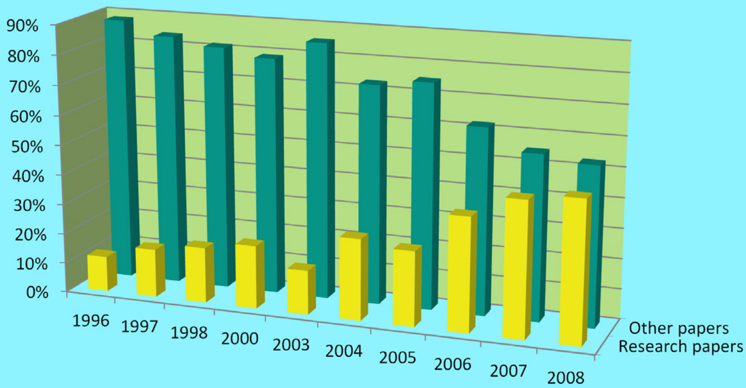
Department of Computer Science

Emergence of computing education as a research discipline

Simon



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 150/2015

Emergence of computing education as a research discipline

Simon

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall U5 of the school on 18 November 2015 at 12 noon.

Aalto University
School of Science
Department of Computer Science
Learning + Technology Group (LeTech)

Supervising professor
Professor Lauri Malmi, Aalto University

Thesis advisor
Professor Lauri Malmi, Aalto University

Preliminary examiners
Associate Professor Mike Joy, University of Warwick, UK
Professor Josh Tenenberg, University of Washington, Tacoma, USA

Opponent
Associate Professor Arnold Pears, Uppsala University, Sweden

Aalto University publication series
DOCTORAL DISSERTATIONS 150/2015

ISBN 978-952-60-6415-4 (printed)
ISBN 978-952-60-6416-1 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6416-1

Unigrafia Oy
Helsinki 2015

Finland

441 697
Printed matter

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Simon

Name of the doctoral dissertation
Emergence of computing education as a research discipline

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 150/2015

Field of research Computing education

Manuscript submitted 17 August 2015 **Date of the defence** 18 November 2015

Permission to publish granted (date) 21 September 2015 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

This thesis investigates the changing nature and status of computing education research (CER) over a number of years, specifically addressing the question of whether computing education can legitimately be considered a research discipline.

The principal approach to addressing this question is an examination of the published literature in computing education conferences and journals. A classification system was devised for this literature, one goal of the system being to clearly identify some publications as research – once a suitable definition of research was established. When the system is applied to a corpus of publications, it becomes possible to determine the proportion of those publications that are classified as research, and thence to detect trends over time and similarities and differences between publication venues.

The classification system has been applied to all of the papers over several years in a number of major computing education conferences and journals. Much of the classification was done by the author alone, and the remainder by a team that he formed in order to assess the inter-rater reliability of the classification system.

This classification work led to two subsequent projects, led by Associate Professor Judy Sheard and Professor Lauri Malmi, that devised and applied further classification systems to examine the research approaches and methods used in the work reported in computing education publications.

Classification of nearly 2000 publications over ranges of 3-10 years uncovers both strong similarities and distinct differences between publication venues. It also establishes clear evidence of a substantial growth in the proportion of research papers over the years in question.

These findings are considered in the light of published perspectives on what constitutes a discipline of research, and lead to a confident assertion that computing education can now rightly be considered a discipline of research.

Keywords computing education research, publications, journals, conferences, empirical research, Simon's system

ISBN (printed) 978-952-60-6415-4 **ISBN (pdf)** 978-952-60-6416-1

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2015

Pages 220 **urn** http://urn.fi/URN:ISBN:978-952-60-6416-1

Acknowledgements

I am deeply indebted to my supervisor, Professor Lauri Malmi, for recognising that some of my past research was worthy of a doctorate, for charting a suitable course through the university's rules and regulations, and for helping to ensure that I completed the work in a reasonable timeframe. Professor Malmi has accompanied me on the path of this research almost since its inception, and it is an honour to complete it under his supervision.

I am also grateful to Associate Professor Raymond Lister. Professor Lister was part of a team that introduced me to computing education research through the BRACE project. As a guest editor of an Australasian special edition of the journal Computer Science Education he encouraged my development of the classification scheme that is central to the research presented here. When I sought to move beyond simple classification, he introduced me to the pivotal work of Peter Fensham. And he took part in the AWPOCCELIPS project, which carried out some of the classification and which permitted the measurements of inter-rater reliability. This journey is unlikely to have taken place without Professor Lister's assistance and guidance. Professor Lister's work is cited in several parts of this thesis, as he has been one of the principal drivers of the emergence of computing education as a discipline of research.

Associate Professor Judy Sheard has been an almost constant companion on the journey. She, too, was part of the AWPOCCELIPS team. Moving on from the classification carried out by that team, she formed a new team to look into the research methods of the papers that were classified by my system as research papers. She then joined with Professor Malmi to develop a more comprehensive system to analyse the theories, methods, and frameworks evident in research papers. I was fortunate to be able to take part in both of these projects.

I thank Professor Valentina Dagienė for urging me to apply my classification system outside the realm of computing education, in particular to the publications in the journal Olympiads in Informatics. This proved to be an excellent test of the system's robustness.

I am grateful for the encouragement and assistance of my co-authors in various phases of this work. In addition to those named above, these are Ahmad Taherkhani, Angela Carbone, Ari Korhonen, Errol Thompson, Jan Lönnberg, Juha Helminen, Juha Sorva, Margaret Hamilton, Michael de Raadt, Niko Myller, and Roman Bednarik.

Many other members of the computing education research community have joined me in helpful discussion and have urged me to undertake a doctorate. Names that spring to mind are Tony Clear, Anders Berglund, and Arnold Pears; and, of course, Sally Fincher, for setting me on the road of this research by snorting (see section 2.1); however, there are so many more that I cannot thank them all individually. I hope for their understanding.

I am grateful to the Board of SIGCSE for the Special Projects grant that facilitated the expansion of the project from a single researcher to a team.

The pre-examiners of this thesis, Professor Josh Tenenberg and Associate Professor Mike Joy, have provided carefully considered suggestions to improve the work. I thank them. And while at the time of writing I have not yet defended my work, I am confident that with Associate Professor Arnold Pears as my distinguished opponent, the defence will be a memorable event.

I am grateful to my dear wife, Susan Snowdon, for her occasional assistance with statistics and formulas; and I am forever indebted to Susan and to our children, Alexis, Benjamin, and Calum, for ungrudgingly sparing me the time to undertake this and all of my other research.

Helsinki, 1 October 2015
Simon

Contents

List of abbreviations

ACE	Australasian Computing Education conference, held annually in Australia or New Zealand
ACM	Once called the Association for Computing Machinery, now called simply ACM: an association of computing practitioners, based in the USA but with members around the world
AWPOCCELIPS	Australasian Working Party on Classifying Computing Education Literature in Published Sources; a group of researchers who met to undertake joint classification of papers according to Simon's system
CER	Computing Education Research, also known as Computer Science Education Research or Informatics Education Research
CS	Computer Science
CSEd	Computer Science Education, also called Computing Education
ICER	International Computing Education Research conference/ workshop, an annual computing education research conference of SIGCSE
IEEE	Institute of Electrical and Electronics Engineers; based in the USA but with members around the world
ITiCSE	Annual conference on Innovation and Technology in Computer Science Education, held by SIGCSE outside the USA, generally in Europe
JERIC	Journal on Educational Resources in Computing; see TOCE
Koli Calling	Koli Calling International Conference on Computing Education Research (formerly known by other, similar, names)
NACCQ	National Advisory Committee on Computing Qualifications, in New Zealand; and the annual conference of that committee; both the committee and the conference are now known as CITRENZ (Computing and Information Technology Research and Education New Zealand)

SIGCSE	The ACM’s Special Interest Group on Computer Science Education; *also* the annual US technical symposium of that group
TOCE	ACM Transactions on Computing Education; formerly JERIC
TMMCER	Theories, Methods, and Models in Computing Education Research; the classification project co-led by Lauri Malmi and Judy Sheard

List of publications

This thesis consists of a summary and of the following publications, which are referred to in the text by their numerals.

P1. Simon. 2007. A classification of recent Australasian computing education publications. Taylor & Francis, UK. *Computer Science Education*, volume 17, issue 3, pages 155-169. ISSN 0899-3408 (print)/ISSN 1744-5175 (online). DOI 10.1080/08993400701538021.

P2. Simon. 2008. Koli Calling comes of age: an analysis. In: Raymond Lister, Simon (eds). *Seventh Baltic Sea Conference on Computing Education Research (Koli Calling 2007)*, Koli, Finland, November 2007. Australian Computer Society. CRPIT, Australia, volume 88, pages 119-126. ISSN 1445-1336. ISBN 978-1-920682-69-9.

P3. Simon, Judy Sheard, Angela Carbone, Michael de Raadt, Margaret Hamilton, Raymond Lister, Errol Thompson. 2008. Eight years of computing education papers at NACCQ. In: Samuel Mann, Mike Lopez (eds). *21st Annual Conference of the National Advisory Committee on Computing Qualifications*, Auckland, New Zealand, July 2008. NACCQ, New Zealand, pages 101-107. ISSN 1176-8053.

P4. Simon, Angela Carbone, Michael de Raadt, Raymond Lister, Margaret Hamilton, Judy Sheard. 2008. Classifying computing education papers: process and results. In: Raymond Lister, Michael Caspersen, Mike Clancy (eds). *Fourth International Computing Education Research Workshop (ICER 2008)*, Sydney, Australia, September 2008. ACM, USA, pages 161-171. ISBN 978-1-60558-216-0. DOI 10.1145/1404520.1404536.

P5. Simon. 2009. Ten years of the Australasian Computing Education Conference. In: Margaret Hamilton, Tony Clear (eds). *Eleventh Australasian Computing Education Conference (ACE 2009)*, Wellington, New Zealand, January 2009. Australian Computer Society. CRPIT, Australia, volume 95, pages 157-163. ISSN 1445-1336. ISBN 978-1-920682-76-7.

P6. Simon. 2009. Informatics in Education and Koli Calling: a comparative analysis. Institute of Mathematics and Informatics, Lithuania. *Informatics in Education*, volume 8, number 1, pages 101-114. ISSN 1648-5831.

