

Trends, Tribes, and Territories in Computer Science Education Research

Justus J. Randolph, Ph.D.
Walden University

George Julnes, Ph.D.
University of Baltimore

Erkki Sutinen, Ph.D.
University of Joensuu

Abstract

Understanding the trends, tribes, and territories of computer science education research can help its researchers gain insights into the disciplinary identity of their field and add meaning to their professional lives. To that end, this article presents the results of a large scale, empirical review of the recent computer science education literature. This review's focus was on the identification of the research-related trends, tribes, and territories within computer science education and across the fields of educational technology and education research proper. It was shown that there are several prominent trends occurring in computer science education research, that researchers from different geographic regions tend to do research differently, and that research practices across fields differ.

1. Introduction

Although computer science education research is characterized as “an emerging field, just giving rise to a literature” (Fincher & Petre, 2004, p. 1), it is mature enough to have already gained a unique disciplinary identity and history. Part of that identity manifests itself in what Becher and Trowler (2001) call the “academic tribes and territories” of computer science education. (In this paper, by *tribes* we mean the various institutions or research groups in which computer science education researchers function. By *territories* we mean the general geographical region in which the research communities live and work.) This identity also manifests itself in the ebb and flow of research trends. Understanding the trends, tribes, and territories in computer science education research can help computer science education researchers gain insights into the identity of their discipline and, thus, add meaning to their professional lives. On this point, Becher and Trowler (2001) state

The ways in which particular groups of academics organize their professional lives are related in important ways to the intellectual tasks in which they are engaged. We argue that the ways in which academics engage with their subject matter, and the narratives they develop about

this, are important structural factors in the formulation of disciplinary cultures. Together, they represent features that lend coherence and relative permanence to academics' social practices, values and attitudes across time and places. (p. 21)

Further, by understanding the research practices common in other tribes, territories, or time periods, researchers can improve the variety of research methods they choose to adopt. Young researchers can, for example, use the information to find their “research family” and chart their research lineage. The gatekeepers of computer science education research (e.g., editors, reviewers, and funders) can put policies in place to guide the discipline on an appropriate course.

Because of the benefits mentioned above, we empirically reviewed the recent computer science education research literature to identify its major trends, tribes, and territories. Specifically, we used the method of quantitative content analysis to review a random sample of 352 papers published in mainstream computer science education forums between 2000 and 2005. In addition, we compared and contrasted research practices between the fields of computer science education research, educational technology research, and education research proper.

As we report in the following section, there have been reviews of the computer science education literature that identify trends over time (Valentine, 2004) or report on the proportion of regions from which computer science education conference papers came (Randolph, Bednarik, & Myller, 2005). However, in our investigation we improve on the breadth, depth, and reliability of those previous reviews. To improve the breadth, we select samples from a representative set of eight mainstream computer science education journals and conference proceedings. The previous reviews had only taken samples from one journal or conference proceeding. To improve the depth, we code each article using up to 120 variables. Valentine, in contrast,

coded only on one variable. To improve reliability, we assess the interrater reliabilities for each variable.

2. Previous Reviews

Two previous reviews have empirically investigated trends or regional differences in computer science education research. The first is Valentine's (2004) well-known review that examined 20 years' worth of articles dealing with first-year computer science instruction published in the SIGCSE Technical Symposium. In that review, Valentine categorized each of 444 articles into one of six categories: (a) *Experimental*, defined as an author's making "any attempt at assessing the 'treatment' with some scientific analysis" (p. 256); (b) *Marco Polo*, "I went there and saw this" types of papers (p. 256); (c) *Tools* - papers dealing with teaching and learning tools; (d) *John Henry* - papers where the author describes "a course that seems so outrageously difficult...that one suspects it is telling us more about the author than it is about the pedagogy of the class" (p. 257); (e) *Philosophy* - papers "where the author has made an attempt to generate debate on an issue, on philosophical grounds, among the broader community" (p. 257); and (f) *Nifty* - papers that come up in "the [nifty] panels that are now a fixed feature of the [SIGCSE Technical Symposium Proceedings]" (p. 257).

Valentine's review identified several trends in the computer science education literature from 1984 to 2003. First, the number of proceedings published in the SIGCSE Technical Symposium Proceedings had been increasing yearly. Second, the percentage of experimental articles had increased over time. Third, the percentage of Marco Polo articles decreased over time. Valentine did not examine how research practices differed across regions.

Randolph, Bednarik, and Myller (2005) completed a methodological review of all of the articles published in the proceedings of the *Koli Calling: Finnish/Baltic Sea Conference on Computer Science Education* between 2001 and 2004. They examined the regions from which the articles published in the Koli Calling proceedings originated. However, they did not examine how research practices differed across regions.

While the Valentine (2004) review sheds some light on the trends over time, the reliability and utility of the Valentine review is questionable. In terms of reliability, Valentine was the sole rater so it is unknown if other raters would have categorized the articles the same way Valentine did. In fact, there is evidence that Valentine's coding system is only marginally reliable. In Randolph (2007), two raters completed a systematic replication of Valentine's coding system, but could not arrive at a high level of agreement about which category should be applied to an article (the value of *kappa*, a chance-adjusted measure of agreement, when using Valentine's system was .62). Putting the

question of the reliability of Valentine's review aside, Valentine's review nevertheless lacks enough depth to provide recommendations with enough detail to significantly improve policy or practice - Valentine only coded each article on one variable (i.e., which of the six categories listed above did each article belong).

Although the Randolph, Bednarik, and Myller review (2005) was shown to be reliable, it lacked breadth. Their results applied only to a small subsection of the computer science education literature (i.e., papers published in a small, regional conference in eastern Finland). In addition, they did not examine how research practices covaried between regions.

3. Purpose and Research Questions

Because the previous reviews of this kind had lacked reliability, depth, or breadth, we conducted a reliable, deep, and broad review of the computer science education literature in order to identify and analyze how research practices have recently varied over years and regions. We expanded on the depth of Valentine's (2004) coding system by concentrating on five variables that we consider to be important indicators of methodological quality, and analyzing how those variables covaried over time and region. (The five variables that we focused on are described in detail in the Method section). In addition, we also made comparisons across fields - to educational technology research and to education research proper. We improved upon the previous reviews by creating a reliable and replicable coding system. Finally, because we did careful sampling of articles, our results should be generalizable to all of the mainstream computer science education literature published between 2000 and 2005.

We list the research questions of the current review below:

1. Is there a yearly trend (from 2000-2005) in terms of:
 - the frequency of articles providing only anecdotal evidence,
 - the frequency of articles using experimental/quasi-experimental research methods,
 - the frequency of articles using explanatory descriptive (i.e., qualitative) research methods, frequency of articles in which the one-group posttest-only design was exclusively used, and the frequency of articles in which attitudes were the sole dependent variable?
2. Is there an association between:
 - the region of the first author's institutional affiliation and frequency of articles providing only anecdotal evidence,
 - the frequency of articles using experimental/quasi-experimental research methods,

- the frequency of articles using explanatory descriptive research methods,
- the frequency of articles in which one-group posttest-only designs were exclusively used, and
- the frequency of articles in which attitudes were the sole dependent variable?

In addition to investigating differences over time and region within the field of computer science education, we also investigated islands of practice among the related fields of computer science education, educational technology, and education research proper. Our research question in this area follows:

3. How do the proportions of quantitative, qualitative, and mixed methods articles in computer science education compare to those proportions in the fields of educational technology and education research proper?

4. Method

In this section, we give an overview of the methods used in this investigation. More specific information about the method can be found in Randolph (2007), which is the dissertation on which this article is based.

