# CS Educational Research: A Meta-Analysis of SIGCSE Technical Symposium Proceedings

Dr. David W. Valentine
Department of Computer Science
Slippery Rock University
Slippery Rock, PA
david.valentine@sru.edu

## Abstract

A meta-analysis is performed on the last twenty years of SIGCSE Technical Symposium Proceedings, looking for the kind of CS Educational Research that has been done at the CS1/CS2 level. A six-point taxonomy of articles types is described. It is shown that about one in five of all CS1/CS2 presentations have used some kind of experimental model, albeit "experimental" is defined quite broadly. Over the last ten years both the number of experimental models, and the percentage of experimental models among CS1/CS2 has significantly increased. SIGCSE members are challenged to adopt a research model for their presentations to the Technical Symposium.

## Categories and Subject Descriptors

K.3.2 [COMPUTERS AND EDUCATION]: Computer and Information Science Education --- *computer science education*

## General Terms

Experimentation, Measurement

## Keywords

CS Education Research, CS1/CS2.

## Introduction

In her farewell column in the SIGCSE Bulletin last December [2002], Nell Dale did an informal survey of our conference Proceedings and (with justifiable pride) reported a "definite increase in CS Ed Research related papers" over the last decade, and concluded "CS Ed Research has arrived in the main stream where it belongs."[1] She mentioned some two dozen papers, panels and workshops that have been presented since 1994. I suspected that there were, in fact, many more than that. So I set out to do a more formal survey of the SIGCSE Technical Symposium Proceedings (TSP) to discover the current state of, and the longitudinal development of CS Educational Research among our members.



## The Method

I limited the survey to the last twenty years of the TSP. Our field moves so quickly that reading articles from even 20 years ago can seem "quaint": one finds text-based 'animation' tools for Turbo Pascal, or debates about 'training vs. education' in college, etc. Yet two decades should give a good look at any trends that are developing. Fortunately, the ACM Portal has all 34 of the TSP on-line. Since 1974 the vast majority of articles are available in full text PDF, along with abstracts, citations and index terms.

Furthermore, I limited my search to presentations dealing with the First Year of CS courses: the traditional CS1/CS2 domain[2]. First Year topics are an important part of our discussion. Indeed some people have said that teaching the First Year is the hardest part of teaching the CS Curriculum. It is one of the seven "Topics of Interest" in the current Call for Papers, is the first of six "Pedagogy Focus Groups" (PFG) of Curricula 2001, and has always been a prominent part of the TSP. This gave me a particular domain to track for the survey.

The first step was to find presentations dealing with the First Year. Following Nell Dale's preliminary survey, I included papers, workshops and panels as presentations. The ACM Portal does not list the "Session Titles", so I could not rely upon that organization. Many authors put "CS1" or "CS2" in their titles, but others were more cryptic, which meant reviewing the full text of a great many articles. If the presentation dealt with a traditional CS1/CS2 topic (from core language debate to teaching recursion, simple data types to grading introductory programs), then it was counted. The topic of "algorithms" was often difficult: the term is used for the normal CS2 Big-Oh type of thing as well as the more advanced Analysis of Algorithms. If the majority of the algorithms discussed in the presentation were traditional CS2 topics (sorts, tree traversal, backtracking, etc.), then it was included.

## General Observations

To state the obvious, the TSP is growing (Figure 1), and it is growing at a statistically significant rate (linear regression $F=25.96$, $p<0.001$). The Proceedings of the last decade (1994-2003) have an average number of presentations that is 39% higher than the average of the previous decade (1984-1993). There has also been in increase both in the number of First Year presentations, and the percentage of presentations dealing with the first year. The last decade saw the average number of First Year presentations rise from 25% to 30% of all citations in the Proceedings. This would indicate that our membership is investing more energy into improving the First Year of CS instruction.