P7. Judy Sheard, Simon, Margaret Hamilton, Jan Lönnberg. 2009. Analysis of research into the teaching and learning of programming. In: Mike Clancy, Michael Caspersen, Raymond Lister (eds). *Fifth International Computing Education Research Workshop (ICER 2009)*, Berkeley, USA, August 2009. ACM, USA, pages 93-104. ISBN 978-1-60558-615-1. DOI 10.1145/1584322.1584334.

P8. Lauri Malmi, Judy Sheard, Simon, Roman Bednarik, Juha Helminen, Ari Korhonen, Niko Myller, Juha Sorva, Ahmad Taherkhani. 2010. Characterizing research in computing education: a preliminary analysis of the literature. In: Michael Caspersen, Mike Clancy, Kathryn Sanders (eds). *Sixth International Computing Education Research Workshop (ICER 2010)*, Aarhus, Denmark, August 2010. ACM, USA, pages 3-11. ISBN 978-1-4503-0257-9. DOI 10.1145/1839594.1839597.

P9. Simon. 2013. Olympiads in Informatics – the journal's first six years. Institute of Mathematics and Informatics, Lithuania. *Olympiads in Informatics*, volume 7, pages 113-122. ISSN 1822-7732.

P10. Lauri Malmi, Judy Sheard, Simon, Roman Bednarik, Juha Helminen, Ari Korhonen, Niko Myller, Juha Sorva, Ahmad Taherkhani. 2014. Theoretical underpinnings of computing education research – what is the evidence? In: Quintin Cutts, Beth Simon, Brian Dorn (eds). *Tenth International Computing Education Research Workshop (ICER 2014)*, Glasgow, Scotland, August 2014. ACM, USA, pages 27-34. ISBN 978-1-4503-2755-8. DOI 10.1145/2632320.2632358.

Author’s contribution

Publication P1: A classification of recent Australasian computing education publications

I was the sole author of this paper.

Publication P2: Koli Calling comes of age: an analysis

I was the sole author of this paper.

Publication P3: Eight years of computing education papers at NACCQ

I obtained a grant for a workshop to introduce the classification scheme to other researchers; led the workshop; drove the classification process; conducted the inter-rater reliability tests; analysed the data; and did most of the writing.

Publication P4: Classifying computing education papers: process and results

I drove the classification process; led the Delphi process that was used to achieve consensus in classifications; analysed the data; and did most of the writing.

Publication P5: Ten years of the Australasian Computing Education Conference

I was the sole author of this paper.

Publication P6: Informatics in Education and Koli Calling: a comparative analysis

I was the sole author of this paper.

Publication P7: Analysis of research into the teaching and learning of programming

This work was led by Judy Sheard as a follow-on from the work carried out for Publications 3 and 4 and other unpublished classification. Associate Professor Sheard had the initial idea and did most of the analysis. All four authors carried out the classification. I did most of the writing.

Publication P8: Characterizing research in computing education: a preliminary analysis of the literature

This work arose from a collaboration between Lauri Malmi and Judy Sheard. Together they formed the ideas, then introduced them to the other team members at a workshop. I took equal part in the classification, and conducted the inter-rater reliability tests. I did not lead the writing process, but took a more than proportional share.

Publication P9: Olympiads in Informatics – the journal's first six years

I was the sole author of this paper.

Publication P10: Theoretical underpinnings of computing education research – what is the evidence?

My share in this paper is about the same as for Publication 8, except that I took a substantially larger share of the writing.

1. Publications as a window into a research discipline

1.1 The emergence of disciplines

By its nature, the exploration of knowledge becomes steadily more specialised. As natural philosophy developed into modern science, it separated into mathematics and the natural sciences such as physics and astronomy, and later into the social sciences. As a discipline develops, it undergoes a gradual separation from its parent, until eventually becoming recognised as a discipline in its own right. However, that recognition is almost inescapably retrospective. People do not so much observe that a field of study is becoming a discipline; rather, they note that it has been a discipline for several decades.

Many of the recently emerged disciplines are in the education of students in specific subject matter. While education has been a recognised discipline for more than a century, some teachers have come to recognise that the general principles of education are not necessarily sufficient to ensure quality learning in their specific subject areas; that more benefit can be obtained by moulding those principles to better suit the subject matter. With enough interest, this leads to the emergence of new disciplines from the combined parent disciplines of education and the specific subjects being taught.

An overarching goal of this research is to investigate the nature and position of computing education research (CER), also known as computer science education research or informatics education research: does CER exist as a research discipline, and how would we determine this? In line with the discussion above, if CER does exist as a discipline, it has presumably emerged from the combined disciplines of computing research and education research (which includes educational psychology).

Why is disciplinary status important? Because, as Tedre (2015) points out, a discipline can have a recognised image, control of its own curriculum, the power to select its own students and staff, representation in policy-making, and access to directed funding. A subdiscipline of some parent discipline typically has few of those things.

Computing is generally regarded as a member of the STEM disciplines: Science, Technology, Engineering, and Mathematics. In exploring CER, it would therefore be worth looking at what has been done in the cognate areas of science education research, engineering education research, and mathematics education research.

A foundational discussion of science education research is that of Fensham (2004). In his book *Defining an identity: the evolution of science education as a field of research*, Fensham traces the development of science education research, from areas of science on the one hand and educational research on the other, observing that it took place over many decades. In charting this progress, he proposes a number of criteria for determining that science education has become a field of research, criteria such as research journals, research conferences, research methodologies, and implications for practice. Fensham's criteria will be applied to the field of computing education research in a later chapter of this thesis.

Work in other cognate areas tends to focus on particular aspects, which might correspond to one or two of Fensham's criteria. Some examples are provided here for illustration.

A report on discipline-based education research prepared for the US National Academy of Sciences (Singer et al 2012) discusses six specific domains of education research: physics, chemistry, engineering, biology, the geosciences, and astronomy. The report provides a comprehensive summary of research in each of those areas, describes some of the salient findings, and discusses future research directions. The analysis of the current state is based on one group of Fensham's criteria, and concludes that most of the six educational research fields under consideration show evidence of meeting those particular criteria.

In engineering education, Wankat (1999, 2004) analyses the papers published over ten years in the Journal of Engineering Education, widely regarded as the leading journal in the field, with the goal of establishing the proportion of those papers that constitute research. To this end he analyses keywords, citations, financial support, collection of data, use of theories of learning, and bibliometrics. Osorio and Osorio (2002) expand on this approach with an analysis focusing in part on another question: why is it not easy to find publications in engineering education research? Taking a broader approach, Borrego (2007) considers engineering education in the light of Fensham's criteria, and finds that engineering education research falls well short of meeting all of them.

In mathematics education a number of papers focus on the use of theories in research, a focus that we did not find in other areas. Papers include *Theories of mathematics education: a global survey of theoretical frameworks/trends in mathematics education research* (Sriraman & English 2005); *Diversity of theories in mathematics education – how can we deal with it?* (Bikner-Ahsbahs & Prediger 2006); and *The concept and role of theory in mathematics education* (Niss 2007).

In physics education, McDermott and Redish (1999) list and categorise more than 200 publications as resources both for physics education researchers and for physics educators who wish to apply research results in their teaching. Heron and Melzer (2005) discuss the sometimes tenuous relationship between empirical studies in education and specific theoretical frameworks, but acknowledge that significant advances can flow even from studies that lack a solid foundation in theory.

The necessary common factor in all of these approaches is an examination of the literature, although some, such as Fensham (2004), include substantially more than the literature analysis. The works cited here vary widely in how they analyse the literature of their disciplines; but every one of them is based on an analysis of the literature – necessarily, because it is in the literature of a discipline that its approaches, theories, and findings are described and shared with others.

1.2 Publication in computing education

In the latter half of the twentieth century the discipline of computer science emerged from mathematics on the one hand and electrical engineering on the other. Tedre (2015) paints an excellent picture of this emergence. Somewhat in parallel, information systems emerged from mathematics and business. The teaching of these disciplines at universities and colleges led to the gradual emergence of computing education as a discipline of interest.

Publication in computing education has been led largely by SIGCSE, the ACM's Special Interest Group on Computer Science Education. For more than 40 years it has held the annual SIGCSE conference in the USA, and for 20 years it has also held the annual ITiCSE conference outside the USA. Both SIGCSE and ITiCSE have been described as 'swap meets' (Goldweber et al 2004). Valentine (2004) elaborates: "Colleagues describe how their institution has tried a new curriculum, adopted a new language or put up a new course … [These] presentations serve an important function: we are a community of educators and sharing our successes (and failures) enriches the whole community."

However, there was a growing interest in conducting sound educational research and blending this with the swap-meet papers, and this move was welcomed by some commentators.

Dale (2002) examined the titles of papers presented at computing education conferences, and concluded that "beginning in 1998, there is a definite increase in CSEd research related papers . . . CSEd research has arrived in the main stream where it belongs." As a direct consequence of that conclusion, Valentine (2004) looked beyond the titles to the content of selected papers from a single major conference, and confirmed Dale's view. Looking at papers from the SIGCSE Technical Symposium, and only at those that deal with the first two programming courses, he concluded that there was a clear and significant increase in the proportion of what he called experimental papers, and therefore that CS educational research had arrived.

Lister (2006) argued strongly that computing academics need to conduct research into their students' learning and their own teaching if their discipline is to survive. "If we bring a research mentality to our teaching, if we approach our teaching as the formal study of how novices become experts within computer science, and if we study ourselves as much as we study the students, then the best of times are ahead . . . If we choose instead to ignore the type of research summarized in Moby Book [Pascarella and Terenzini (1991)], then I fear students will never return to the computer science major. Instead, the students will learn the minimum necessary about computers (not computer science) while studying within other disciplines."

This argument was not made in a vacuum. The number of students attending universities and other educational institutions had undergone a dramatic increase. Rather than catering to an academic elite, these institutions were now trying to educate nearly half of the school-leaving population. They were expected to maintain their educational standards, but at the same time to pass about the same number of students as before, so as not to damage the reputations of the institutions. In this climate, it was observed that computing courses were often ill-regarded by their students, and some of the courses had pass rates low enough to attract individual attention from their institutions. Many educators had proposed new initiatives in the hope of improving the student experience or student pass rates; but without evidence of their effectiveness, these initiatives were nothing more than suggestions. Rigorous research was seen as a way to validate some of these initiatives, giving educators a convincing reason to adopt them, and promising real change in experience or pass rates.