4.1 Sample

We chose a random sample of 352 computer science education articles from the 1306 full papers published in eight of the major computer science education research forums:

- *SIGCSE Bulletin* [hereafter *Bulletin*],
- *Computer Science Education* (the journal), [hereafter *CSE*],
- *The Journal of Computer Science Education Online* [hereafter *JCSE*],
- *Proceedings of the Koli Calling: Finnish/Baltic Sea Conference on Computer Science Education* [hereafter *Koli*],
- *Proceedings of the SIGCSE Technical Symposium on Computer Science Education* [hereafter *SIGCSE*],
- *Proceedings of the Innovation and Technology in Computer Science Education Conference* [hereafter *ITiCSE*],
- *Proceedings of the Australasian Computing Education Conference* [hereafter *ACE*], and
- *Proceedings of the International Computer Science Education Research Workshop* [hereafter *ICER*].

By *full papers* we mean peer-reviewed papers that were three or more pages long. We excluded invited columns, poster summaries, demo papers, working group reports, editorials, summaries of panel sessions, and similar types of reports. A complete list of articles included in the sample can be found in Randolph (2007). We stratified the sample by year and forum to ensure that articles from each forum and year were represented. Of the 352 articles, we focused primarily on the articles that reported on investigations dealing with human participants. Table 1 shows the number of articles that were sampled from each forum and year.

Table 1
Number of Articles Sampled From Each Forum and Year

<i>Year/forum</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>Total</i>
Bulletin	8	6	6	11	10	10	51
CSE	5	5	5	5	5	4	29
JCSE	0	0	2	1	0	0	3
KOLI	0	4	3	3	6	7	23
SIGCSE	21	21	20	20	25	28	135
ITiCSE	12	12	11	11	12	18	76
ICER	0	0	0	0	0	4	4
ACE	0	0	0	9	13	9	31
Total	46	48	47	60	71	80	352

4.2 Coding Procedure

Two individuals were involved in the coding of the 352 articles. The primary coder, the first author of this article, coded all of the articles. A second coder independently coded a simple random sample of 54 of the 352 articles so that estimates of interrater reliabilities could be established.

At the time of coding, the primary coder was a PhD candidate in education research and evaluation; the secondary coder was a PhD student in computer science (specifically, human computer interaction). Both coders were familiar with research methodology and with the computer science education research literature. The primary

coder's data were used in this analysis; the secondary coder's data were only used to calculate estimates of interrater reliability.

To analyze each article, the coders used a coding book and coding sheet developed by the first author. The coding book and coding sheet explains in detail how to code each article in terms of up to 120 variables. The coding book and coding sheet can be found in Randolph (2007).

4.3 Variables Examined

Although each article was coded for up to 120 variables, eight variables are of particular interest in this article. Those variables are (a) the year the article was published, (b) first author institution and geographical region of that institution, (c) whether the article dealt with human participants, (d) whether the article presented only anecdotal evidence, (e) whether the article reported on an explanatory descriptive (i.e., qualitative) investigation, (f) whether the attitude of teachers or students was the sole dependent variable, (g) whether an experimental/quasi-experimental research design was used, and (h) whether an experimental/quasi-experimental design was used, if the one-group posttest-only design was used. We chose to focus on these variables because we consider them to be important indicators of methodological quality in computer science education research articles. The use of experimental/quasi-experimental or explanatory descriptive methods enables researchers to make causal conclusions. Although attitudes are an important variable of interest, especially in a discipline with such a high degree of student attrition, attitudes (especially self-reports of learning) are historically unreliable as measures of learning. The one-group posttest-only design is subject to almost all threats to internal validity (Shadish, Cook, & Campbell, 2002) and, therefore, the validity of causal conclusions based on data gathered from this design is often in question. In addition to the eight main variables described above, to answer the research question involving differences across fields, we also categorized each empirical article that dealt with human participants into three categories: quantitative, qualitative, or mixed-methods.

Types of articles that did not deal with human participants included descriptions of specific programs, interventions, or tools; theoretical, methodological, or philosophical papers; literature reviews; and technical papers. The papers that dealt with human participants were divided into two categories – those that presented anecdotal evidence only and those that presented more than anecdotal evidence. In this case, we considered anecdotal evidence to be evidence based on a researcher's informal reports of their informal, post-hoc personal observations. For example, if a researcher's only evidence was, "We noticed that the students liked using our innovative tool and learned a lot from it," we considered that to be anecdotal evidence. We

did not consider evidence to be anecdotal if it was clear that the researchers were using some sort of structured observation (e.g., direct observation of operationalized behaviors) or intentionally adhered to a recognized qualitative method, such as the case study, ethnography, etc. For example, if a researcher had verbally interviewed students and reported a synthesis of the interview results, that would not have been considered to be anecdotal evidence because the researcher adhered to a structured, recognized qualitative method and reported more than his or her informal, post-hoc personal observations about a phenomenon.

The articles that reported on investigations with human participants and that presented more than anecdotal evidence were coded on four additional aspects. First, they were coded on whether they used what Yin (2003) called an explanatory descriptive method. In this method the researchers probably, but not necessarily, carried out qualitative modes of inquiry to explain a phenomenon.

Second, those articles were also coded on whether an experimental/quasi-experimental mode of inquiry was used. By *experimental/quasi-experimental* we mean an investigation in which a researcher manipulates an independent variable and compares a counterfactual condition to a factual condition. By *experimental* we refer to designs in which participants were randomly assigned to control and treatment conditions. By *quasi-experimental* we refer to a design in which participants were assigned to contrast or treatment conditions by some method other than random assignment.

Third, if an article was an experiment or quasi-experiment, we coded whether the experimental/quasi-experimental design used was the one-group posttest-only design. In the one-group posttest-only design, the researcher introduces an intervention to one group and then takes a measurement only after the intervention had been introduced. There is no measurement before the intervention is given and there is no control or contrast group. We included studies that used retrospective posttests in the one-group posttest-only design category. (In a retrospective posttest, after the intervention has been given participants are asked to estimate, for example, how much they knew before the intervention or asked to estimate how much they learned.)

Fourth, articles were coded on whether the only measured dependent variable was students' or teachers' attitudes. If attitudes *and* some other dependent variable, such as achievement in computer science, were measured, then the article was not considered an *attitudes-only* article.

We made a distinction between quantitative, qualitative, and mixed-methods types of research. An article was considered to be quantitative if it reported only numerical results,

qualitative if it reported only qualitative results, and mixed-methods if it reported both quantitative and qualitative results.

In terms of the data about the proportions of human participant articles and mode of inquiry, the data concerning the field of educational technology came from a synthesis of 13 methodological reviews, covering 999 articles, of the educational technology research literature. Detailed information about that synthesis can be found in Randolph (2007). The data concerning the field of education research proper came from Gorard and Taylor (2004), which is based on a review of a purposive sample of 94 articles from leading British education research journals. Although Gorard and Taylor's sample of articles is relatively small, they validate their results through

- interviews with key stakeholders from across the education field, including researchers, practitioner representatives, policy-makers and policy implementers;
- a large-scale survey of the current methodological expertise and future training needs of UK education researchers; [and a]
- detailed analysis and breakdown of 2001 RAE [Research Assessment Exercise, 2001]. (p. 114)

4.4 Data Analysis

To analyze the data, we cross tabulated them, examined the adjusted residuals, and, for categorical variables calculated X^2 (see Agresti, 1996), and for ordinal variables, such as year, we calculated M^2 (Agresti) using SPSS 11.0, which treats the sample as if it had been simple random. Because we used a stratified sample and because stratification is known to have an effect on variability estimates, to double check our results we used bootstrapping analyses including the stratification layers. (We found that the differences between the bootstrapped and SPSS-generated probabilities were negligible.) The resampling codes for calculating X^2 and M^2 from a proportionally stratified random sample can be found in Appendix F of Randolph (2007). Since we planned to conduct 12 inferential tests in total (five tests to examine trends, five contrasts to examine regional differences, and two contrasts to examine differences across fields), we set the alpha level, a priori, for individual tests to .004 to arrive at an overall alpha level around .05 (see Stevens, 1999).