Figure 1: Total Number of TSP Presentations

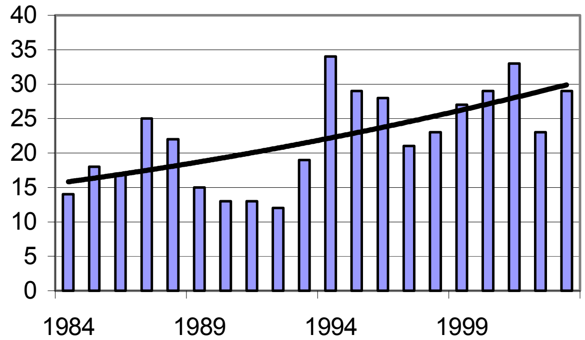



Table 1: Comparing Last Two Decades of TSP

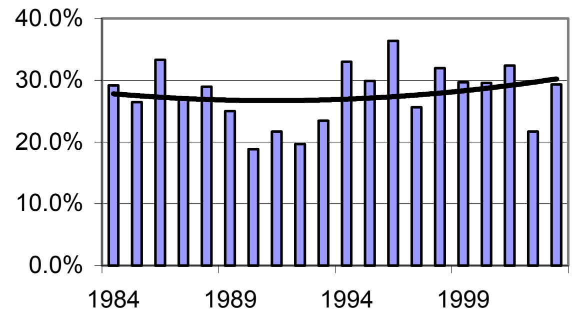
| SIGCSE Technical Symposium Proceedings (TSP) | 1984-1993 | 1994-2003 |
|---|---|---|
| Average Number of Presentations | 67 | 93 |
| Average Number of Presentations dealing with First Year | 17 | 28 |
| Percentage of Presentations dealing with First Year | 25% | 30% |

Two other general observations are in order. Concerning the debate over language (C++/Java), it is worth noting that the language issue is apparently an eternal religious war for SIGCSE. Writing some fifteen years ago, Luker [1989] said, "For those of us who have been teaching Computer Science for nearly twenty years, there is a feeling of deja vu about debates on the first programming language for our majors." And Luker was commenting on the Ada/Modula2 debate, and comparing it to the earlier debate in England over Algol/Fortran!

The second comment would be SIGCSE's concern over attracting and retaining women in Computer Science. The earliest presentation (in my twenty year window) was Wieckert & Dale [1984], "Women in Science and Academe." The issue has appeared in roughly half of the TSP in the last twenty years. That also means that half the time we did not have a single article on the topic.

Looking at First Year presentations as a percentage of all TSP citations, the trendline (which was not significantly significant in a nonlinear regression) shows a shallow parabolic shape with a minimum in the very early 90's. I suspect that the SIGCSE membership was pretty well at ease with the aging Curriculum '78 by then and difficulties in the First Year had been addressed. Then Curriculum '91 "changed the rules" and our members responded with a resurgence of interest in teaching the First Year with the new curriculum.

Figure 2: First Year Presentations as a Percentage of all TSP




## The Taxonomy

As I reviewed the presentations, a six-fold taxonomy to classify the type of articles at TSP emerged. Because I was interested in seeing the Educational Research done for the First Year, that was an obvious first category. A purist might require that there be an explicit research question, conveyed in a series of hypotheses, tested with a variety of experimental and control groups, with a strict statistical analysis of results. But I don't think you need such a strictly quantified, statistical model to prove significant educational results. I wanted to set as inclusive (and yet reasonable) a bar as possible for this category, and I settled on a very simple rubric: if the author made any attempt at assessing the "treatment" with some scientific analysis, I counted it as an "**Experimental**" presentation. For a minimal example, Bagert, et. al. [1995] showed that after a New Breadth-First CS1 course, the number of CS majors earning a 'C' or better in CS2 doubled at Texas Tech. At the other end of the category, Dey & Mand [1986] did a complete statistical analysis of 500 introductory students at two institutions to show the impact of math background and prior programming to success in CS1. Another, less quantitative example (because not all knowledge is quantifiable) would be Fleury [1991] who, through a series of interviews, developed an ethnography of how students develop their own (often faulty) cognitive rules about parameter passing. Clancy & Linn [1999] in a philosophical discussion of pedagogy did a review of existing research literature, so they were also included here. Please note that this was a preemptive category, so if the presentation fit here and somewhere else (e.g. a quantified assessment of some new Tool), it was placed here.