Notwithstanding the push to increase research, Dale (2002), Valentine (2004), Fincher and Petre (2004), and others acknowledged the continuing value of the swap-meet nature of computing education conferences, and Goldweber argued (Goldweber et al 2004) that there was a danger in an increasing expectation that computing education papers all be research papers.

1.3 Research questions

While there is thus clear evidence that ten years ago there was a growth of interest and publication in computing education research, what is not clear is whether computing education research can since be said to have emerged as a research discipline.

The principal research question of this work is therefore

> ***RQ1: Has CER emerged as a research discipline?***

This question will be addressed using the criteria proposed by Fensham (2004) for a research discipline. As mentioned earlier, these criteria include research journals and research conferences. It is readily confirmed that computing education has journals and conferences. However, are they all essentially swap meets, or can any of them be confidently described as research venues? The question cannot be answered until we have a way of identifying which publications can be considered as research publications. This leads to two subsidiary research questions,

> ***RQ2: How can it be determined whether a publication constitutes research?***

RQ3: What other aspects of computing education publications would lead to an informative broad picture of complete corpuses of them?

These questions will be addressed through the development and application of a purpose-designed system for classifying computing education publications. The system was created because no existing system appeared able to answer the questions.

Assuming that there is a satisfactory answer to RQ2, it should then be possible to observe the growth in research publications alluded to by Dale (2002), Valentine (2004), and others:

RQ4: Do publishing venues for computing education papers show an increase in the proportion of their papers that can be described as research?

This question will be addressed by applying the classification system to a number of computing education venues over several years and measuring the proportions of their papers that are classified as research papers.

Independently of whether there is a perceived growth in the proportion of research papers in computing education, two of Fensham's criteria are the existence of research journals and research conferences within the field.

RQ5: Does computing education have one or more research conferences and one or more research journals?

This question will be addressed by way of the same classifications as RQ4, but considering the variations between venues rather than variation with time.

One further research question emerged as the work progressed:

RQ6: What theories, methods, and approaches are used in computing education research papers, and in what disciplines do they originate?

This question will be addressed by a deeper analysis of the research papers than is possible with the classification system described above.

1.4 Thesis structure

Chapter 2 of this thesis establishes the background for a new classification system for computing education publications. Chapter 3 describes the classification system that was developed, and summarises publications **P1** and **P2**, in which the system was introduced and applied to two different corpuses of computing education publications.

Chapter 4 notes the value of measuring inter-rater reliability for a classification system and describes how a number of other researchers were trained to use the system, after which they applied it to several more corpuses of publications. The chapter then summarises publications **P3** and **P4**, which arose from the work of the team. Chapter 5 summarises further work carried out by the team, work that has not been published.

Chapter 6 discusses further classification work that I carried out alone. In this work the classification system was adapted so that it could be applied to publications that are not necessarily in computing education, and bibliometric analysis was introduced to supplement the classification work. The chapter then summarises publications **P5**, **P6**, and **P9**.

Chapter 7 describes two further projects that in part arose from the classification system and its findings, projects that delved more deeply into the research aspects of computing education publications. These projects gave rise to publications **P7**, **P8**, and **P10**.

Chapter 8 draws these threads together and uses them to help answer the initial question: has computing education emerged as a research discipline?

Finally, chapter 9 summarises the work and its significance for the research endeavour that it has set out to describe.

2. Classifying computing education publications and identifying those that constitute research

2.1 Practice versus research

In early 2004, at a workshop designed to induct computing academics into computing education research, workshop co-leader Sally Fincher snorted and said “practice paper!” With those two words she dismissed a paper that one of the workshop members had proposed for discussion. Until then I had never heard the term, but as the workshop proceeded I quickly came to appreciate the distinction between practice papers and research papers in computing education.

Fincher subsequently made the same distinction in print: “To date, much of what is published in CSEd (called ‘research’ or not) has been concerned with noticing phenomena: ‘This is what happens when I teach *x* in a particular way.’ What moves recognition of phenomena to evidence is purposeful investigation and a relationship to theory.” (Goldweber et al 2004).

As discussed in the preceding chapter, there is clearly a place for practice papers in the computing education literature. An academic area without innovation stagnates, and education in the highly dynamic subject matter of computing can ill afford to be stagnant. Practice papers are the principal means of sharing innovations, which is presumably why Goldweber (Goldweber et al 2004) urged the community not to sacrifice these papers in the quest for a stronger research community. On the other hand, rigorous research is also vital for an academic discipline, and this is the point being made by Fincher and others.

For example, Holmboe et al (2001) express the desire for “more empirical research and comparative evaluation”, so as to “strengthen the case for Computer Science Education Research to be taken seriously as an academic discipline, and counter the criticism . . . that it is merely a way for ‘teachers to write papers’.”

The practice-research distinction was also evident in the calls for papers of some conferences. In 2004, Koli Calling explicitly introduced a submission category for research papers

> “to make a clearer distinction between papers that present novel ideas, approaches and systems for CS education, and papers in

> which these issues have been elaborated further in some rigid research setting. Both types of papers are, however, equally necessary for the whole CS education community. New ideas and tools are the fuel for research work, and research is needed to convince us that we are really making progress towards our goal of improving learning." (Malmi 2004).

In 2007 the journal Computer Science Education announced a special issue on Computing Education – the Australasian Perspective ('Australasia' meaning Australia and New Zealand). The call for papers made it clear that submissions were expected to review developments in specific areas of computing education such as curriculum, introductory programming, capstone projects, gender issues, etc. In contrast, I proposed to review all of the papers in the relevant publications, partly to see which of these areas they fell into, but also to see how many were practice papers and how many were research papers. The proposal was accepted by the guest editors, and I proceeded to develop a system for classifying complete corpuses of computing education publications.

2.2 Previous classifications of related publications

Some of the literature already discussed formed a clear context for the development of this new system. Valentine (2004) had examined 20 years of proceedings from the annual SIGCSE conference, identified those papers that related to introductory programming, and classified them into six groups, one of which, *experimental*, was the group of papers that he considered to represent computing education research. Although this was a systematic approach, he looked only at a subset of papers (those set in the first two programming courses); his *experimental* category is probably too broad, as it includes papers with any mention of evaluation; and his other categories are somewhat problematic, as they measure different aspects of a paper. For example, *tools* concerns the subject matter of a paper, while *experimental* assesses the nature of the work carried out.

Randolph (2007a; also Randolph et al 2005, Randolph 2007b, Randolph et al 2007) was somewhat more rigorous, but limited the bulk of his study to particular papers, those that he had already deemed to be research. Within that grouping, his interest was in the methodology applied in the research, and his principal finding was that many research papers applied methodology poorly or not at all. Whenever papers are preselected for study, the preselection itself is open to question. Lister (2007) praised Randolph's thesis, but suggested that Randolph had included many papers that Lister did not consider to be research. This has the clear potential to influence Randolph's findings: if non-research papers are included when assessing the methodology that they employ, the proportion of papers that appear to employ no methodology will be artificially elevated.

Fincher and Petre (2004) did not classify any publications, but they did provide a list of ten 'broad areas that motivate researchers in CS education'. It might then be reasonable to expect that many publications can be placed into one of these broad areas: student understanding; animation, visualisation and simulation; teaching methods; assessment; educational technology; transferring professional practice into the classroom; incorporating new developments and new technologies; transferring from campus-based teaching to distance education; recruitment and retention; construction of the discipline.

Pears et al (2005, 2007) explored the establishment of a core literature for computing education research. In accordance with their goal, they classified papers as influential, seminal, or synthesis; but they also classified them as small-scale, institutional, dealing with problems/solutions, and dealing with computing education research, the first three of these being reduced from the ten areas of Fincher & Petre (2004). In their specific analysis of introductory programming papers (Pears et al 2007) they also nominated topics of curriculum, pedagogy, language choice, and tools.

Beyond the realm of computing education, Glass and colleagues had analysed corpuses of publications in computer science (Ramesh et al 2003), software engineering (Glass et al 2002), and information systems (Vessey et al 2002), and had compared the three areas (Glass et al 2004). Their analysis (Vessey et al 2004) identified the academic topic in which each publication was set, and then focused on aspects of the research: the research approach taken, the research method used, the discipline in which that method originated, and the level of analysis of data.

Research question RQ3 asks what other aspects of computing education publications might permit informative analysis of complete corpuses of them. Answering this question was an immediate goal of the research reported in this thesis. Classification systems designed for specific subsets of a corpus are unlikely to apply readily to the full corpus. Indeed, it is likely that some further classification can be inferred by generalising from the subset, asking what other types of publication are found in the corpus. For example, Valentine classified only publications that he identified as dealing with introductory programming; a broader classification system would identify the courses in which all papers are set, with introductory programming being one possible value. Likewise, Randolph and colleagues classified only publications that they identified as research; a broader classification system would require values by which to classify the publications that do not appear to represent research.

2.3 The difficulty of defining research

Research question RQ2 asks how we can establish whether a publication constitutes research. Valentine (2004) answered this question by defining his *experimental* category as papers incorporating any sort of scientific analysis; this would include, for example, papers that focused on presenting a teaching innovation and then briefly analysed a student survey conducted at the end of the course.

Most other definitions of research are either succinct and somewhat trivial or extended and more informative. The difficulty of clearly defining research can be seen in this attempt by the British Research Assessment Exercise:

> 'Research' . . . includes work of direct relevance to the needs of commerce, industry, and to the public and voluntary sectors; scholarship; the invention and generation of ideas, images, performances, artefacts including design where these lead to new or substantially improved insights; and the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction. It excludes routine testing and routine analysis of materials, components and processes such as for the maintenance of national standards, as distinct from the development of new analytical techniques. It also excludes the development of teaching materials that do not embody original research. (RAE 2005)

In the context of educational publications, it is interesting to note the explicit exclusion in the final sentence of that definition.

In his chronicle of the emergence of computer science as a discipline, Tedre (2015) notes that even within the realm of science, research can be

- exploratory: developing an initial understanding of an uncharted phenomenon;
- descriptive: systematically recording and modelling a phenomenon and its relationships to other phenomena;
- predictive: using previous understandings to predict phenomena that are yet to come; or
- explanatory: clarifying the causes, relationships, and consequences of a phenomenon.