We used Brennan and Prediger's (1981) free-marginal *kappa*, a chance-adjusted measure of agreement, as the

measure of interrater reliability. The free-marginal *kappa* statistic varies from -1.0 to 1.0, where 1.0 is perfect agreement, 0.0 is as much agreement as would be expected by chance, and -1.0 is perfect disagreement. Although there are a variety of opinions on this subject (see Neundorf, 2002), we have adopted a rule of thumb that variables with free-marginal values of *kappa* .80 or above are certainly reportable, variables with values between .60 and .79 are reportable in most cases, and variables with values less than .60 should not be reported in most cases or should be very carefully qualified. Free-marginal *kappa* was calculated by measuring the number of agreements and disagreements between the primary and secondary raters.

5. Results

5.1 Complications

Because there were no papers in the sample from South America or Central America, and only one paper from Africa, that dealt with human participants, we ultimately decided to exclude these regions from our analysis. In addition, it turned out that there were some cells in the crosstabulations that had less than five cases, which is a violation of the assumptions for conducting conventional X^2 and M^2 tests (Agresti, 1996). However, we are confident that our probability estimates are accurate because (a) bootstrap analysis of the same data led to practically the same results as those from conducting conventional X^2 and M^2 tests, and (b) Randolph's (2007) regression analysis of the same data, but using grouped categories, resulted in essentially the same overall findings.

5.2 Sample Characteristics

Of the 352 articles in the sample, 233 dealt with human participants. Of those articles that dealt with human participants, 144 presented more than anecdotal evidence. Explanatory descriptive methods were used in 38 of the 144 cases, Experimental/quasi-experimental designs and explanatory descriptive methods were used in 93 and 38 of the 144 studies that presented more than anecdotal evidence, respectively. One-group posttest-only designs were used in 45 of the 93 experimental/ quasi-experimental studies. Explanatory descriptive methods were exclusively used in 22 studies. Mixed-methods were used in 15 studies. Of the 123 studies that reported quantitative results, 38 measured only students' or teachers' attitudes about an intervention. Figure 1 shows the total number of articles that originated from each region. See Randolph (2007) for additional information about sample characteristics.

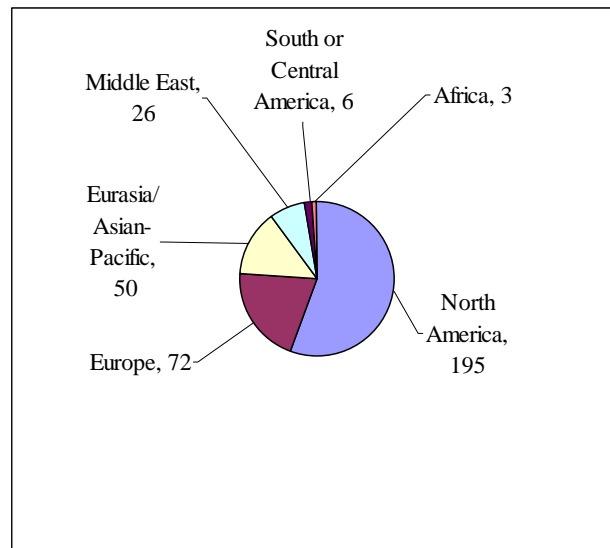


Figure 1. Number of articles by region of first author's affiliation.

5.3 Interrater Reliabilities

Table 2 shows the interrater reliabilities and their associated confidence intervals for the main variables in this study. Interrater reliabilities for factual variables, like year published, were not calculated. The reliability for all but one of the variables was high and the explanatory descriptive

variable had a marginally acceptable value of *kappa*—.65. On the explanatory descriptive variable, a dichotomous variable, the coders agreed in 14 of 17 cases (about 82% agreement). Based on the binomial distribution, the probability of agreeing of on at least 14 of 17 cases given chance is .00064.

Table 2

Interrater Reliabilities for Key Variables

Variable	<i>n</i>	<i>Kappa</i>	Lower CI 95%	Upper CI 95%
Explanatory descriptive	17	.65	.29	1.00
Experimental/quasi-experimental	17	.88	.65	1.00
Human participants	53	.81	.66	.96
Anecdotal-only	34	.94	.82	1.00
Attitudes-only	15	1.00	-	-

5.4 Trends

Out of the five planned contrasts involving yearly trends, two were statistically significant. The number of anecdotal articles and the number of explanatory descriptive articles had decreased from 2000 to 2005.

5.4.1 Anecdotal-only articles

Table 3 shows that there was a decreasing trend in the number of anecdotal-only articles from 2000-2005. The fact that the adjusted residuals in the *Percentage Yes* column

transition from large positive values in 2000 to large negative values in 2005 and that the percentages, more or less, transition from larger to smaller support the finding that there was a trend. (The degree to which the observed value differs from the expected value in a cell, given chance, is indicated by the adjusted residual. In this case, if there were no trend, the adjusted residuals would all have been around zero.) The trend was statistically significant, $M^2(1, N = 233) = 9.00, p = .003$; bootstrapped $p = .003$.

Table 3

Anecdotal-Only Papers by Year

Year	Anecdotal-only		Total	Percentage Yes	Adjusted residual
	Yes	No			
2000	18	13	31	58.1	2.4
2001	15	15	30	50.0	1.4
2002	9	17	26	34.6	-0.4
2003	14	25	39	35.9	-0.3
2004	18	34	52	34.6	-0.6
2005	15	40	55	27.3	-1.9
Total	89	144	233		

Figure 2, which shows the number of articles per year by region, provides additional insight into the nature of the trend. It shows that in 2000, North American papers clearly had the highest proportion of anecdotal-only papers, but over time proportions of those papers decreased to a level that is about the same as in other regions. In short, the overall decrease in anecdotal papers appears mainly to be due to a decrease in North American anecdotal-only papers.

In a regression analysis of these data, Randolph (2007) showed that there was indeed an interaction between year and region ($p = .07$). Note that to simplify the graph we combined regions with low numbers of articles. Specifically, we combined the Asian Pacific/Eurasian and Middle Eastern categories to create an Asian-Pacific/Eurasian et al. category.

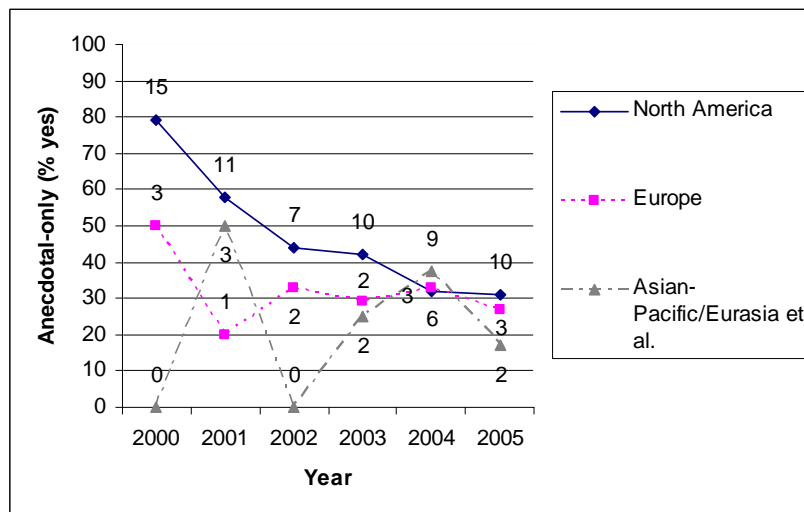


Figure 2. Anecdotal-only papers by combined region and year.

