19 | 21 | 11 |
| Tools and Technologies in Computing Education | | | | 4 | | |
| Poster | | | | | 16 | 18 |
| Special Session | | | | | 1 | |
| *ITiCSE* | | | | | | |
| **Publication Type** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** |
| Paper | 54 | 54 | 56 | 56 | 66 | 72 |
| Panel | 1 | 3 | 2 | 1 | 3 | 2 |
| Poster | 24 | 26 | 24 | 21 | 37 | 18 |
| Working Groups | 7 | 7 | 9 | 9 | 10 | 9 |
| Tips, Techniques and Courseware | | 9 | 9 | 9 | 9 | 15 |
| Tutorials | | 1 | 2 | | | |
| Doctoral Consortium | | | | | 10 | 11 |

Table 2: Types of publications per year at ICER and ITiCSE

### 4.2 Participation by Country

Analyzing the data from the three sources, there were a total of 1,099 publications. Among these, 646 publications came from ITiCSE, and 310 publications came from ICER, and 143 papers came from TOCE. Table 3 describes the amount of publications produced from authors by continent, the amount of authors that generated the publications, the amount of countries that have publications in the time frame we analyzed, and the amount of different institutions that had something published within our time-frame. The amount of publications from our three sources and the amount of publications from Table 3 are different because each publication was associated to all of its authors, so one publication could be associated with multiple continents, countries. It was evident that the majority of publications came from North America (Canada and USA), Europe, and Oceania. Asia, Latin America, and Africa also appeared, but their contributions were one magnitude smaller than the other three continents. In total, 61 different countries participated.

The top countries in terms of the (amount of publications/amount of authors) per continent were:

- **Africa:** South Africa (4/9), Ghana (2/1), Kenya (1/1), Nigeria (1/1)
- **Asia:** India (21/23), Israel (12/18), China (11/21), Japan (8/21)
- **Europe:** United Kingdom (154/174), Finland (68/64), Germany (59/90), Sweden (50/55), Ireland (39/27), Italy (31/31), Spain (29/47), Netherlands (25/29)
- **Latin America:** Peru, (21/17), Brazil (15/31), Colombia (6/10), Argentina (6/7), Chile (5/8)
- **North America:** United States (102/79), Canada (42/45)
- **Oceania:** Australia (102/79), New Zealand (42/45)

| Continent | Amount of Publications | Amount of Authors | Number of Countries | Amount of Institutions |
|---|---|---|---|---|
| Africa | 8 | 12 | 4 | 8 |
| Asia | 78 | 120 | 17 | 59 |
| Europe | 587 | 700 | 26 | 251 |
| North America | 685 | 1049 | 2 | 284 |
| Oceania | 144 | 124 | 2 | 32 |
| Latin America | 59 | 83 | 10 | 43 |
| **Grand Total** | **1561** | **2088** | **61** | **677** |

Table 3: Amount of publications, authors, and countries that published on ITiCSE, ICER and TOCE in the last six years

### 4.3 Publications by Institution

We also analyzed the amount of publications by institution. Table 4 shows the institutions that have more than 12 publications at ITiCSE, ICER, or TOCE in the last six years.

### 4.4 Conference Location Relation to Participation

Conferences change venues every year, so to determine if location affected participation, we explored the relationship between where the conference was located (continent) with the origin of the publications. Table 5 details the breakdown by continent, as well as the years and continent where the conference took place.

ICER was held primarily within the USA and Oceania during the years analyzed. Yet ITiCSE is a conference that frequently is placed in Europe.

In 2016, ITiCSE was held in Peru (Latin America). During that year, the participation from Latin America increased by a factor of 3. It is also noteworthy that this was the only year that the participation of Europe was not at the top. The results illustrate that the local continent where ICER took place increased the publication of locals from the hosting continent, but it did not generate any relationship nor show a specific trend. Irrespective of where ICER was located, more publications came from North America than from other continents.