The second category is what has been called by others "**Marco Polo**" presentations: "I went there and I saw this." SIGCSE veterans recognize this as a staple at the Symposium. Colleagues describe how their institution has tried a new curriculum, adopted a new language or put up a new course. The reasoning is defined, the component parts are explained, and then (and this is the give-away for this category) a conclusion is drawn like "Overall, I believe the [topic] has been a big success." or "Students seemed to really enjoy the new [topic]". Now, Marco Polo presentations serve an important function: we are a community of educators and sharing our successes (and failures) enriches the whole community. Yet, it seems that with just a little more effort at some educational assessment, we could wring a great deal more benefit from the exercise.

A third classification would be "**Philosophy**" where the author has made an attempt to generate debate of an issue, on philosophical grounds, among the broader community. An example here would be Reed, et. al [2002] who put up a panel discussion on "Integrating Empirical Methods into CS", and said, "This panel is designed to promote discussion …within the traditional computer science community." Or we would include an article like McCraken [1992] who tried to stimulate the core language debate along philosophical and educational lines. Of course the "Denning Report" [1988] on "Computing as a Discipline" was a foundational work that still guides our philosophical understanding to the present day. It is a matter of fact that the Task Force on the Core of Computer Science made its report to SIGCSE before the Communications of the ACM published the results. I did not count the Denning Report, because it wasn't primarily about CS1/CS2, but I mention it because of the overwhelming number of times it is referenced by articles that did fit our study.

Next there was a large collection of presentations that I classified "**Tools**". Among many other things, colleagues have developed software to animate algorithms, to help grade student programs, to teach recursion, and to provide introductory development platforms. For example, Studer et. al [1995] developed a tool so novice programmers could use pictograms rather than syntax to create programs. Rambally [1985] built a tool to graphically represent linked data structures for students. Not all tools were software; an author could present a paradigm or an organizing rubric to be a tool for an entire course. Carrasquel et. al. [1989] presented a combination of a visual design tree and data flow diagrams as an effective teaching tool for CS1.

The most whimsical category would be called "**Nifty**", taken from the panels that are now a fixed feature of the TSP. Nifty assignments, projects, puzzles, games and paradigms are the bubbles in the champagne of SIGCSE. Most of us seem to appreciate innovative, interesting ways to teach students our abstract concepts. Sometimes the difference between Nifty and Tools was fuzzy, but generally a Tool would be used over the course of a semester, and a Nifty assignment was more limited in duration. Ginat [1995] related loop invariants to mathematical games. Fell and Proulx [1997] showed how to use Martian planetary images in CS1. Cigas [1992], in a real gem, shows how to use finite state automata in traditional CS1/CS2 problems to improve student success.

The last, and (happily) the smallest category of presentations would be "**John Henry**" papers. Every now and then a colleague will describe a course that seems so outrageously difficult (in my opinion), that one suspects it is telling us more about the author than it is about the pedagogy of the class. To give a silly example, I suppose you could teach CS1 as a predicate logic course in IBM 360 assembler – but why would you want to do that? Yes, every once in a while somebody can beat the steam engine, but most of us try to avoid that situation. Yet John Henry's are valuable to our community, too. We should continually be touching that upper limit of our pedagogy (which means occasionally we'll push over the line).