The value nearest to a data point shows the n-size for that data point.

5.4.2 Explanatory Descriptive Articles

Table 4 shows that there was a somewhat decreasing trend in the number of explanatory descriptive articles that were published each year. Although the trend was not consistent

(2002 was an exception to the trend), it was statistically significant, $M^2(1, N = 144) = 11.54, p = .001$; bootstrapped $p < .000$.

Table 4

Explanatory Descriptive Papers by Year

Year	Explanatory descriptive		Total	Percentage Yes	Adjusted residual
	Yes	No			
2000	7	6	13	53.8	2.4
2001	4	11	15	26.7	0.0
2002	8	9	17	47.1	2.1
2003	7	18	25	28.0	0.2
2004	9	25	34	26.5	0.0
2005	3	37	40	7.5	-3.2
Total	38	106	144		

Figure 3 shows the proportion and number of explanatory descriptive papers by year and region, and reveals considerable variability over years and regions. However,

there seemed to be a decreasing trend in North American papers. This trend was not found to be significant in Randolph's (2007) regression analysis.

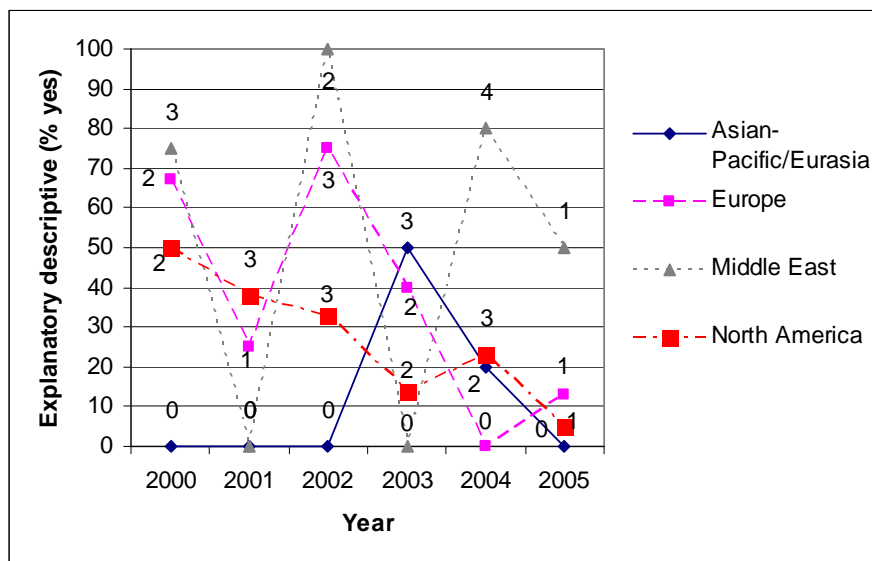


Figure 3. Explanatory descriptive papers by year and region.

The value nearest to a data point shows the *n*-size for that data point.

5.4.3 Other types of articles

Crosstabulations for the types of articles where there was not a statistically significant trend (e.g., experimental/quasi-experimental articles, one-group posttest-only articles, and attitudes-only articles) are presented below. Table 5 shows

that there was not a strong trend in the number of experimental/quasi-experimental papers that were published each year. Likewise, no trend is shown in Table 6, which shows the number of one-group posttest-only articles per year, or in Table 7, which shows the number of attitudes-only papers by year.

Table 5
Experimental/Quasi-Experimental Papers by Year

Year	Experimental		Total	Percentage Yes	Adjusted residual
	Yes	No			
2000	8	5	13	61.5	-0.2
2001	11	4	15	73.3	0.7
2002	10	7	17	58.8	-0.5
2003	14	11	25	56.0	-1.0
2004	22	12	34	64.7	0.0
2005	28	12	40	70.0	0.8
Total	93	51	144		

Note. $M^2(1, N = 144) = 0.17, p = .676$; bootstrapped $p = .676$.

Table 6
One-group posttest-only papers by year

Year	One-Group Posttest-Only		Total	Percentage Yes	Adjusted residual
	Yes	No			
2000	6	2	8	75.0	1.5
2001	6	5	11	54.5	0.4
2002	4	6	10	40.0	-0.6
2003	4	10	14	28.6	-1.7
2004	15	7	22	68.2	2.0
2005	11	17	28	39.3	-1.3
Total	46	47	93		

Note. $M^2(1, N = 93) = 0.97, p = .326$; bootstrapped $p = .315$.

Table 7
Attitudes-Only Papers by Year

Year	Attitudes-Only		Total	Percentage Yes	Adjusted residual
	Yes	No			
2000	1	8	9	11.1	-1.3
2001	6	7	13	46.2	1.3
2002	3	9	12	25.0	-0.5
2003	5	17	22	22.7	-0.9
2004	12	17	29	41.4	1.4
2005	11	27	38	28.9	-0.3
Total	38	85	123		

Note. $M^2(1, N = 123) = 0.20, p = .657$; bootstrapped $p = .665$.

5.5 Tribes

Table 8 shows the 12 institutions whose papers were most often randomly selected for the sample. The number of articles that should correspond with the number of articles in the population can be estimated by multiplying the number of articles in the sample for each institution by 3.71, which is the ratio of the number of articles in the population to the number of articles in the sample. The University of Joensuu, with 13 articles included in the sample, was an outlier. Of

those 13 articles, 11 were from the Koli Calling conference, a conference held in a remote location near Joensuu. In addition to the institutions mentioned above, several authors stood out as prolific contributors to the body of computer science education research literature between 2000 and 2005. Yifat Ben-David Kollikant was the most prolific author, with four articles in the sample. Other prolific authors were A.T. Chamillard, Orit Hazzan, David Ginat, H. Chad Lane, and Richard Rasala, each with three articles.

Table 8

Institutions with the Greatest Number of Computer Science Education Articles

<i>Institution</i>	<i>n (articles)</i>	<i>%</i>
University of Joensuu	13	3.7
Technion – Israel Institute of Technology	6	1.7
Drexel University	5	1.4
Northeastern University	5	1.4
Tel-Aviv University	5	1.4
Weizmann Institute of Science	5	1.4
Helsinki University of Technology	4	1.1
Michigan Technological University	4	1.1
Trinity College	4	1.1
University of Arizona	4	1.1
University of Technology, Sydney	4	1.1
Virginia Tech	4	1.1
Other institutions	289	82.4
Total	352	100.0

5.6 Territories - Within Computer Science Education

Of the five contrasts that dealt with the region of the first author's affiliation, three were statistically significant. The statistically significant findings are described below. Note that the total sums to 143, instead of 144 as in Tables 9 and 10 because the one African article that dealt with human participants was excluded.

5.6.1 Experimental/quasi-experimental Articles

Table 9 shows that first authors who were affiliated with institutions in North America tended to write and get published articles that used experimental or quasi-experimental articles. In contrast, first authors who were affiliated with institutions in Europe or the Middle East

tended not to write, or get published, experimental or quasi-experimental articles. In fact, based on the data from these crosstabulations, the odds of a first author affiliated with a North American association having published an experimental paper were more than 3.9 times greater than a first author affiliated with a European institution and more than 7.6 times greater than a first author affiliated with a Middle Eastern institution. The differences between observed and expected cell values in Table 9 were statistically significant, $\chi^2(3, N = 143) = 15.54, p = .001$; bootstrapped $p < .000$. In the regression analysis of Randolph (2007), there was not a statistically significant year by region interaction for this variable, but there was a forum type by region interaction.