| **Institution Name** | **Country** | **Amount of Publications** Total publications (conference-TOCE) |
|---|---|---|
| University of Toronto | Canada | 54(52-2) |
| The University of Adelaide | Australia | 38(35-3) |
| Monash University | Australia | 35(35-0) |
| Georgia Institute of Technology | US | 31(29-2) |
| North Carolina State University | US | 30(28-2) |
| University of Helsinki | Finland | 28(26-2) |
| Aalto University | Finland | 27(25-2) |
| University of Newcastle | Australia | 26(26-0) |
| Uppsala University | Sweden | 26(21-5) |
| University of Auckland | New Zealand | 25(25-0) |
| University of California, San Diego | US | 25(24-1) |
| University of Washington | US | 22(18-4) |
| University College Dublin | UK | 22(22-0) |
| Carnegie Mellon University | US | 20(17-3) |
| Indian Institute of Technology Bombay | India | 20(20-0) |
| Michigan Technological University | US | 19(18-1) |
| The Robert Gordon University | UK | 19(18-1) |
| University of Colorado at Boulder | US | 18(14-4) |
| Knox College | US | 17(17-0) |
| Virginia Polytechnic Institute and State University | US | 16(15-1) |
| Norwegian University of Science and Technology | Norway | 16(13-3) |
| University of Florida | US | 16(13-3) |
| Rochester Institute of Technology | US | 15(12-3) |
| University of Chicago | US | 14(11-3) |
| University of Warwick | UK | 14(11-3) |
| University of North Carolina at Charlotte | US | 14(13-1) |
| KTH Royal Institute of Technology | Sweden | 14(11-3) |
| Michigan State University | US | 14(12-2) |

Table 4: Top universities publishing

### 4.5 Collaboration Analysis

Collaboration is critical to gathering different perspectives and forming a community, and in an interdisciplinary field like computing education, it is even more

| Conference (ICER-ITiCSE) | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|
| Africa | | 0-4 | 0-1 | 0-1 | | 1-0 |
| Asia | 0-15 | 3-10 | 0-6 | 2-6 | 3-12 | 2-10 |
| Europe | 8-50 | 11-40 | 11-84 | 19-79 | 17-81 | 19-93 |
| North America | 39-38 | 33-55 | 36-72 | 28-64 | 49-81 | 42-66 |
| Oceania | 1-18 | 7-15 | 3-24 | 2-20 | 6-13 | 7-21 |
| Latin America | 1-4 | 0-22 | 0-7 | 0-7 | 1-7 | 2-4 |

Table 5: Number of publications (continent and year) ICER/ITiCSE

important in generalizing knowledge. For research in the field to move forward, with clear evidence for what was done, interventions, ideas, and results must be validated in multiple contexts, and muse be generalizable beyond the institution or country where the research was conducted. Furthermore, collaboration builds communities, through sharing and refining ideas. Here, we wanted to evaluate how authors and institutions work together, global publications (between different countries) and national (authors in the same country) collaborations. As such, we searched for publications with authors from more than one institution and from more than one country.

Fig. 1 illustrates collaborations over the 6 years analyzed. We noted a positive trend in the number of collaborations (national and international) in publications over the years. The atypical results from 2018 could be explained because both conferences were based in Europe (Cyprus and Finland), which could have led to less participation from other continents (Table 1).

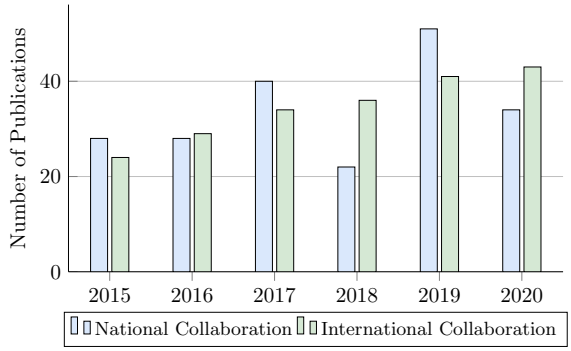


Fig. 1: General collaboration

**National Collaboration** To evaluate the level of collaboration that occurred at the national level (collaborations between institutions within the same country), we examined the institutional affiliations of all the authors on a paper. We observed that there were publications with up to 12 authors (a poster) and also 10 authors (one poster and one paper) from the same institution. In Fig. 2, we demonstrate the national collaborations observed in the sample. Looking at each venue individually, we note that at ICER there were publications with up to 4 different institutions from the same country, but the majority of collaboration over the years entailed authors from 2 local institutions. However, when considering the data from ITiCSE, sans working group (WG) publications

(since WG's have international collaborations), and ignoring 2018, the trend was positive overall. Yet ITiCSE has up to 5 different local institutions in the same publication, but as seen with ICER, the majority of local collaboration occurs within 2 institutions. When looking at the amount of collaborations in articles from TOCE, we observed there is not a general trend, and that the majority of collaboration is also between 2 local institutions.