Quoting Wegner (1976), Tedre adds that when the area of interest is expanded to include engineering, research can also be

- constructive: developing tools that accomplish classes of tasks more efficiently.

Fincher and Petre (2004) devote a chapter of their book to examining the features of many forms of research, discussing such matters as the method of science and the scientific method, predictive and explanatory theories, empirical laws, models, and conceptual frameworks. They acknowledge that the scientific method is seldom an appropriate approach for research involving humans. They then list and expand on six principles for computing education research, which were first proposed in a report on scientific research in education for the US National Academy of Sciences (Shavelson & Towne 2002):

- pose significant questions that can be answered empirically
- link research to relevant theory
- use methods that permit direct investigation of the question
- provide a coherent and explicit chain of reasoning
- replicate and generalise across studies
- disclose research to encourage professional scrutiny and critique

Literature in the philosophy of science seldom appears to define scientific research; rather, it poses definitions for science itself, with the implied assumption that doing science, which some authors call 'sciencing' to identify it as an activity rather than a body of knowledge (Azevedo 1997), is the same as doing scientific research. The definitions vary considerably depending on which model of science is favoured. The many different models of science, including induction, falsificationism, paradigms, rationalism, relativisim, objectivism, anarchism, and realism, all lend themselves to different definitions of science (Chalmers 1982). However, one of the key components is the development and application of theories, as opposed to a simple listing of findings (Fawcett 1999, Okasha 2002).

To the extent that computing education research is based on education research, it is pertinent to ask how research is defined in education. While there are many books about research in education and the social sciences, and many of them include chapters or sections called 'What is research?', again there are few clear answers; instead there are often many pages describing aspects of research without directly defining it (Blaxter et al 2001, Brew 2001, Creswell 2003).

Slavin (2007) offers a one-sentence definition: "Research is organized, systematic inquiry that seeks to answer well framed questions". However, such a definition offers little detail of the research process, and provides little clarity as to whether a particular publication embodies research.

The process is encapsulated in a little more detail by Kervin et al (2006), who suggest that the process consists of five stages: problem definition, research design, data collection, data analysis, and communication of results.

Still more detail is suggested by Bouma and Ling (2004), who define the research process as comprising three phases:

- Phase 1: Essential first steps – clarifying the issue to be researched and selecting a research method
  - select, narrow, and formulate the question to be studied
  - select a research design
  - design and devise measures for variables
  - set up tables for analysis
  - select a sample
- Phase 2: Data collection – collect evidence about the research question
  - collect data
  - summarise and organise data
- Phase 3: Analysis and interpretation – relate the evidence to the research question, draw conclusions about the question, and acknowledge the limitations of the research
  - relate data to the research question
  - draw conclusions
  - assess the limitations of the study
  - make suggestions for further research (Bouma & Ling 2004)

However, when a definition reaches this level of detail, it is necessarily descriptive rather than prescriptive. For example, if a piece of work were to include all but the final step in the process, not suggesting directions for future research, it would seem a little harsh to disqualify it as a piece of research.

2.4 Research – an alternative perspective

It was noted in the previous section that there is no clear and consensual definition of science. Why is this? Why do the generally accepted definitions change (Chalmers 1982)? Because science is a socially defined activity, and its definition varies according to time and audience. In the same way, research is a socially defined activity. It is defined by governments, research universities, funding agencies, and other individuals and groups with various degrees of interest. It is apparent in the wording of calls for papers, in review criteria promulgated to reviewers of papers, and in the decisions made by conference chairs and journal editors.

It can readily be argued that everything accepted for publication in a journal or a conference can be considered as research. For example, the Australian government's Higher Education Research Data Collection, which collects data on research income and outputs from all Australian universities, effectively defines research as what is published, so long as it shows evidence of:

- substantial scholarly activity, properly situating the work in the relevant literature;
- originality (it is not just a compilation of existing works);
- peer review;
- an increase in the stock of knowledge;
- being in a form that enables dissemination. (HERDC 2015)

Under the interpretation that everything accepted for publication can be considered research, research questions RQ2, RQ4, and RQ5 become both trivial and unhelpful: a publication can be considered research by virtue of being a publication; computing education venues cannot show an increase in the proportion of their papers that can be considered research, because that proportion is always 100%; and any computing education conference or journal is necessarily a computing education research conference or journal.

The interpretation of research that is required to make these questions non-trivial can perhaps best be described as that which is present in research papers but not in practice papers when they are so divided; or as that in which writers such as Fincher and Petre, Dale, Valentine, and Lister (section 1.2) are either observing or urging an increase. In the next five chapters, this thesis will propose and use a definition of research that does not include all publications, and that therefore permits an examination of whether the proportion of research papers in computing education is increasing.

However, if all publications are to be considered research, an alternative set of research questions would ask how the nature of the research has changed over the years. For example, has the proportion of descriptive research dropped in favour of explanatory research? The analysis from section 7.2 onward takes this alternative view, that all publications can be regarded as research, and delves into the nature of that research.

Neither perspective is right; neither perspective is wrong; they are two sides of the same coin, two different approaches to answering what is at heart the same question.

It is clear that there is no single simple prescriptive definition that can be applied to large numbers of papers to determine easily and quickly whether they are research papers. Indeed, it is not possible for there to be such a definition. Therefore a pragmatic classification system must synthesise the available definitions into one that is brief, easily understood, and easy to apply to publications – with the understanding that it is descriptive and convenient, with no pretensions to being authoritative. Such a definition will be developed in the next chapter.

It is also clear that no single dimension of classification can capture all the features of interest of a paper. Considering again the classification of Valentine (2004), a single paper might belong to three of the six categories: it might be experimental, about tools, and indeed John Henry (referring to doing a reasonably simple task in a difficult way). Valentine resolved such issues by ordering the classification values and classifying each paper into the highest-ordered category to which it applied. This was adequate for his purposes, but is not helpful when one is trying to gather all of the information. Therefore the new classification system would require several dimensions, with a paper's classification in one dimension having no direct influence on its classification in the others.

In the next chapter the new classification system will be described, and then compared with the previous classification systems discussed in this chapter.

3. A new classification scheme for computing education publications

Publication P1 was written for a special issue of Computer Science Education subtitled Computing Education – the Australasian Perspective, and was intended to provide an overview of computing education in Australia and New Zealand by way of an analysis of the complete corpus of relevant publications.

It would have been impractical to seek out all publications world-wide that have Australasian authors, so the corpus was defined as the publications at ACE, the Australasian Computing Education conference, and the computing education publications at NACCQ, the conference of the National Advisory Committee on Computing Qualifications, now known as CITRENZ (Computing and Information Technology Research and Education New Zealand). The conference of this body incorporates general computing as well as computing education, and only the computing education publications were chosen for analysis.

It was decided that to provide an overall picture of the corpus, four distinct dimensions would be required. A paper would be classified according to what it was about, what kind of course it was set in, what level of collaboration was evident in it, and whether it was a research paper. These dimensions are described in the following sections, with reference where appropriate to the earlier classification systems discussed in section 2.2. Following the description of the new system (which has now come to be known as Simon's system), it will be compared with the previous systems to illustrate why none of them could provide as comprehensive a picture of the state of computing education publications.

3.1 The nature dimension

Valentine's (2004) system includes three values that appear to describe the nature of a paper. The first is *experimental*, which we have already discussed, and which Valentine took to indicate that this was a research paper. The second is *Marco Polo*, named for the explorer, which Valentine characterises as 'I went there and I saw this'. These are the papers that Fincher and others call practice papers or experience reports; they might report on a new tool developed for use in the classroom, an innovative teaching technique, a form of assessment, or one of many other matters. The third is *philosophical*, 'where the author has made an attempt to generate debate of an issue, on philosophical grounds, among the broader community'. I believe that Valentine's use of the word 'philosophical' is at times generous, as many papers of the sort he describes owe everything to their authors' opinions and nothing to philosophy. Nevertheless, the category is useful.

While Randolph et al (2005) analyse only the papers they consider to be research, they do list the other types of paper that they encountered, describing them as

- literature reviews, meta-analyses
- programme descriptions without anecdotal evidence
- programme descriptions with anecdotal evidence
- theoretical, methodological or philosophical papers
- technical investigations
- other

Glass et al (2004) adopt the view (see section 2.4) that all of the papers they examined are research papers, so they do nothing to distinguish research papers from other papers.

Simon's system initially had four possible values to describe the nature of a paper.

A *position* paper is very like Valentine's and Randolph's philosophical papers. In a position paper the authors express their opinions, but generally present no experience report or empirical work to back up those opinions. Such papers can be valuable in a research community, but it seems reasonable to expect that they will form just a small minority of the papers.

A *report* paper is one of those papers variously called practice papers, experience reports, or Marco Polo papers. Essentially, the purpose of such a paper is to share an experience with the rest of the community, to say 'Here's something I've done; you should try it.' If I had wanted to use quirky names for the categories, I would have called these papers Genesis papers: 'and he saw what he had made, and it was good'. One distinction between these and Valentine's Marco Polo papers is that a brief evaluation of the experience, such as an analysis of a student satisfaction survey, is not enough to elevate a paper to a higher category. If the apparent purpose of the paper is to present the innovation to the community, it is classified as a report paper.

An *analysis* paper conducts substantial analysis of some type on data that is already available, such as student performance over several years. This analysis is clearly the purpose of the paper, not just a postscript to a paper whose purpose is to descriptively report on something such as an innovation in the classroom.

An *experiment* paper reports on a project in which data is collected for the specific purpose of the research and then analysed. This data might be as simple as student surveys designed and conducted to address a specific research question, or might be more involved, such as interview or focus group transcripts, artefacts produced by students or teachers, etc. The data will not be something that was going to be collected in the normal course of teaching, such as student assignments, standard satisfaction surveys, etc.

3.2 Research – a definition of convenience

Notwithstanding the difficulty of defining research, as described in sections 2.3 and 2.4, it appears reasonable to assert that both analysis papers and experiment papers can be considered as research. They pose a question, although not always explicitly; they gather data relevant to that question; they analyse the data; and they infer conclusions.