Table 9
Experimental Papers by Region of First Author's Affiliation

Region	Experimental /Quasi-experimental		Total	Percentage Yes	Adjusted residual
	Yes	No			
Eurasia/Pacific	20	10	30	66.7	0.3
Europe	14	16	30	49.7	-2.3
Middle East	4	9	13	30.8	-2.6
North America	54	16	70	77.1	3.1
Total	92	51	143		

5.6.2 Explanatory descriptive articles

Table 10 shows that first authors who were affiliated with a Middle Eastern institution tended to write and get published explanatory descriptive articles. The odds of a first author affiliated with a Middle Eastern institution having written and had published an explanatory descriptive articles was more than 13 times greater than the odds of an author

affiliated with a North American institution having written and had published an explanatory descriptive article. The differences were statistically significant, $\chi^2(3, N = 143) = 20.13, p < .000$; bootstrapped $p < .000$. See Figure 3 for a breakdown of explanatory descriptive articles by region and year.

Table 10
Explanatory Descriptive Papers by Region of First Author's Affiliation

Region	Explanatory descriptive		Total	Percentage Yes	Adjusted residual
	Yes	No			
Eurasia/Pacific	5	25	30	16.7	-1.4
Europe	9	21	30	30.0	0.5
Middle East	10	3	13	76.9	4.3
North America	14	56	70	20.0	-1.7
Total	38	105	143		

5.6.3 Attitudes-only Articles

Table 11 presents the number and percentage of attitudes-only articles by region. The odds of a first author affiliated with an institution in the Asian Pacific or Eurasia having written and published an article in which attitudes were the sole dependent measure were more than 12 times greater

than the odds of a first author affiliated with an institution in Europe having done the same. The differences were statistically significant, $\chi^2(3, N = 122) = 17.39, p = .00$; bootstrapped $p < .000$.

Table 11
Attitudes-Only Papers by Region of First Author's Affiliation

Region	Attitudes-only		Total	Percentage Yes	Adjusted residual
	Yes	No			
Eurasia/Pacific	16	10	26	61.5	3.9
Europe	3	24	27	11.1	-2.5
Middle East	1	4	5	20.0	-0.5
North America	17	47	64	26.9	-1.0
Total	37	85	122		

5.6.4 Other types of Articles

Crosstabulations for the types of articles in which there were no statistically significant regional differences (e.g., anecdotal-only papers and one-group posttest-only papers) are presented in Tables 12 and 13 below. (Note that Randolph's [2007] logistic regression analysis showed, however, that region was a statistically significant predictor of an article's being an anecdotal-only article when other

factors are controlled for. Figure 2 illustrates that the decline in anecdotal articles was mainly exhibited in North American papers.) Figure 4 below shows a summary of how papers from different territories differed in their reported computer science education research practices.

Table 12

Anecdotal-Only Articles by Region of First Author's Affiliation

<i>Region</i>	<i>Anecdotal-Only</i>		<i>Total</i>	<i>Percentage Yes</i>	<i>Adjusted residual</i>
	<i>Yes</i>	<i>No</i>			
Eurasia/Pacific	10	30	40	25.0	-1.9
Europe	14	30	44	31.8	-1.0
Middle East	5	13	18	27.8	-.9
North America	59	70	129	45.7	2.7
Total	88	143	231		

Note. $\chi^2(3, N = 231) = 7.65, p = .054$; bootstrapped $p = .059$.

Table 13

One-Group Posttest-Only Papers by Region of First Author's Affiliation

<i>Region</i>	<i>One-group posttest-only</i>		<i>Total</i>	<i>Percentage Yes</i>	<i>Adjusted residual</i>
	<i>Yes</i>	<i>No</i>			
Eurasia/Pacific	13	7	20	65.0	1.6
Europe	8	6	14	57.1	0.7
Middle East	3	1	4	75.0	1.1
North America	21	33	54	38.9	-2.3
Total	45	47	92		

Note $\chi^2(3, N = 92) = 5.71, p = .127$; bootstrapped $p = .128$.

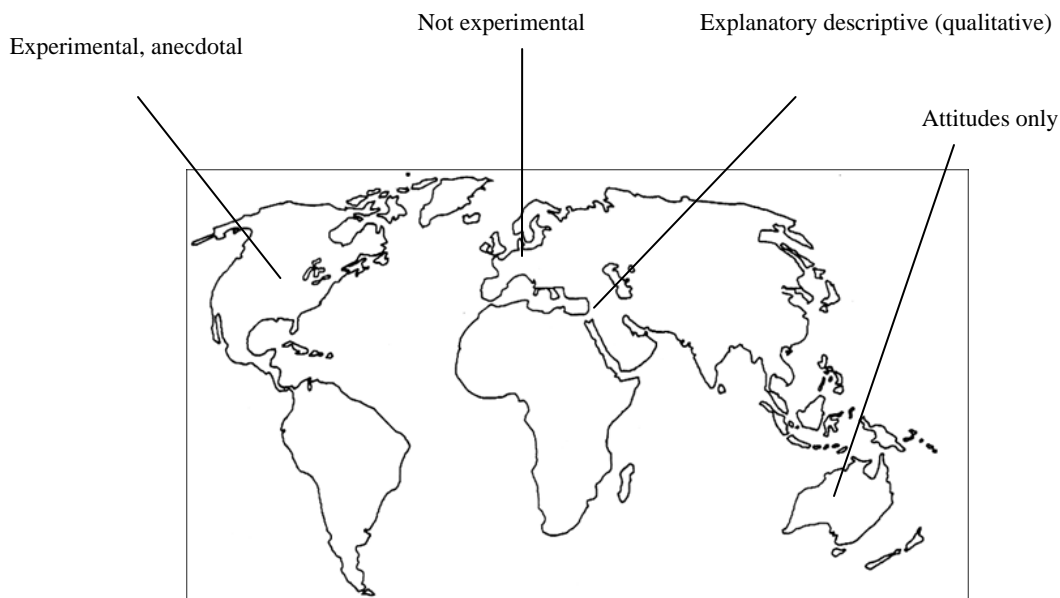


Figure 4. Summary of research tendencies across regions.

5.7 Differences between Fields

Up to this point, we have presented results within the field of computer science education. In this section, we present results concerning the proportions of empirical (i.e., not anecdotal) articles dealing with human participants and proportions of quantitative, qualitative, and mixed methods research between fields. Note that the proportions for the field of education research proper come from Gorard and Taylor (2004) and the proportions for the field of educational technology come from the review of methodological reviews of educational technology, which was originally presented in Randolph, (2007). Please note the Gorard and Taylor articles only include British research articles.

5.7.1 Proportions of Empirical Articles Dealing With Human Participants

Table 14 shows that the proportions of nonanecdotal articles dealing with human participants decreased monotonically from education research proper to educational technology and from educational technology to computer science education. Assuming that those fields are ordinal in terms of the degree to which they have a computer science tradition (where computer science education has the largest degree of the computer science tradition and education research proper has the least), indicated by the number of articles that do not deal with human participants, the results of the M^2 test, indeed, showed that there was a statistically significant linear (monotonic) relationship, $M^2(1, N = 1,351) = 52.32, p < .000$. The adjusted residuals in Table 14, which ranged from 6.2 for education research proper and -5.3 for computer science education, showed that the linear relationship was pronounced.