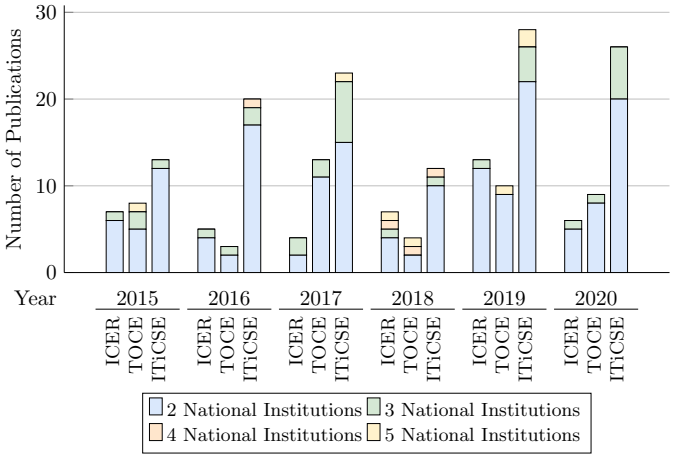


Fig. 2: National collaboration

**International Collaboration** Fig. 3 shows the data regarding international collaborations. ITiCSE was analyzed without WGs (separate analysis), since it otherwise distorts the data. We also did not consider data from doctoral consortia (DC), since they are by nature solitary pursuits. Comparing collaborations to the amount of work being accepted at ICER (in regard to new tracks that appeared), we observed the values of international collaboration were stable. In 2015, international collaboration represented 16% of the publications, and in 2016 they rose to 19%. Then, in 2017 they decreased to 7%. In 2018 they represented 13%, in 2019 they represented 16%, and in 2020 also 20%.

Looking at the data from ITiCSE (no WGs), we observed that in 2015 the international collaborations represented 11% of the publications. Yet, in 2016 they represented 19%, and in 2017 they represented 13%. Also, in 2018 they represented 25%, in 2019 they represented 12%, and in 2020 they were also 15%.

At ITiCSE, a WG is within a different track where collaboration is a necessity. The values show that they are distinct, and in 2015 international collaboration on WG was 86% of the total publications. In 2016 they were 43%, and in all other years, from 2017 to 2020 there were 100% of international collaborations in all WGs publications.

Looking at TOCE data, in 2015 the international collaborations represented 30% of the total papers published. In 2016, they represented 16% and in 2017 they represented 22%. However, by 2018 they only represented 6%. Yet in 2019 they went up to 24%, and then in 2020, they were at 11%.

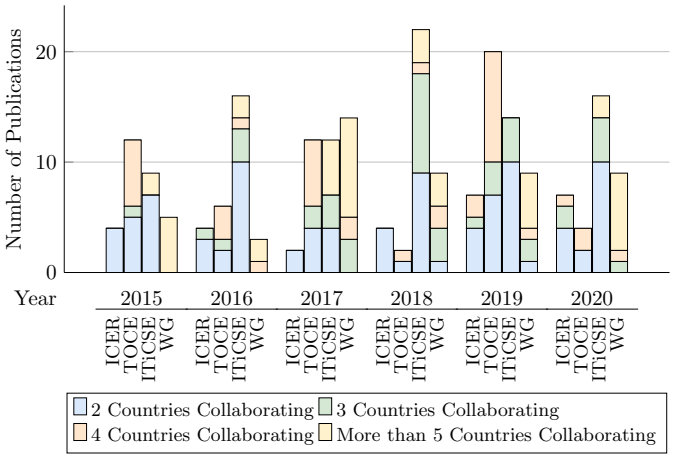


Fig. 3: International collaboration

## 5 Discussion

In this research, we gathered information about the landscape of research and collaborations in CEd. However, we do acknowledge that this is not a complete picture of the entire CEd community, but rather, a snapshot that gives us a sense of where CER is happening and with what level of collaboration.