Research is not restricted to the analysis and experiment papers. A report paper might report on the design and construction of a completely novel approach to addressing research questions. A position paper might consider all the existing theory and then propose a novel theory, method, model, or framework to be used in research. It would be harsh not to consider these as research papers. Consequently we cannot assert that all of the research papers will be in either the analysis or the experiment category; but we can assert with some confidence that all of the papers in those categories can be considered as research; and, indeed, that they are in that particular class of research that is generally recognised as empirical research. While this is not the only kind of research, it is the kind that authors such as Fincher & Petre (2004), Valentine (2004), and Lister (2006) are pressing educators to engage in, so it is a useful thing to measure.

This, then, is the initial definition of research that has been used in the analysis of computing education publications by way of Simon's system:

> ***A research paper is a paper that in Simon's system is given a nature classification of analysis or experiment.***

Note, however, that the definition was subsequently changed, when the experiment category was split into study and experiment. The new definition will be found in section 4.1.

3.3 The topic dimension

What is a paper about? One of Valentine's six categories, *tools*, addresses a paper's subject matter. Clearly there are more subjects being addressed in computing education papers. Fincher and Petre (2004) list ten, such as student understanding, student retention, teaching methods, and assessment. They make it clear that their list is not intended to be comprehensive, but rather is indicative of the topics that were generating research papers at the time they wrote. In the context of papers on introductory programming, Pears et al (2007) acknowledge papers about curriculum, pedagogy, language choice, tools, and computing education research.

A paper's topic is orthogonal to its nature: a paper about student retention could be a position paper (this is how we think the problem could be addressed), a report (this is what we have done about student retention), an analysis (this is what the enrolment figures over the past ten years tell us), or experimental (we interviewed students who had dropped the degree to find out why they did so). Therefore a topic list arising from an observation of research papers should apply equally well to other papers.

The list of topics in Simon's system began with something very like the topics listed by Fincher and Petre, but was essentially data-driven: when a paper is found that is not about any of the items on the current list, the topic is condensed into a word or phrase (for example, assessment tools, teaching/learning techniques, language/culture issues, etc) and added to the list.

Following the analysis of three years of Australasian computing education publications, the topics were as listed in Table 1. However, the list must be considered dynamic, as new topics will presumably be encountered when further papers are analysed; the current list is included in an appendix to this thesis.

Table 1. Topics identified in analysing three years of Australasian computing education papers. Each topic is briefly explained in Table 1 of publication P1.

ability/aptitude	curriculum	recruitment
about research	distance/online delivery	teaching/learning
assessment techniques	educational technology	teaching/learning techniques
assessment tools	ethics/professional issues	teaching/learning tools
cheating & plagiarism	gender issues	tutors & demonstrators
credit for prior learning	language/culture issues	

3.4 The context dimension

Most computing education papers are set in the context of a particular course (also known as a subject, a unit, or a paper): a unit of teaching, typically one semester long, at the end of which students are awarded a final grade. This might be a course in programming, database design, software engineering, computer graphics, etc. This setting provides the *context* of the paper.

It is easy at first to confuse a paper's context and its topic. One might observe that a paper is about data mining, only to appreciate on reading it that is about ethical issues as taught in a data mining course. Such a paper would have a topic of ethical issues and a context of data mining.

It follows that context is orthogonal to both nature and topic. In the context of data mining one might equally find a report on the topic of assessment techniques, an analysis paper on the topic of student retention, or an experiment paper on the topic of gender issues.

There are curriculum specifications that try to provide comprehensive lists of all possible topics in computing; for a current example, see Computer Science Curricula 2013 (CSC 2013). However, these can run to many pages, and are so detailed that any given course will cover many topics. For simple classification a system requires contexts that encompass potentially large numbers of courses (for example, data structures), rather than individual topics of study within those courses (for example, linked lists, red-black trees, etc). Therefore comprehensive curriculum topic lists are not a useful model for the context dimension.

For their classification, Glass et al (2004) use a list of 49 'topics' based on earlier work by Glass (1992) that was in in turn derived from a number of curriculum specifications. It is clear that they use the word 'topic' to refer not to what the paper is about, but to the academic topic being taught in the work described by the paper: that is, to the context of the work.

In the development of my system the list of contexts, like the list of topics, was essentially data-driven. Beginning with a list of known contexts, try to give each paper a context from that list; if it is clear that the context of a paper is none of those in the list, choose a word or phrase to describe the context and add it to the list. When classification is complete, remove from the list any contexts for which no papers were found.

Some papers were found that were not set in the context of particular courses, or where the course was not the important aspect of the work, and several contexts were devised to accommodate these papers. An analysis of the literature, even of the literature within a particular classroom context, does not represent work carried out in that context; for such work, a context of *literature* is more appropriate. For a paper focusing on group work, regardless of whether it is in a programming course or an e-commerce course, the context of *group work* helps to make more sense of the paper. And work carried out in no particular course, but pertinent to many courses – for example, the development of a tool for recording students' marks – would be considered *broad-based*.

The list of contexts is necessarily dynamic. Even if it were feasible at any one time to list all possible contexts, new contexts would arise as new areas of computing were added to the syllabus. Table 2 lists the contexts found in the analysis of three years of Australasian computing education papers. Again, the complete current list can be found in the appendix.

Table 2. Contexts identified in analysing three years of Australasian computing education papers. Each topic is briefly explained in Table 2 of publication P1.

broad-based	group work	professionalism
basic skills	hardware/architecture	programming
capstone project	html	programming languages
communication skills	information systems	systems analysis
database	interface design	theory of computation
design	literature	work experience
e-commerce	networks	

3.5 The scope dimension

The final dimension, *scope*, was devised to assess the breadth of collaboration evident in the paper. This measure does not appear explicitly in any of the earlier classification systems discussed in section 2.2, but Fincher and Petre (2004) allude to a distinction between small-scale and institutional works, and Pears et al (2007) list 'small scale' and 'institutional' as two of four possible 'areas' into which the papers they examined might fall. Furthermore, much of the work in computing education appears to be based in single-classroom settings, but some recent multi-institutional studies (McCracken et al 2001, Lister et al 2004) had proven to be milestones in addressing important problems. It would therefore be interesting to see how widely these collaborations are happening; whether they are outliers or a trend.

Once more informed by the data, five concrete values of scope emerged. The scope of *subject* indicates work done within a single subject (course/unit/paper). This corresponds to some extent with the 'small scale' of Pears et al (2007), and appears to count for a majority of publications. *Program/department* refers to work done in two or more courses within the same academic department or degree program, and tends to necessitate a greater breadth of collaboration. *Institution*, akin to the 'institutional' of Pears et al (2007), indicates collaboration between two or more academic departments within the same institution. And *many institutions* indicates a collaboration between researchers at two or more institutions.

Not all papers have an identifiable scope. For example, a position paper offering suggestions for improving success rates among ethnic minorities is not based in a subject, a department, an institution, or many institutions. Instead it transcends those values. However, a full classification of a set of papers requires that no value be left blank; so an additional value of *not applicable* was included in this dimension.

As with the other dimensions, there will sometimes be some ambiguity in a classification. If researchers at three institutions collaborate in a study, but the participants in the study are all students in a single course at a single institution, the work will probably be accorded a scope of subject, as it did not necessitate the involvement of the researchers at the other two institutions. On the other hand, if it is clear that their involvement was in some way necessary to the work, and particularly to the data collection, a scope of many institutions might be chosen.

While scope is not a measure of the nature of a paper, some correlation between these dimensions is expected. For example, papers broader in scope, indicating more collaboration, might be more likely to be research papers.

3.6 Comparison of Simon's system with previous classifications

The dimensions of Simon's system have been described with reference where possible to prior classification systems. Why, then, was it not possible to use these prior systems to classify the Australasian computing education papers? Why was a new system required? To help address this question, Table 3 summarises the dimensions and classifications of all of the systems considered.

The clear separation into four dimensions, as illustrated in Table 3, highlights some of the strengths of Simon's system. Valentine's system has only six possible classification values, but three of them concern the nature of a paper, one concerns its topic, and the remaining two refer to properties that would lie outside most other systems of classification. Randolph's system addresses a paper's nature fairly thoroughly, but pays no heed to the topic, context, or scope, thus limiting its value in presenting a big picture of a corpus of publications. Fincher and Petre provide a limited list of topics, and allude to an element of scope. Pears et al offer some coverage of three of the four dimensions, but not so clearly delineated. For example, one of their categorisations comprises small scale and institutional, which describe scope; problems/solutions, which appear to describe an aspect of nature; and computing education research, which appears to describe a topic.

Simon's system does not capture everything addressed by the other systems. Both Randolph et al and Glass et al delve deeply into the research methods employed within papers, an analysis that was beyond the scope of the current work. And in accord with their goal of establishing a core literature for computing education research, Pears et al explicitly classify papers as influential, seminal, or synthesis – properties that cannot necessarily be identified in a paper that has been published only recently.

In the light of these comparisons, Simon's system does appear to hold some promise for the classification of complete corpuses of publications.

Table 3. Elements of six different classification systems for computing and computing education papers.

	Nature	**Topic**	**Context**	**Scope**	**Beyond**
Simon	• experiment • analysis • report • position	• 20 values	• > 30 values	• subject • program/department • institution • many institutions • not applicable	
Valentine	• experimental • Marco Polo • philosophical	• tools	*all introductory programming*		• nifty • John Henry
Randolph et al	• research • literature reviews, meta-reviews • programme descriptions • theoretical, philosophical, methodological • technical investigations • other				• analysis of research method
Fincher & Petre		• 10 values		*small-scale vs institutional*	
Pears et al	• problems/ solutions	• curriculum • pedagogy • language choice • tools • computing education research		• small scale • institutional	• influential • seminal • synthesis
Glass et al	*all assumed to be research*		• > 45 values (called topic)		• analysis of research method

3.7 Publications P1 and P2

Publication P1 is the paper that appeared in the special issue of Computer Science Education. It introduced and explained the classification system and, in keeping with the theme of Computing Education – the Australasian Perspective, provided the results of analysing three years of papers at the most pertinent Australian and New Zealand conferences.