Table 14

Comparison of the Proportion of Nonanecdotal Research with Human Participants across Fields

Field	<i>Nonanecdotal research with human participants</i>		Total	Percentage Yes	Adjusted residual
	Yes	No			
Ed. proper	79	15	94	84.0	6.2
Ed. tech.	494	411	905	54.6	1.6
CSE	144	208	352	40.9	-5.3
Total	717	634	1,351		

Note. Ed. proper = education proper, Ed. tech. = educational technology, CSE=computer science education.

5.7.2 Proportions of Research Traditions between Fields

Table 15 shows that there was a statistically significant difference, $\chi^2(2, N=638) = 20.84, p < .000$, between the proportions of quantitative, qualitative, and mixed methods articles in computer science education and educational technology forums. The adjusted residuals show that authors

of computer science education articles tended to write, and get published, quantitative articles and tended to not write, or get published, qualitative-only articles, compared to authors of papers published in educational technology forums. The percentage of mixed-method articles in each field was about the same however.

Table 15

Comparison of the Proportion of Nonanecdotal Research with Human Participants Articles in Computer Science Education and Education Technology

Method	Field		Total	Percent CSE	Percent Ed. tech	Adjusted residual (CSE)
	CSE	Ed. tech				
Quantitative	107	280	387	74.3	56.7	3.8
Qualitative	22	174	196	15.3	35.2	-4.6
Mixed	15	40	55	10.4	8.1	0.9
Total	144	494	638			

Note. CSE=computer science education, Ed. tech = educational technology.

Table 16 shows that there was also a statistically significant difference, $\chi^2(2, N = 223) = 18.12, p < .000$, between the proportions of quantitative, qualitative, and mixed methods articles between the fields of computer science education and education research proper. The adjusted residuals show

that the authors of computer science education research articles tended to use quantitative methods and tended not to use qualitative methods. Again, the proportions of mixed methods articles were about the same across fields.

Table 16

Comparison of the Proportion of Nonanecdotal Research with Human Participants in Computer Science Education and Education Research Proper

Method	Field		Total	Percentage CSE	Percentage Ed. proper	Adjusted residual (CSE)
	CSE	Ed. proper				
Quantitative	107	43	150	74.3	54.4	3.0
Qualitative	22	32	54	15.3	40.5	-4.2
Mixed	15	4	19	10.4	5.1	1.4
Total	144	79	223			

Note. CSE=computer science education, Ed. proper = education research proper.

6. Discussion

In this section, we discuss where there were or were not differences in research practices - across regions, years, and fields. A summary of findings is provided below:

1. Controlling for other factors, North American articles had a higher proportion of anecdotal-only articles than most other regions.
2. North American articles had a higher proportion of experimental/quasi-experimental articles than other regions.
3. Middle Eastern articles had a much higher proportion of explanatory descriptive articles than articles from any other region.
4. Asian-Pacific/Eurasian articles had a higher proportion of attitudes-only articles than did articles from other regions.
5. The proportion of anecdotal-only articles had decreased each year; the strongest decrease was seen in North American articles. Also, the proportion of explanatory descriptive articles had decreased every year.
6. Articles from computer science education research forums had a higher proportion of articles that did not deal with human participants than did articles from educational technology or education research proper forums.
7. When doing empirical investigations involving human participants, computer science education researchers tended to use quantitative methods more than their counterparts in educational technology or education research proper.

6.1 Trends

Valentine (2004) identified several encouraging trends in the computer science education research from 1984 to 1999. First, the number of technical symposium proceedings had been increasing each year. Second, the percentage of experimental articles (loosely defined as the author having made “any attempt at assessing the ‘treatment’ with some scientific analysis” [p. 256]) had increased since the mid 1990s. Third, the percentage of Marco Polo articles (which probably would correspond with what we called anecdotal-only articles) had shown a yearly decrease.

The findings of this methodological review show that two out of the three trends identified by Valentine (2004), from 1984 to 1999, continued in the years from 2000 to 2005. First, as is evident from Table 1, the number of articles in the SIGCSE Technical Symposium (and in computer science education forums in general) was still on the rise. Second, the decline in the number of anecdotal-only/Marco Polo articles had continued to decline in the years from 2000 to 2005. The decline was most pronounced for North American articles. In contrast to what Valentine found, this study did not find the proportions of experimental articles had continued to increase in the years from 2000 to 2005. However, it is important to note here that we used a more conservative definition of *experimental* than did Valentine.

We assumed that in addition to true experiments or quasi-experiments, Valentine would have included explanatory descriptive, exploratory descriptive, correlational, and causal comparative investigations in the “experimental” category. We, on the other hand, only included actual experiments or quasi-experiments in the experimental category.

6.2 Tribes

By grouping the institutions who contributed the most to the computer science education literature, three regions of significant influence become clear—the U.S., Finland, and Israel. However, the degree to which the institutions within those regions work together to create a “tribal family” is not clear.

6.3 Territories within Computer Science Education

There are several differences in the way that computer science education researchers from institutions in different regions conduct research. First, computer science education researchers from North American institutions tended to do experimental research, while their European and Middle Eastern counterparts tended not to do experimental research. Second, computer science education researchers from Middle Eastern institutions strongly tended to do explanatory descriptive (qualitative) research. Third, North American researchers tended to do anecdotal-only research more than their peers in other regions, but the proportions of North American anecdotal research articles had been on the decline while the proportions had been stable across time for the other regions. Fourth, computer science education researchers from Asian-Pacific or Eurasian institutions tended to measure attitudes only.

Disentangling the relationship between the factors related to the environment in which a group of scientists work and how they carry out their research is difficult (see Depaepe, 2002). It is like speculating how the work of the Vienna School, for example, would have been different had they been the Toledo (Ohio) School instead. Nonetheless, we describe some of our hypotheses below, which might be used to inform further investigations, about why the results may have turned out as they did.

One of our findings was that there was a tendency for North American researchers to do experimental research and a tendency for European researchers not to do experimental research. Van Helden, Johnsen, and Vakkuri (2006), in a review of the public sector performance measurement research, reported findings that are congruent with our own. They found that European researchers conducted 16% more qualitative types of research (e.g., case studies or field studies) than their U.S. counterparts. Conversely, U.S. researchers conducted 19% more quantitative (e.g., survey research) research than their European counterparts.

One possible reason for the tendency of North American education researchers to do experiments may be that the

worth attributed to randomized field trials by the U.S. Department of Education, a major source of funding for U.S. education researchers, has something to do with the tendency of North American researchers (of whom most are from U.S. institutions) to do experimental research. The U.S. Department of Education (2002) made the following statement about the relative importance they give to descriptive studies and to “rigorous field trials of specific interventions”:

Descriptive implementation studies play a crucial role in understanding the impact of policy changes, but they are no substitute for rigorous field trials of specific interventions.

Even with high-quality fast-response surveys, annual performance data, and descriptive studies, we still cannot answer the question on the minds of practitioners: “What works?” To be able to make causal links between interventions and outcomes, we need rigorous field trials, complete with random assignment, value-added analysis of longitudinal achievement data, and distinct interventions to study.

This approach might be considered “research” rather than “evaluation.” Whatever the name, the Department’s evaluation agenda would be incomplete without it. It is a fair use of evaluation dollars because federal program funds are paying for the interventions to be studied. (para. 24-26)

This policy is a hotly debated topic in U.S. research and evaluation circles (see Donaldson & Christie, 2005; Julnes & Rog, in press; Lawrenz & Huffman, 2006). Regardless of the propriety of this policy, the quote above shows that U.S. educational policymakers give value and funding priority to true experiments, so it is not surprising that many U.S. education researchers choose to do experimental research.