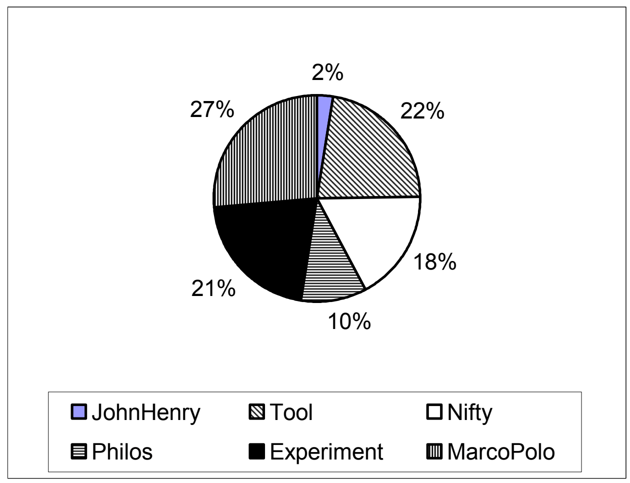
## The Results

Table 2 shows the total number of First Year articles in each category for the last twenty years. Figure 3 shows the breakdown of First Year presentations over the last 20 years using the 6-point taxonomy. Half of all TSP presentations dealing with the First Year of instruction have fallen into the areas of "Marco Polo" (27%) and Tools (22%). This was to be expected. What may seem surprising is that the third most common form of presentation was some form of Experiment (21%). Again, we cast a very broad net for this category. But even so, this seems like good news.

**Table 2: Count of First Year Presentations**

| | Number of First Year Articles 1984-2003 |
|---|---|
| Marco Polo | 117 |
| Tools | 99 |
| Experimental | 94 |
| Nifty | 78 |
| Philosophy | 45 |
| John Henry | 11 |
| **Total of All First Year** | **444** |

**Figure 3: Distribution of First Year Presentations**



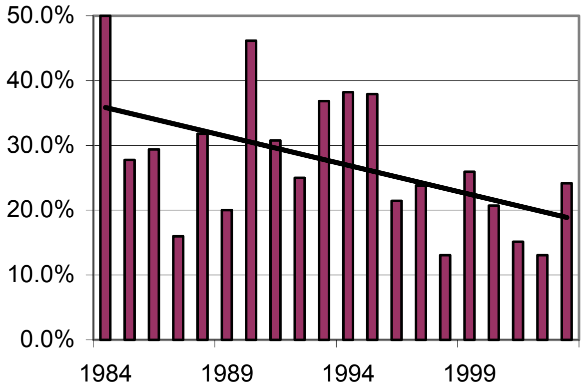

Now if we compare the two decades (84-93, 94-03), we see that four of the categories have remained quite stable (Table 3). The major difference is about a 6% shift from Marco Polo presentations to Tools. So these two still account for half of all First Year presentations. The stability of the Experimental group is encouraging: about one in five First Year presentations included at least some kind of quantitative assessment in each of the last two decades.

**Table 3: Comparison of First Year over Two Decades**

| | 1984-1993 | 1994-2003 |
|---|---|---|
| Marco Polo | 30.4% | 23.9% |
| Tools | 18.5% | 24.6% |
| Experimental | 19.6% | 22.1% |
| Nifty | 17.3% | 17.8% |
| Philosophy | 11.9% | 9.1% |
| John Henry | 2.4% | 2.5% |

I believe that the "We tried this and we think it is good" kind of paper (Marco Polo) is the antithesis of the experimental presentation. Figure 4 shows encouraging news in that regard. The percentage of First Year presentations in this category is showing a clear falling off (linear regression F=6.15, p<0.025). The decline would be even more dramatic but for the rebound of Marco Polo articles in the wake of Curriculum 91's release (see Figure 5). That Curriculum, with its "knowledge units" rather than specific course outlines, understandably generated a great deal of "this is how we try to implement it at our school" type articles. A quick look at the 1994 TSP, for example, sees significant numbers of presentations on OOPS, on Closed Labs, and on Breadth First models. There was a lot that our members were trying to adapt and bring into the First Year at that time; Marco Polo type presentations are certainly understandable given that situation. But it seems we are moving beyond that now.