The system was used to classify 175 papers, 129 from ACE and 46 from NACCQ, which appeared between 2004 and when the analysis was carried out. The principal findings were:

- the dominant topic was teaching/learning techniques, with curriculum, teaching/learning tools, assessment techniques, and ability/aptitude each accounting for more than 5% of the papers;
- about 35% of the papers were set in the context of programming courses, about 22% were broad-based, and no other context accounted for more than 7% of the papers;
- report papers accounted for more than half the corpus, but analysis and experiment papers together made up some 35%;
- subject was the dominant scope with nearly 60% of the papers; some 15% had a scope of program/department and some 13% had a scope of many institutions.

Of the SIGCSE papers that he analysed, Valentine classified 21% as experimental. By comparison, 35% is a good proportion of analysis and experiment papers, those that can clearly be described as research. Nevertheless, the dominance of subject as a scope and report as a nature suggests that ACE and NACCQ at that time were far from being research conferences.

Publication P2 reports on a similar analysis of the 102 papers from all six years of Koli Calling, the computing education conference held annually in Finland. Six years was long enough to look for trends, and it was pleasing to see that while the first three years of the conference were dominated by reports, the next three were dominated by analysis and experiment papers, as shown in Figure 1. This is clear evidence that computing education research was a growing force in the papers presented at this conference.

Other aspects of Koli Calling were fairly similar to the Australasian papers:

- the dominant topics were teaching/learning techniques and teaching/learning tools, with most other topics accounting for less than 10% of the papers each;
- about 35% of the papers were set in the context of programming courses, about 22% were broad-based, and no other context accounted for more than 7% of the papers;
- subject was the dominant scope with some 53% of the papers; some 27% were classified as not applicable, often because they reported on the development of pedagogical tools.

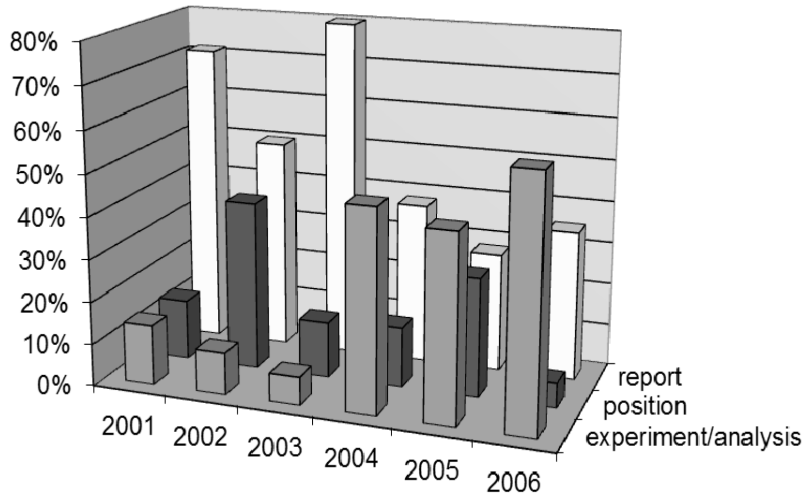


Figure 1. Proportion of papers by nature and year at Koli Calling.

While there were clear differences between the corpuses of papers analysed in publications P1 and P2, the similarities are also interesting, suggesting that the Australasian conferences and the Finnish conference have a great deal in common.

Another point of interest is that in the later years, Koli Calling invited papers in three different submission categories: research, discussion, and system. It might reasonably be expected that most research papers would be classified as analysis or experiment, most system papers as report, and most discussion papers as position. While this was found to be the case, it was not as clearcut as might be expected. Figure 2 plots the natures of Koli Calling papers against their submission categories: it shows a substantial number of reports among the papers accepted as research, and substantial numbers of analysis and experiment papers among those accepted as discussion papers.

This is not particularly surprising. Authors submitting papers to a conference are not asked to familiarise themselves with a classification system before selecting the appropriate category for their paper, so their classification is unlikely to precisely match the findings of a researcher conducting a purposive classification. Furthermore, the submission classifications included an element of paper length, meaning that a short research paper was more likely to be submitted as a discussion paper than as a research paper.

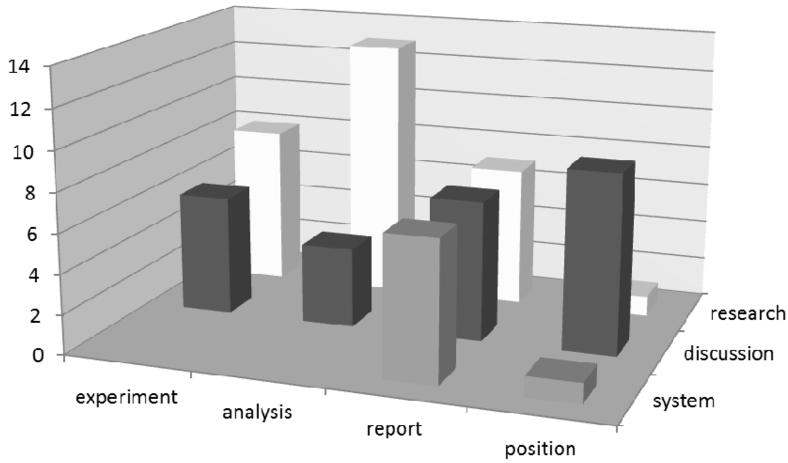


Figure 2. Paper counts of nature classification vs submission category at Koli Calling.

3.8 Research questions RQ2 and RQ3

Research question RQ2 asks how it can be determined whether a publication constitutes research. Using the definition of convenience given in section 3.2, it has been possible to distinguish research papers from other papers, and, for Koli Calling, to detect an increase in the proportion of research papers. This thesis makes the case that that Simon's system therefore offers a useful and usable way of identifying papers that can be considered as research.

Research question RQ3 asks what other aspects of those publications would lead to an informative broad picture of complete corpuses of them. This thesis and its associated papers make the case that combined with nature, the dimensions of topic, context, and scope present a useful and informative picture of a corpus of computing education papers.

4. Validating the scheme with multiple classifiers

4.1 New classifiers and adjustments to the scheme

The classifications for publications P1 and P2 were carried out by a single researcher. The classification system had not been validated in any way, and readers of those publications were asked to accept that the classifications were meaningful.

When items of any sort form a continuum, classifying them into a finite set of categories necessarily involves subjective judgements. If a number of distinct classifiers classify most of the items in the same way, the system is robust; if every classifier chooses different classifications for the same items, the system is useless.

It was now time to formally measure the inter-rater reliability of the classification system. If this was sufficiently high, meaning that different classifiers tend to choose the same classifications for the same papers, the system could be considered validated.

I applied for and obtained a small project grant from SIGCSE, which I used to conduct a workshop prior to ACE 2008. The Australasian Working Party on Classifying Computing Education Literature in Published Sources (AWPOC-CELIPS) attracted six further participants, increasing the number of classifiers from one to seven. While the goal of the workshop was to examine the system's inter-rater reliability, it was decided to focus on computing education papers from the New Zealand NACCQ conference, as the deadline for submissions to that conference was a reasonable time after the workshop.

Once the system had been thoroughly explained, the participants classified two years of NACCQ papers by consensus arising from group discussion. In the beginning it was sometimes difficult to reach consensus, but as the discussion continued, the group's members formed a better picture of what the dimensions and their values were attempting to capture, and agreement was reached more easily.

The papers from a third year were classified individually, with each member classifying each paper. This was followed by further group discussion, which resolved any differences and again led to consensus.

The papers from a fourth year were again classified individually, but this time there was no group discussion. Instead, the individual classifications were recorded and the inter-rater reliability was determined using Fleiss's kappa (Fleiss 1971), following the lead of Randolph et al (2005).

As the workshop drew to a close, participants discussed the classification system and made some adjustments to it.

The *topic* dimension, which had taken its name from the list of Fincher & Petre (2004), was clearly causing some confusion. While participants understood that the dimension was intended to capture what the paper is about, some continued to confuse it with the *context*, the topic that was being taught in the classroom when the work was carried out. The confusion is further exemplified by the fact that Glass et al (2004) use the word 'topic' for what Simon's system calls context. The topic dimension was therefore renamed *theme*.

Within the *nature* dimension, the *experiment* category was split into two. Some of the participants were of the impression that the word 'experiment' should be used only for scientific-style experiments, with control and experimental groups and dependent and independent variables. Other studies, designed and implemented to gather data to answer a research question, but without the features mentioned above, could not reasonably be called experiments. To address this concern, the category was split into *study* and *experiment* categories.

This change gives rise to a revision of the definition of research:

> ***A research paper is a paper that in Simon's system is given a nature classification of analysis, study, or experiment.***

Also within *nature*, the *position* category was renamed *position/proposal*. In earlier classifying, papers describing proposals to undertake work had been classified as position papers because they had no implementation to report on. However, workshop participants felt that there was a clear difference between expressing an opinion and describing work to be undertaken, so the category name was changed to reflect that.

The other adjustments were as expected: in the course of classifying, we had encountered themes and contexts not previously met, and had expanded the lists to accommodate the new values.

These adjustments could be of some concern when comparing fresh classifications with classifications already completed and published. For example, a paper is seen to be set in the context of a cryptography course, and *cryptography* is added to the list of contexts. The question that must then be asked is whether there might have been other papers that belong in this context, but that had, for example, been placed into the more general *programming* context. The answer to this question comes in three parts. First, so long as the classifiers remain alert to the possibility of new contexts, this is probably the first cryptography paper that has been classified. The context of a paper is generally made explicit by the authors, and where possible their information is preserved in the classification. Second, this classification is unlikely to be reported in its own right. A report on the themes or contexts of a corpus of papers will explicitly list the categories with the greatest numbers of papers, and will gather the remaining categories into a general *other* group. It will therefore require not just one or two papers in the new category, but a substantial number, to merit an explicit mention of that category in any report. Third, the classifications are generally applied to a single corpus, which is classified at one time. When we do compare with previous analyses, it is generally in the fixed dimensions, not the dynamic ones. When this is not the case, it is necessary to identify any affected papers that have already been classified, and confirm or alter their classification. For example, *experiment* was split into *study* and *experiment* when some NACCQ papers and some ACE papers had been classified and others were yet to be classified. When this happened, those papers that had already been classified as *experiment* were checked to see if they should now be classified as *study*. Every one of the NACCQ experiment papers was changed, because there were no scientific-style experiments reported in that corpus; most of the ACE papers were also reclassified as study, but two retained the *experiment* value.