Second, the tendency of European researchers to not do experimental research is congruent with the contemporary European decline in the popularity of the study of quantitative research methods. Rautopuro and Väisänen (2005), well-known Finnish quantitative-research-method educators, wrote the following about the state of quantitative research methods, at least in Finland:

The level of skills in the quantitative methods seems to be worrying. In educational science, too, the level of method used as well as how they are used in quantitative research in all levels—from master theses to dissertations—is getting out of hand. The students do not get excited about taking voluntary quantitative research methods courses and therefore are not capable to use them in their own research. Compulsory statistics courses, as well, are only a necessity for the students and sometimes for the researcher, too. Moreover, one generation of educational researchers, at least partially, has lost the competence of applying quantitative research methods and because of

this they have also lost the possibility to pass on the tradition of the use of these methods. (p. 273)

If Rautopuro and Väisänen’s (2005) findings can be generalized to the rest of Europe (and there is reason to believe that it does - see European Science Foundation, 2004), then it is no surprise that there is a tendency for European computer science researchers not to do experimental research.

One possible reason for this could be that the resurgence of the qualitative research tradition has had a greater influence in Europe than in North America, according to Fielding (2005). Fielding speculates that the

The American quantitative approach was influential during this period [i.e., the resurgence of the qualitative method since the publication of Glaser and Strauss’s *Discover of Grounded Theory* in 1967, Strauss and Corbin’s revision of it in 1990, and Turner’s influential 1981 paper on qualitative data analysis] too but qualitative methodology was arguably more secure in the European curriculum due to the import of hermeneutics in German social philosophy and the life history method in French and Italian sociology. (2005, para. 12)

Fielding (2005) also mentions that since the 1980s, qualitative research has become increasingly legitimized and institutionalized in the European social science research curriculum. An example of this institutionalization of qualitative research that Fielding provides are the postgraduate training guidelines written by the United Kingdom’s Economic and Social Research Council (ESRC). According to Fielding, those curriculum guidelines

strongly emphasize qualitative methods and require that students understand archival, documentary and historical data, life stories, visual images and materials, ethnographic methods, cases studies and group discussions, at least one qualitative software package, and a range of analytic techniques including conversation analysis and discourse analysis. Since the guidelines are written by senior academics, they clearly index the institutionalization of qualitative methods. (para. 21)

Concerning the finding that computer science education researchers affiliated with Middle Eastern institutions tended to do explanatory descriptive research, a quick examination of the Middle Eastern institutions from which the Middle Eastern articles came sheds light on this finding. Three Israeli institutions accounted for over half of the Middle Eastern computer science education articles. Those institutions were the Technion – Israel Institute of Technology, the Weizmann Institute of Science, and Tel-Aviv University, which contributed 23.1, 23.1, and 11.5%, respectively, of the total number of Middle Eastern computer science research articles included in this sample.

Another interesting finding was that North American papers had a significantly higher proportion of anecdotal-only papers than other regions, but that this proportion had been declining over time. As Figure 2 shows, in 2000 the proportion of North American anecdotal-only papers was about 80%; in 2005 the proportion was about equal with the proportions of other regions at about 30%. Although we cannot state an informed hypotheses about why the proportion of anecdotal-only North American papers would have been so much higher than in other regions in 2000, we do have a hypothesis about why the proportion of anecdotal-only articles had been declining steadily only in North America, apart from the fact that extreme scores tend to regress towards the mean.

Given that more than one third of the total computer science education research articles came from the SIGCSE conference proceedings (i.e., *Proceedings of the SIGCSE Technical Symposium on Computer Science Education*) (Randolph, 2007), which were held in the United States from 2000 through 2005, one possible explanation is that the decline in North American conference papers is heavily correlated with a decline in anecdotal-only papers in SIGCSE conference proceedings. (In fact, the Spearman correlation of the anecdotal-only percent by year between the SIGCSE conference proceedings and North American articles in general was quite high, $r(6) = .87, p < .02$.) In addition, that decline in the proportion of anecdotal-only SIGCSE conference papers could be a result of the increased interest in the methodological qualities of the articles published in SIGCSE conference proceedings, which is evident in recent SIGCSE conference proceedings articles, such as Valentine (2004), and working group reports, such as Almstrum, Ginat, Hazzan, and Clement (2003) and Almstrum, Ginat, Hazzan, and Petre (2005). One flaw with this hypothesis though is that there has also been a recent interest in the methodological quality of computer science education research articles across the range of computer science publication forums, which is evident in articles such as Almstrum, Ginat, Hazzan, & Morely (2002); Bouvier, Lewandowski, and Scott, (2003); Carbone and Kaasbøll (1998); Clear (2001); Daniels, Petre, and Berglund (1998); Fincher and Petre (2004); Greening 1997); Lister (2005); Pears, Seidman, Eney, Kinnunen, and Malmi (2005); Pears, Daniels, and Berglund (2002); Randolph, Bednarik, and Myller (2005), and Sandström and Daniels (2000), among others.

6.4 Differences between Fields

Tedre (2006) explains that computer science is a field that is mainly comprised of three traditions: a formalist tradition, an engineering tradition, and an empirical tradition. Assuming that the proportion of papers that do not empirically deal with human participants are, more or less, indicators of engineering and/or formalist traditions lingering in computer science education, then we speculate that computer science education is a field in which the traditions of computer science research proper, especially

the engineering tradition, bleed through to the practice of computer science education research itself. Computer science education researchers, as a whole, publish more “I engineered this intervention to certain specifications” types of articles and less “I empirically evaluated the effects of this intervention on student learning” types of articles than their counterparts in educational technology. In turn, educational technologists, as a whole, publish more engineering types of articles and less empirical-evaluation types of articles than their counterparts in educational research proper.

In terms of the proportions of qualitative, quantitative, and mixed-methods research, computer science educators tended to use quantitative methods more frequently and qualitative research less frequently than their counterpart researchers in educational technology or education proper. This might come as a source of concern to the factions of computer science education researchers who call for more qualitative research, such as Ben-Ari, Berglund, Booth, and Holmboe (2004); Berglund, Daniels, and Pears (2006); Hazzan, Dubinsky, Eidelman, Sakhnini, and Teif (2006); and Lister (2003).

7. Summary

In this article, we conducted an empirical review of the recent computer science education research literature to gain insights into its trends, tribes, and territories. We also compared research practices across fields. In terms of trends, we found that there was a decrease in the percentage of both anecdotal-only and explanatory descriptive papers being published each year. In terms of tribes, we found that certain institutions from the U.S., Finland, and Israel tended to have the greatest influence on the recent computer science education literature. Finally, in terms of territories, we found that North American researchers had a tendency to write and get published experimental and anecdotal-only papers; however, the percentage of North American anecdotal-only papers had decreased over the years and is now about equal with the percentage of anecdotal-only papers in other regions. European researchers had a tendency *not* to write or get published experimental/quasi-experimental papers. Middle Eastern researchers had a tendency to write and get published explanatory descriptive papers. Researchers from Asian-Pacific/Eurasian regions tended to investigate attitudes only. Articles from authors affiliated with African or Central/South American institutions made up a very small percentage of the mainstream, computer science education literature. Lastly, it was found that computer science education researchers in general tended to write more papers that do not deal with human participants and tended to do more quantitative research (when empirical research on human participants is done) than educational technology researchers or education researchers proper.