**Figure 4: Percentage of Marco Polo among First Year**




Over the same period, Figure 6 shows the percentage of First Year articles done on the line of an Experiment. The parabolic shape is quite interesting (although the quadratic coefficient of nonlinear regression was not statistically significant). Articles done along experimental lines clearly fell off between 1989 and 1994 (with a small spike in 1991). This closely tracks a corresponding spike increase in Marco Polo presentations. Causes for this could be the large increase in SIGCSE membership, where the newcomers did not appreciate the traditional interest in the area (or plain ignorance of educational research). It could also be a swamping out of research as we tried to cope with Curriculum '91, OOPS, Breadth First and Closed Labs in CS1/CS2. The cause is not clear, and there are probably a variety of reasons.

Yet it is undeniable that the Experimental model, as a percentage of all First Year articles, is on the rise over the last five or six years. If we look at a linear regression analysis of the last decade (the domain of Nell Dale's informal survey), there is a clear, statistically significant increase in both the number of First Year Experimental papers (F=5.42, p=0.03) and in the percentage of First Year presentations of the Experimental type (F=7.96, p<0.025).

Figure 6 shows perhaps the most surprising result of this study: the high percentage of experimental papers in the late 1980's. Were we doing more experimental papers early on, and drifted away from the experimental model?

**Figure 5: Total Number of Marco Polo per Year**

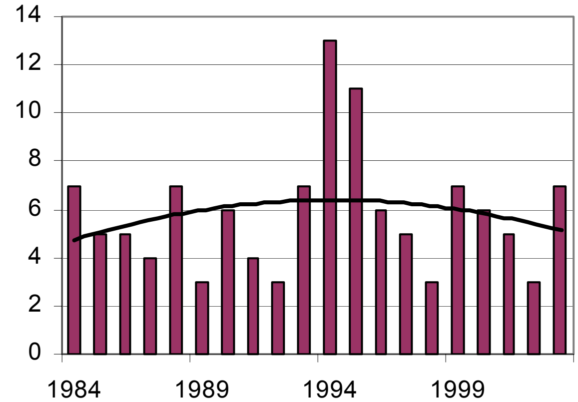
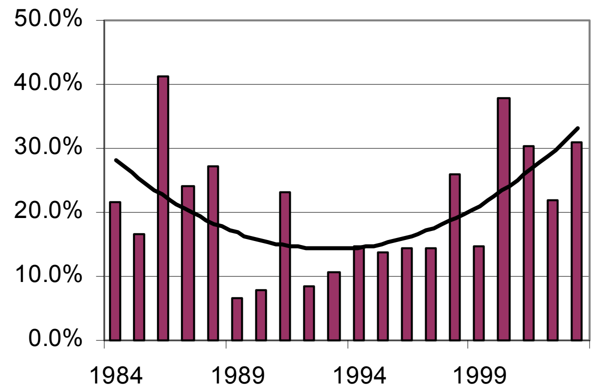



**Figure 6: Percentage of Experimental Among First Year**




## What have we "proven" at SIGCSE?

Because this was a longitudinal meta-analysis of the SIGCSE TSP, I thought it would be illuminating to see the kinds of results colleagues have "proven" in the sense of reporting an experimental study with significant results. My goal was to report a breadth of results, both in time and in domain, and no slight is intended toward experimental studies not listed here. We, as a community of educators, have experimentally explored the whole domain of teaching computer science. From the nearly 100 experiments reported to TSP over the last twenty years, here are a dozen presentations that represent a cross-section of the kinds of things SIGCSE had done at the CS1/CS2 level.

- Campbell [1984] proved that at-risk students did significantly better in CS1 if they took a specially-designed preliminary course first.
- Sidbury [1986] proved that students did better in the first year of CS if they took a discrete math sequence over an algebra-calculus math sequence.
- Austin [1987] reported his doctoral work on predicting student success in CS1 at the community college level.