An appendix to this thesis shows the complete current set of values for each dimension.

4.2 Inter-rater reliability, Publication P3

Following the workshop, the participants classified a fifth year of NACCQ papers individually, and the inter-rater reliability was measured again. Finally, they formed pairs to discuss their classifications of those papers and achieve consensus within each pair, and the inter-rater reliability was measured for the pairs.

According to Banerjee et al (1999), agreement between raters can be considered fair to good if the measure lies between 40% and 75%, and excellent if it is over 75%. Researchers unfamiliar with these measures might be somewhat alarmed at the thought that 40% agreement can be called fair to good, so some explanation is warranted.

There is a great difference between raw measures of agreement and a calculated kappa. It is not uncommon in CER papers to see words to the effect of: 'We each classified 20 items, and agreed on 16 of them; satisfied with this 80% agreement, we classified the remaining items individually.'

One problem with this is that even if two people were to classify the same items completely randomly, their classifications would agree for some of the items. The inter-rater reliability measures are expressly designed to counter this chance agreement, and indeed are known as chance-corrected measures. Further, if there is a prevalent classification (for example, a context of programming or a nature of report), the measures effectively weaken the weight of this prevalent value by more strongly penalising any deviation from it. As expressed by Banerjee et al (1999), "agreement studies conducted . . . in populations known to have a high prevalence of [a particular classification value] do not necessarily reflect on the agreement between the raters". Further, the more categories that are available to a classifier, the lower the calculated agreement will be, even if not all of the categories are used. This has a clear potential to reduce the measured agreement on the theme and context dimensions.

Another problem is simply the number of classifiers. Complete agreement is clearly more likely between two classifiers than among seven. The more people there are, the more scope there is for one of them to differ from the rest.

This is why there are validated measures of inter-rater reliability, and why percentage agreements that appear low are generally considered to represent fair to good agreement.

The agreement itself is far from trivial to measure. The Fleiss kappa is expressed by the following formula, where N is the number of items to be classified, n is the number of classifications made of each item, k is the number of categories into which the items can be classified, n_{ij} is the number of raters who classify item i into category j, and p_j is the number of classifications into category j as a proportion of all the classifications:

$$\kappa = \frac{\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn[1 + (n-1)\sum_{j=1}^{k} p_j^2]}{Nn(n-1)(1 - \sum_{j=1}^{k} p_j^2)}$$

(Fleiss 1971)

The first time we applied Fleiss's kappa, during the training process, agreement was poor on *nature* and fair to good on the remaining dimensions. On the next application, once the workshop was over, agreement was poor on *theme* and fair to good on the remaining dimensions. On the final application, when pairs took the place of individual raters, agreement was fair to good on all dimensions.

All participants acknowledged that paired classification tends to eliminate simple errors, unintentional selections of the wrong value. It does not, however, eliminate disagreement when the classifiers see the paper as falling into two different categories, as will occasionally happen. Such disagreements are resolved by discussion, although not always to the satisfaction of both partners.

It is important to understand that while there are disagreements, they are on individual dimensions of individual papers, and they are often because the paper could easily fall into either of two categories. The goal of this work is not to classify each paper unequivocally, but to form an overview of a corpus of papers. Therefore even if a number of classifications are seen as wrong by some classifiers, it will not dramatically affect the big picture of the results. For example, one phase of the paired classification required 61 papers to be classified in all four dimensions, leading to 244 classifications from each member of the pair. Of these 244, 79 (32%) were initially different. However, 67 of these were agreed in the first pass of discussion, leaving only 12 (5%), for serious discussion. After one more pass of discussion there remained only one classification on which there was some disagreement: one member of the pair accepted the other's arguments, but somewhat reluctantly. Had the question been resolved the other way, one of the 61 papers would have been classified in a different category for one of the four dimensions. This would not have made a substantial difference to the proportions of the papers that fall into each category.

At this point the system had been validated for paired classification. Our individual agreement on nature was poor when first measured but fair to good the second time; our individual agreement on theme was fair to good when first measured but poor the second time. Therefore we had at some point established fair to good reliability for individual classification and for paired classification on each of the four dimensions: the system had been validated for both individual and paired classification.

While we did not measure inter-rater reliability again, our next approach to establishing agreement (section 4.3) exposed each of us to prolonged discussions on differences in our classification; this led to an improved understanding as to how we as a group were interpreting the classification values, and thus presumably to an improved inter-rater agreement.

Inter-rater reliability measures have two principal purposes. One, discussed above, is to provide some sort of validation for a classification system. The other is to measure the reliability of a particular group of raters. Even if a system has been validated by inter-rater reliability measurement, this does not mean that any individual can use it reliably: there will generally be an assumption that people must be trained to use the system. In this case, the people who were trained to use Simon's system, and whose inter-rater reliability had helped to validate the system, were also confirmed as reliable classifiers.

A further three years of papers were divided among participants for individual classification, giving a total of eight years of NACCQ papers that had now been classified.

Publication P3 presents the method and results of this classification.

In line with previous classifications, programming and broad-based (here called multiple contexts) were the dominant contexts; however, hardware/ architecture and capstone project were also strongly represented, accounting for about 15% and 10% of the papers that had clear contexts.

The dominant themes were teaching/learning techniques and curriculum, with teaching/learning tools and educational technology also contributing substantially.

In line with prior results, subject was clearly the dominant scope. The relationship between scope and nature was examined to test the expectation that there might be a correlation, and it was found that papers of narrower scope (subject or program/department) are significantly less likely to fall into the research group of analysis or study (there were no experiment papers in this corpus).

Report was the dominant nature, but at 40% it was not a long way clear of the combined analysis (8%) and study (28%). Separating the eight years into three ranges, it was clear that study papers, and to a lesser extent analysis papers, showed a clear rise over the period, while report papers showed a comparable fall. This adds to the evidence of an increase in the proportion of research publications over recent years.

4.3 The Delphi method, Publication P4

The Delphi method (Powell 2003, Yetim & Turoff 2004) is a means of achieving consensus among a group of experts. The experts all send their initial classifications to a facilitator, along with brief reasons for their choices. The facilitator summarises the classifications and reasons and sends the summary to all members. Then follow several rounds in which the experts either agree to change their classifications or try to persuade others to change. This is all conducted anonymously, through the facilitator, so that no participant will concede to others because of their status or reputation.

Our next undertaking was to classify the papers that had appeared in all three years of ICER, the International Computing Education Research conference, using the Delphi method to try to achieve consensus. As project leader, I facilitated the process, and was thus aware of who was proposing which classifications. For the other five classifiers (one person had left the team), the process was anonymous.

As with paired classification, the number of disagreements, while never high, fell sharply after a single round, and then dropped more gradually over several rounds. In the context of meeting the deadline for a conference publication, the process was extremely time-consuming: it ran for seven rounds including the initial classification, at an average of nine days for a round. While it did eventually result in consensus at the level we had agreed in advance, four or more of the six classifiers agreeing on a single value, it is definitely not recommended when there are other viable approaches, such as classifying in pairs.

Publication P4 presents the process and the findings from this stage of the work.

One finding from the classification of these papers is that ICER is indeed a research conference rather than a computing education conference with some research papers included. Of the 43 papers in the corpus, only five (12%) were reports or position/proposal papers, with the remaining 88% being analysis, study, or experiment papers.

Programming was once more the dominant context, but with a substantially greater proportion than in previous corpuses: 74% of all ICER papers were set in the context of programming courses. This suggests that most of the research being conducted in computing education is focused on the perennial question of why many students find programming difficult to learn.

Unlike the other corpuses, the dominant themes were ability/aptitude and teaching/learning theories and models, at 26% each, with teaching/learning techniques close behind on 23%. This reflects the difference in preferred subject matter between computing education practitioners and computing education researchers: the research focuses not so much on how we teach as on the ability and aptitude and understandings of our students, and how this relates to theories and models of education.

4.4 Obstinate classifiers and alternative classifications

In both pair classification and the Delphi process, a sufficiently obstinate participant has the potential to cause problems. Assuming for a moment that there is such a thing as a 'right' classification for a dimension of a paper, it is possible for one or more proponents of that classification to give in to an opponent who is clearly unwilling to concede. It is hoped that if this were to happen, the number of cases would be very small, and would therefore have minimal impact on the overall findings of the classification.

A more frequent cause of disagreement is that in some instances there is not a single 'right' classification. Rather, there will be two or more categories into which a single paper can fall for one or other of the dimensions. Glass et al (2002) make the same point:

> "To promote consistent classification and analysis it was important that for every paper examined we select one dominant category for most of the five classifications. For example (and with certain exceptions), it was necessary to choose a single reference discipline, or a single topic, or a single research approach/method. Most SE papers touch on multiple matters within a single category, and it sometimes became an exercise in compromise to select the dominant category."

Once again, if the goal of this research were to unequivocally classify each individual paper, this could be a serious problem: there would be some papers that simply cannot be unequivocally classified in all dimensions. However, these papers would be very few in number, as indicated by the rapid approach to agreement described for pair classification in section 4.2 and for the Delphi process in section 4.3.

Therefore when the goal is to form a big picture of a complete corpus of papers, as opposed to an unequivocal classification of every paper, the variation due to such uncertainties is unlikely to make a substantial difference to the overall findings.

5. Further paired classification

Members of the AWPOCCELIPS team went on to classify corpuses of papers from several years of the ITiCSE conference and the SIGCSE technical symposium. Unfortunately, the submissions reporting on these classifications were rejected by those conferences, and because the work was so targeted to each conference there was no point in seeking alternative publication venues for the work.

The essential reason for the rejections is that reviewers did not see what impact the paper would have on the conference audience. Here are some excerpts from the reviews of the second SIGCSE submission:

- “I don't see the interest or significance of the paper.”
- “The topic has very little to contribute.”
- “This paper should be of great interest to SIGCSE organizers. However, I expect that many CS educators or SIGCSE participants will not find the paper so interesting.”
- “Since this paper is a paper about papers, it may have some effect on authors who are interested in publishing especially in SIGCSE. The reviewer can hardly see the effect beyond that.”