References

- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.
- Almstrum, V. L., Ginat, D., Hazzan, O., & Clement, J. (2003). Transfer to/from computing science education: The case of science education research. In *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education* (pp. 303-304). New York: ACM Press.
- Almstrum, V. L., Ginat, D., Hazzan, O., & Morely, T. (2002). Import and export to/from computing science education. The case of mathematics education research. In *Proceedings of the 7th Annual Conference on Innovation and Technology in Computer Science Education ITiCSE '02* (pp. 193-194). New York: ACM Press.
- Almstrum, V. L., Hazzan, O., Guzdial, M., & Petre, M. (2005). Challenges to computer science education research. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education SIGCSE '05* (pp. 191-192). New York: ACM Press.
- Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. Buckingham, UK: SRHE and Open University Press.
- Ben-Ari, M., Berglund, A., Booth, S., & Holmboe, C. (2004). What do we mean by theoretically sound research in computer science education? In *Proceedings of the 9th Annual Conference on Innovation and Technology in Computer Science Education ITiCSE '04* (pp. 230-231). New York: ACM Press.
- Berglund, A., Daniels, M., & Pears, A. (2006). Qualitative research projects in computing education research. In *Proceedings of the 8th Australian Conference on Computing Education*, 52, (pp. 25-33). Darlinghurst, Australia: Australian Computer Society.
- Bouvier, D., Lewandowski, G., & Scott, T. (2003). Developing a computer science education research program. *Journal of Computing Sciences in Colleges*, 19(1), 218.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687-699.
- Carbone, A., & Kaasbøll, J. (1998). A survey of methods used to evaluation computer science teaching. *SIGCSE Bulletin*, 30(3), 41-45.
- Clear, T. (2001). Thinking issues. Research paradigms and the nature and meaning of truth. *SIGCSE Bulletin*, 33(2), 9-10.
- Daniels, M., Petre, M., & Berglund, A. (1998). Building a rigorous research agenda into changes to teaching. In *Proceedings of the 3rd Australasian Conference on Computer Science Education* (pp. 203-209). New York: ACM Press.
- Depaepe, M. (2002). A comparative history of the educational sciences: The comparability of the incomparable? *European Education Research Journal*, 1(1), 118-122.
- Donaldson, S. I., & Christie, C. A. (2005). The 2004 Claremont Debate: Lipsey vs. Scriven. Determining causality in program evaluation and applied research: Should experimental evidence be the gold standard. *Journal of Multidisciplinary Evaluation*, 3, 60-77.
- European Science Foundation. (2004, October). *Quantitative methods in the social sciences (QMSS): An ESF scientific programme*. Retrieved August 24, 2006, from <http://www.esf.org/publication/193/QMSS.pdf>
- Fielding, N. (2005, May). The resurgence, legitimation and institutionalization of qualitative methods. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), Article 32. Retrieved September 25, 2006, from <http://www.qualitative-research.net/fqs-texte/2-05/05-2-32-e.htm>
- Fincher, S., & Petre, M. (Eds.). (2004). *Computer science education research*. London: Taylor & Francis.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory*. New York: Aldine.
- Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research*. Berkshire, England: Open University Press.
- Greening, T. (1997). Paradigms for educational research in computer science. In *Proceedings of the 2nd Australasian Conference on Computer Science Education* (pp. 47-51).
- Hazzan, O., Dubinsky, Y., Eidelman, L., Sakhnini, V., & Teif, M. (2006). Qualitative research in computer science education. In *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education* (pp. 408-412). New York: ACM Press.
- Julnes, G., & Rog, D. J. (Eds.). (in press). Informing federal policies on evaluation methodology: Building the evidence for method choice in government sponsored evaluation. *New Directions for Evaluation*.
- Lawrenz, F., & Huffman, D. (2006). Methodological pluralism: The gold standard of STEM evaluation. In D. Huffman & F. Lawrenz (Eds.), *New directions for evaluation: Critical issues in STEM evaluation* (pp. 19-34). Hoboken, NJ: Wiley.
- Lister, R. (2003). A research manifesto, and the relevance of phenomenography. *ACM SIGCSE Bulletin*, 35(2), 15-16.
- Lister, R. (2005). CS research: Mixed methods: Positivists are from Mars, constructivists are from Venus. *SIGCSE Bulletin*, 37(4), 18-19.
- Neuendorf, K. A. (2002). *The content analysis handbook*. Thousand Oaks, CA: Sage.
- Pears, A., Daniels, M., & Berglund, A. (2002). Describing computer science education research. An academic process view. In *Proceedings of the Conference on Simulation and Multimedia in Engineering Education, ICSEE 2002* (pp. 99-104).
- Pears, A., Seidman, S., Eney, C., Kinnunen, P., & Malmi, L. (2005). ITiCSE 2005 working group reports: Constructing a core literature for computing education research. *SIGCSE Bulletin*, 37(4), 152-161.
- Randolph, J. J. (2007). *Computer science education research at the crossroads: A methodological review of computer science education research: 2000-2005*. Dissertation, Utah State University, Utah. Retrieved May

- 15, 2009, from http://www.archive.org/details/randolph_dissertation
- Randolph, J. J., Bednarik, R., & Myller, N. (2005). A methodological review of the articles published in the proceedings of Koli Calling 2001-2004. In T. Salakoski, T. Mäntylä, & M. Laakso (Eds.), *Proceedings of the 5th Annual Finnish / Baltic Sea Conference on Computer Science Education* (pp. 103-109). Finland: Helsinki University of Technology Press. Retrieved March 19, 2006, from http://www.it.utu.fi/koli05/proceedings/final_composition_b5.060207.pdf
- Rautopuro, J., & Väisänen, P. (2005). DEEP WATER? Quantitative research methods in educational science in Finland. In M-L Julkunen (Ed.), *Learning and instruction in multiple contexts and settings III. Proceedings of the Fifth Joensuu Symposium on Learning and Instruction. Bulletins of the Faculty of Education, 96*, (pp. 273-293). Finland: University of Joensuu Press.
- Sandström, A., & Daniels, M. (2000). Time studies as a tool for (computer science) education research. In *Proceedings of the Australasian Conference on Computing Education* (pp. 208-214). New York: ACM Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Stevens, J. (1999). *Intermediate statistics: A modern approach* (2nd ed.). Mawwah, NJ: Earlbaum.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Newbury Park, CA: Sage.
- Tedre, M. (2006). *The development of computer science: A sociocultural perspective*. Dissertation, University of Joensuu. Retrieved October 24, 2006, from http://joypub.joensuu.fi/publications/dissertations/tedre_development/index_en.html
- Turner, B. (1981). Some practical aspects of qualitative data analysis. *Quality and Quantity, 15*, 225-247.
- U.S. Department of Education. (2002). *Press release: New directions for program evaluation at the U.S. Department of Education*. Retrieved August 24, 2006, from <http://www.ed.gov/news/pressreleases/2002/04/evaluation.html>
- Valentine, D. W. (2004). CS educational research: A meta-analysis of SIGCSE technical symposium proceedings. In *Proceedings of the 35th Technical Symposium on Computer Science Education* (pp. 255-259). New York: ACM Press.
- van Helden, J., Johnson, Å., & Vakkuri, J. (2006, June). *Exploring the USA-Europe divide in public section performance measurement research*. Paper presented at a performing public sector: The second transatlantic dialogue, Workshop 5: Emerging and other strategies for productivity and performance, Leuven, Belgium.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.) Thousand Oaks, CA: Sage.

Acknowledgements

This research was supported by a generous special projects grant from SIGCSE.

Author Information

Justus J. Randolph, Ph.D.

Faculty
K-12 Educational Leadership
College of Education
Walden University

Assistant Professor
Elementary Education Department
Finlandia University

justus@randolph.name
<http://justus.randolph.name>

George D. Julnes, Ph.D.

Associate Professor
School of Public Affairs
University of Baltimore

Erkki Sutinen, Ph.D.

Professor, Department Head
Department of Computer Science and Statistics
University of Joensuu