- Harrington [1988] did a longitudinal study of incoming freshmen and their changing pre-college computer experience.
- Lang and Smith [1993] proved that a CS1 model using a closed lab and quizzes (based on previous programs) improved student performance in CS2.
- Moore, et.al. [1994] developed and validated an instrument to measure student attitude toward the discipline of computer science.
- Sabin and Sabin [1994] proved that a collaborative learning experience increased student learning and improved student attitude toward the course.
- Bagert, et.al. [1995] proved in a five-year study that breadth first plus closed labs significantly improved the retention and passing rates for the next course, as well as improved overall graduation rates in CS.
- Gunsher-Sackrowitz and Parker [1996] proved that women come to college having less computer experience and that statistically puts them at a disadvantage in CS.
- Wu, et.al. [1998] proved the relationship between conceptual models, cognitive learning styles, and student success in learning recursion.
- Chamillard and Braun [2000] provided a statistical comparison of various ways to evaluate student programming ability in CS1.
- Fagin and Merkle [2003] reported (honestly) that a year-long experiment in using robots in CS1 actually lowered student test scores significantly.

## The Challenge

I believe that if CS Educational Research is to grow at TSP, it will come at the "expense" of Marco Polo and Tools types of articles. I want to both challenge and encourage the SIGCSE membership to push their presentations this next step: it isn't really all that difficult. Many of us are caught up in the Assessment movement, and this discussion is fundamentally the same issue: prove that you did what you said that you did! Since Tools is the fastest growing segment of the TSP First Year presentations, let me take one example from the last TSP to share.

Oderick-Hash and Zachary [2003] at the University of Utah designed and built "InSTEP, an online tutoring system for beginning C programmers", and used it in their CS1 courses. They could have presented a Tool/Marco Polo type of presentation to TSP, but they went to the next level. They divided the audience into experimental and control groups, did a pre-test and post-test on one simple lab exercise where the experimental group used their Tool, and measured (i) correctness, (ii) time on task, and (iii) time seeking help from TA's as variables. Compared with the labor put into creating InSTEP, I doubt this 'experiment' was a significant additional burden.

They had hoped that InSTEP would provide improved learning, but the data didn't support that conclusion (as a Marco Polo article they easily could've said, 'We think the students learned more.'). To paraphrase T.H. Huxley, a beautiful hypothesis was murdered by an ugly fact. What the data did support was equivalent learning with less time seeking TA help. That is a more modest claim: but it is verifiable! With confidence, they conclude that InSTEP "would be useful to students who did not have a TA available", and that using InSTEP would free up TA's to help students with more serious needs.

Oderick-Hask and Zachary transformed what could have been a very commonplace Marco Polo/Tools type presentation into a viable Experimental article. The effort involved (I assume) was negligible when compared to the overall size of the task they set for themselves. Their conclusions are more modest than they might have been able to claim otherwise. But those conclusions are based in defensible research, and not mere assumptions. We need more of this in SIGCSE. I challenge the creators of CS1/CS2 Tools, in particular, to step up and prove to us that your Tool actually does what you are claiming that it does. Do the fundamental research necessary to rest your claims upon defensible fact.

## Concluding Remarks

This study was a very rewarding exercise personally. Many of the names leapt out as old friends either because their writings have challenged me (on both sides of the teacher's desk) or because they had stimulated my thinking in person at the Technical Symposium. I am pleased to report there is now a rather large stack of article reprints on my own desk because I kept coming across a paper that I knew had material I wanted to bring into the classroom. The SIGCSE membership has promoted the cause of CS Education faithfully and energetically for more than twenty years, and there is a deep wealth of material that will improve our teaching today. Finally, it was reaffirming to trace the maturity and development of our SIG: the number of and the quality of presentations (experimental or otherwise) has steadily risen over the last two decades. Nell Dale was right to conclude that CS Educational Research has arrived (and now we have proven it!).

---

[1] All references to the SIGCSE Technical Symposium Proceedings will be given inline using the author's name and year of the Proceedings.

[2] Of course, Curricula 2001 has this sequence as CS101I/CS102I/CS103I || CS111I/CS112I || CS101O/CS102O/CS103O || CS111O/CS112O || CS111F/CS112F || CS100B/CS101B/CS102B/CS103B || CS111A/CS112A || CS111H/CS112H. No wonder we still use "CS1/CS2" at the Symposium!