### 5.1 Representation of Countries

In Section 4.2, we analyzed the continents and countries that contributed the most amount of work to the publications considered (RQ1). Primarily, contributions came from North America, Europe and Oceania — in particular the USA, United Kingdom, Australia, Canada, Finland and Germany. This is evident both in terms of number of publications and authors from these venues. There could be a number of reasons, but one potential explanation is a difference in relative wealth. Previous research by others has shown that a country's financial resources could impact participation [3, 14]. Therefore, we also examined the relationship between the countries represented and their GDP (Growth Domestic Product). Looking at the list of countries that published in the three venues, and comparing with the GDP per capita [1] we observed that the majority of the countries who published a total of more than 10 publications, also had a GDP per capita of more than $10,000 in United States Dollars (USD). There were only three exceptions: India (GDP per capita of $2,104.1 USD and 21 publications), Peru (GDP per capita of $6,997.7 USD and 21 publications), and Brazil (GDP per capita of $8,717.2 USD and 15 publications). Our findings serve to further reinforce the relationship between the economy and research presented. Given that many conferences require in person attendance, it is not entirely unsurprising that there could be a financial burden that affects who applied and attends. However, providing options like virtual attendance to conferences, could improve global communication and participation.

During the COVID-19 pandemic, many events and conferences shifted online [4, 17]. While this change did present its own technical challenges, it encouraged

new ways of meeting and communication [4]. As conferences were forced to pivot, expansion of creative ways to interact virtually, through Discord chats or via Zoom breakout rooms encouraged communication while affording researchers the ability to meet remotely. Scholars have described that virtual conferences may not only be more affordable and less time-consuming than in person meetings, but they may be more inclusive to expansion efforts by being more globally accessible [17]. Future gatherings and meetings should consider continuing to investigate the benefits of digital tracks, and could offer them as the primary means of hosting meetings, or in conjunction with in person options, to mitigate attendance costs and encourage greater participation.

### 5.2 Collaborations

In Section 4.5, we described a growing trend in heightened collaborations, particularly when separating out national from international collaborations (RQ2). In 2015, 40% of the publications considered involved collaborations, in 2016 this dropped to 39%, in 2017 it was 47%, in 2018 it was 40%, in 2019 it was 44%, and in 2020 it was 43%. Alternatively, only 20% of the posters published involved collaborations, which makes sense given that posters are frequently used to show work being done that remains in early stages, and to formulate ideas with feedback and discussion. So, when disregarding data from posters in the collaboration analysis, the numbers of collaborations increased further: 49% of all publications in 2015 involved collaborations, 48% in 2016, 56% in 2017, 47% in 2018, 58% in 2019, and 53% in 2020.

Comparing this information against data regarding the institutional affiliations, the top institutions came from countries with more publications, but in a different order. Moreover, at these venues, certain institutions published more. Some of the top institutions tended to focus on conference papers (but since we only considered one journal, this is not necessarily true for other journals) like Monash University. While it would also be interesting to evaluate the authors' department and/or school of origin, we were only able to retrieve this information for less than 20% of the publications, since reporting such information is at the author's discretion. The top institutions in these venues and in our timeframe were: University Of Toronto, Canada; University Of Adelaide, Australia; Monash University, Australia; Georgia Institute Of Technology, USA; University Of Helsinki, Finland; North Carolina State University, USA.

We also wanted to explore which countries offered a PhD in CEd, or in a related field. Since conference publications do not display the professional titles of authors, we instead focused our efforts on the DC, under the assumption that participants from this track would offer insight into which institutions are currently doing research in CEd. The results demonstrated that students from 22 different countries participated in the DC. During the time frame examined, ten countries just sent one student to the DC (Nigeria, Estonia, Czech Republic, Belgium, Saudi Arabia, Netherlands, Denmark, Portugal, Israel). This represents less than 8% of the students that attended. Furthermore, nine countries sent more than one but less than seven students to the DC (New Zealand, Sweden, Italy, Ireland, Norway, India, Australia, Finland and Canada) representing

24% of students that participated. However, three countries (Germany, United Kingdom and United States) represented more than 68% of the students that participated on the DC. This breakdown of the representation of countries in the DC is a bit different than the representation from the number of publications. Interestingly, we observe that countries that tended to have more publications in general, did not appear, or were not particularly well positioned, in the DC participation. Examples of this included France, Italy, Ireland. This raises questions about the root causes, and how to equalize representation. Moreover, it leads us to ask, if students with higher representation are traveling elsewhere to seek degrees, upon achievement of their doctorate, do they remain abroad, or return home to continue their research in the home country? Again, this is a topic which requires additional exploration, and potentially qualitative feedback would help to better discern the relationships observed.

When exploring the international collaboration data, we acknowledge that evaluating only our data as an amalgamation of both conferences and a journal may have its own limitations, since they are not necessarily comparable. Yet, this data did help to provide a snapshot of the work recently completed in computing education. Also, within the venues gathered in this six year time frame, there were 1,099 publications, which corresponded to a total of 2,088 authors. This yielded an average of 1.9 author(s) per paper.

To truly create a global CSEd for all environment, we need to reconsider current practices and how they may serve to limit participation. As mentioned, one potential option is to offer virtual attendance to conferences, rather than solely requiring in person participation. However, there are other options as well. For example, given that the current major conferences in CSEd all require proficiency in English, perhaps alternatives could be established to cater to different populations that speak other languages. One potential idea could be creation of an offshoot conference or session, that allows submissions to be written and reviewed by others fluent in a particular language, e.g., Spanish or Hindi (based on the languages spoken in some of the top countries that contributed hail from Asia and Latin America). Then, these works could be published and presented during their allotted time in the given language, but could also be potentially translated into English after the fact to allow the information to be disseminated to a larger audience. While there may be a huge initial overhead to set up such a program or event, doing so may help to ensure that international voices are able to share their work with the rest of the global CEd community.

## 6 Threats to Validity

Although this research was meant to provide an overview of the work conducted globally in CEd, we acknowledge that there are several threats to validity. Firstly, we only considered two conferences (ICER and ITiCSE data from 2015 to 2020) and one journal (TOCE up to July 2020) which does not represent all the work being published by the CER community. Apart from the fact that CEd researchers may publish in other disciplines, there is also a potential bias in the publications chosen (all ACM - anchored in the USA). While future research should consider expanding to additional venues, it was infeasible to do so here

due to the enormous effort required to gather all the data, clean it, and put it into a database for analysis.

In part, this issue was ameliorated by identifying a canonical list of higher education institutions around the world. While we were able to find a resource for the United Sates, we were unable to find one that represented the full breadth of institutions across the globe. Another concern that arose was institutions with multiple campuses. To standardize our efforts, we opted to align the records with the formal options listed by the Department of Education. If they consolidated locations under a single name, so did we. However, this became more challenging with international institutions. Furthermore, while the researchers of this work examined and manually checked authors' names and affiliations multiple times to note when they switched institutions, it is still possible that some of the instances went undetected.

Finally, we would like to note that participation in CEd is not only represented by publications accepted. Although conferences or TOCE may play a role, researchers can choose other ways to share their work too. However, for the purposes of analysis it was easier to work with information publicly available, and to utilize this information as a first step towards profiling the field.

## 7 Conclusions

Our results demonstrate that collaborations exist in publications submitted to the venues assessed, and that these collaborations have increased in recent years for the conferences (although not in the journal). The amount of international collaborations in ICER and ITiCSE did not change much over the time frame examined, but the number of authors per paper, and the corresponding collaborations, were higher when compared to prior literature. Also, when examining the ITiCSE WGs, it is evident that there was an increase in the amount of authors per publication, and also an increase in international collaborations on it. We also observed inequalities in representation that point to a need to further expand work to ensure a global "CSEd for all" attitude. Collaboration is critical to the development of a CEd community, to sharing knowledge, and to developing formal programs in CEd. However, as we observed in our analysis of collaborations, even though many institutions may collaborate, often the scope of these collaborations are limited. This research provided an overview of the countries and collaborations involved in computing education research and laid the groundwork for expanding this kind of analysis. While this is just a first attempt to profile where CER occurs, it provides useful insight into how we are currently working together — and suggests that we still have a long way to go to establish a global community.

## Acknowledgment

We would like to thank the entire team at SageFox for their assistance on this work. Funding for this research was made possible by the National Science Foundation (Award #1939265). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Bibliography


[1] Gdp per capita (current us$), 2020. URL https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.

[2] ANONYMOUS. Anonymous. In *Proceedings of SIGCSE 2021*, SIGCSE '21, New York, NY, USA, 2021. Association for Computing Machinery.

[3] T. Anwar, A. Jimenez, A. Bin Najeeb, B. Upadhyaya, and M. M. McGill. Exploring the enacted computing curriculum in k-12 schools in south asia: Bangladesh, nepal, pakistan, and sri lanka. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*, ICER '20, New York, NY, USA, 2020. Association for Computing Machinery.

[4] A. Bilas, D. Kostic, K. Magoutis, E. Markatos, D. Narayanan, P. Pietzuch, and M. Seltzer. The eurosys 2020 online conference: Experience and lessons learned. *arXiv preprint arXiv:2006.11068*, 2020.

[5] L. A. DeLyser. The who, what, and why of #csforall, Dec 2017. URL https://medium.com/csforall-stories/the-who-what-and-why-of-csforall-e7b60e2d543f.

[6] A. Education and C. E. Agency. Digital education at school in europe, Aug 2019. URL https://eacea.ec.europa.eu/national-policies/eurydice/sites/eurydice/files/en_digital_education_n.pdf.

[7] S. Fincher and M. Petre. *Computer science education research*. CRC Press, 2004.

[8] M. Foundation. About mariadb server, 2020. URL https://mariadb.org/about/.

[9] Q. Hao, D. H. Smith IV, N. Iriumi, M. Tsikerdekis, and A. J. Ko. A systematic investigation of replications in computing education research. *ACM Transactions on Computing Education (TOCE)*, 19(4):1–18, 2019.

[10] L. Malmi. Supervisor's perspective. *ACM Inroads*, 6(2):27–28, 2015.

[11] L. Malmi, J. Sheard, R. Bednarik, J. Helminen, P. Kinnunen, A. Korhonen, N. Myller, J. Sorva, and A. Taherkhani. Theoretical underpinnings of computing education research: what is the evidence? In *Proceedings of the tenth annual conference on International computing education research*, pages 27–34, 2014.

[12] S. Masapanta-Carrión and J. Á. Velázquez-Iturbide. A systematic review of the use of bloom's taxonomy in computer science education. In *Proceedings of the 49th acm technical symposium on computer science education*, pages 441–446, 2018.

[13] R. McCartney and K. Sanders. 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*, pages 332–337, 2018.

[14] V. L. Meek, U. Teichler, and M.-L. Kearney. Higher education, research and innovation: Changing dynamics. In *A report on the UNESCO forum on*

*higher education, research and knowledge. International Centre for Higher Education Research Kassel, University of Kassel*, 2009.

[15] J. Miró Julià, D. López, and R. Alberich. Education and research: evidence of a dual life. In *Proceedings of the ninth annual international conference on International computing education research*, pages 17–22, 2012.

[16] U. B. of Labor Statistics. Bureau of labor statistics, u.s. department of labor, occupational outlook handbook, Sep 2019. URL https://www.bls.gov/ooh/.

[17] S. Sarabipour. Research culture: Virtual conferences raise standards for accessibility and interactions. *Elife*, 9:e62668, 2020.

[18] Simon. A picture of the growing icer community. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ICER '16, page 153–159, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450344494. doi: 10.1145/2960310.2960323. URL https://doi.org/10.1145/2960310.2960323.

[19] Simon and J. Sheard. Twenty-four years of iticse papers. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, pages 5–11, 2020.

[20] D. W. Valentine. Cs educational research: a meta-analysis of sigcse technical symposium proceedings. *ACM SIGCSE Bulletin*, 36(1):255–259, 2004.