These comments sit well with the proposition that SIGCSE and ITiCSE are venues where computing educators can meet to swap classroom techniques and tools. The reviewers generally accepted that the work had been well done and the papers well written, but they could see nothing in the findings that would benefit computing educators – except perhaps “authors who are interested in publishing”.

Some interesting results from the work are briefly presented here.

The ITiCSE classification was of all 654 full papers published in the first 13 years of that conference. Pairs of classifiers classified sets of papers individually, then discussed their classifications to reach consensus. Thirteen years is the longest period that we had classified, and this corpus was thus well suited to looking for trends.

In keeping with our classifications of other conferences, the most prevalent context was programming, with 36% of the papers presenting work carried out in programming courses.

The dominant themes were teaching/learning tools (29%) and teaching/learning techniques (28%). The first of these is substantially higher than at other conferences we had classified, emphasising the swap-meet aspect of the conference. An observed trend over the lifetime of the conference was an increase in the number of papers dealing with ability/aptitude/understanding: these papers averaged 0.5% of the total in the first four years, 6% in the next five years, and 11% in the next four years.

The analysis of scope revealed no obvious differences from the other conferences classified, and no trends.

The nature dimension revealed some particularly interesting results. As shown in Figure 3, the proportion of research papers (analysis, study, and experiment) had climbed from a low start to 35%, a value that was maintained for the three years 2002-2004. The year 2005 saw the inception of ICER, a conference devoted to computing education research, and in that year the proportion of research papers at ITiCSE dropped sharply to 18%, although it did then gradually recover. ITiCSE also distinguished itself by having an average 18% of position/proposal papers, a higher proportion than at any other conference we had studied.

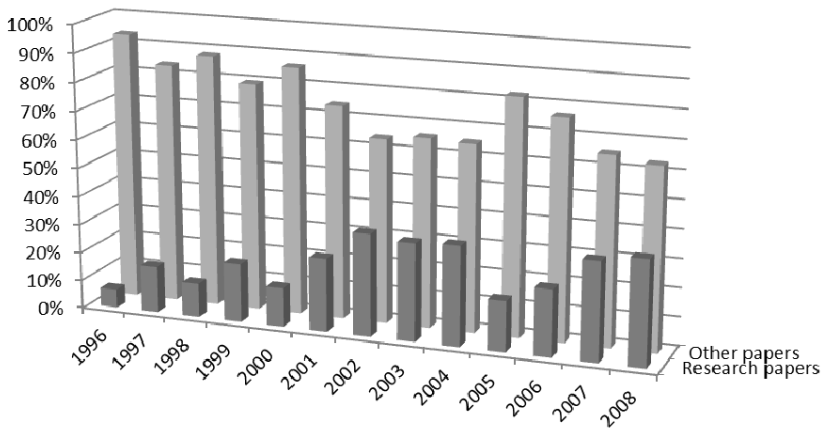


Figure 3. Proportion of research papers and other papers at ITiCSE over the years 1996-2008.

The SIGCSE analysis was of the 514 full papers accepted by that conference in the five years 2006-2010. Most of the findings were somewhat comparable to those from other conferences, but again the nature dimension provided the most interesting result. Figure 4 shows, over the five years, the proportion of papers classified in the combined analysis, study, and experiment categories, which together form the group that we are confidently calling research papers. In just five years the proportion of these papers has climbed steadily to three times its initial value.

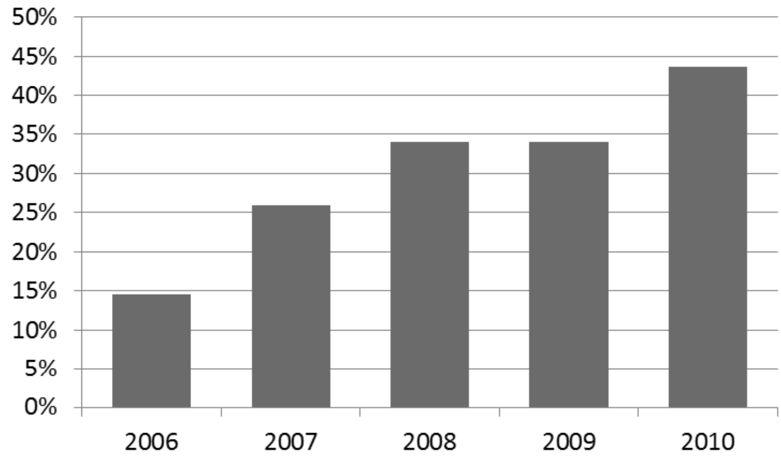


Figure 4. Proportion of SIGCSE papers in the research categories over the years 2006-2010.

There were now three years for which we had classified the papers at six different computing education conferences. Figure 5 shows the proportions of research papers at each of the conferences over the three years. Four of the conferences show a steady increase in the proportion of research papers. ICER has a very high proportion of research papers, and showed a small drop over the years in question. Koli Calling had experienced a rapid growth in the proportion of research papers over several years, and failed to maintain that in the third year examined here. While the research papers still comprised a healthy 46% of the accepted papers, the conference did take an exceptional number of discussion papers that year – 15 of the 28 accepted papers – and this would have contributed to the drop in the proportion of analysis, study, and experiment papers.

While three years is a short period in which to seek trends, it does appear that over those three years there was an overall steady increase in the proportion of computing education papers that can confidently be called research papers. For the conferences where we have classified more than three years, the observation of this trend appears to be supported, as for example in Figure 4, with the occasional slip such as that for Koli Calling in the third year of comparison, and that described above for ITiCSE in 2005.

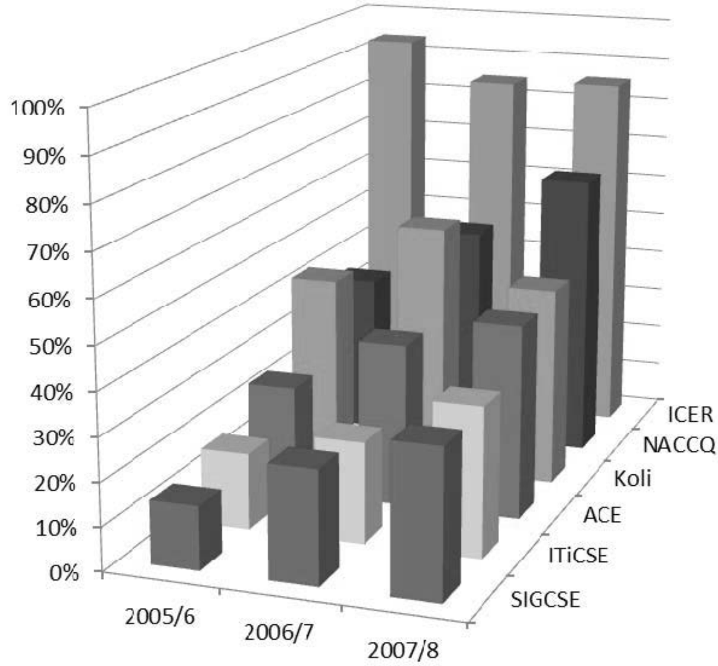


Figure 5. Proportions of research papers at six computing education conferences over three years.

5.1 Research questions RQ4 and RQ5

Research question RQ4 asks if publishing venues for computing education papers show an increase in the proportion of their papers that can be described as research. The analysis of six conferences, as illustrated in Figure 5 and elsewhere, certainly suggests that this question can be answered in the affirmative: yes, publishing venues for computing education papers show an increase in the proportion of their papers that can be described as research. It is true that ICER shows no such trend, but this would not be a reasonable expectation, given its overall high proportion of research papers.

At this point the classification has been applied just to computing education conferences. In section 7.6 we shall expand the picture to include computing education journals.

Research question RQ5 asks whether computing education has one or more research conferences and one or more research journals. Still looking only at Figure 5 and the proportions of research papers, it is clear that ICER is, as it was intended to be, a conference for the presentation of computing education research. SIGCE and ITiCSE, the other two conferences sponsored by the same organisation, have retained some of their earlier character, and have substantial proportions of report and position/proposal papers. The remaining conferences, arguably more regional in their catchments, fall between these extremes. As ICER is a research conference, one part of RQ5 has been answered in the affirmative; the second part will be considered in section 7.7.

The AWPOCCELIPS team had now classified 200 papers leading to two publications (NACCQ and ICER), and a further 1168 papers leading to three rejections (one from ITiCSE and two from SIGCSE). On the basis of its diminishing returns, the team decided not to undertake any further classifying of computing education papers, and disbanded.

6. Subsequent solo classification and bibliometrics

Measurement of inter-rater reliability serves two distinct purposes: it assesses how reliably the raters rate, but it also necessarily determines whether the system itself can be reliably used. Failure to achieve IRR values in the fair to good or excellent range might indicate unreliable raters, but might equally indicate an unreliable system. On the other hand, achieving IRR values in the fair to good or excellent range confirms that the system can be used reliably.

The IRR measurements of the AWPOCCELIPS group not only confirmed that the members of that group could classify reliably; it also confirmed that the system itself could be used reliably by members of that group. It thus validated any solo classification undertaken by members of that group.

While the group was still active I undertook two further sets of solo classification, at the same time adding a bibliometric study to each set. Bibliometrics is essentially a study of authors and authorship patterns. Where do the authors come from? How many papers do individual authors write? How many authors contribute to individual papers? What are the collaboration patterns? And so on. Bibliometric analysis is reasonably common in the library sciences, and has been applied to publications in numerous discipline areas including science education (Tsai & Wen 2005), engineering education (Williams & Neto 2012), crystallography (Behrens & Luksh 2006), accounting (Chung et al 1992), and bioinformatics (Patra & Mishra 2008).

Within the realm of computing education, Randolph et al (2005) noted that about 90% of Koli Calling papers to 2004 had first authors from Finland, and two years later I observed in publication P2 that there had been a clear increase in first authors from other European countries, the Americas, and Australasia: the conference was becoming more international. This was my first foray into bibliometric analysis. However, there was scope for more work in this area, and that work was applied in the next two publications.

6.1 A ten-year time span, Publication P5

For the 11th Australasian Computing Education Conference, held in 2009, I prepared a survey of the first ten instances of the conference (publication P5). (It is not strictly correct to call this a ten-year span, as the conference experienced some problems of continuity in its early years, and ran only ten times in 13 years.) The survey included some details about the conference itself, such as venues and chairs; a full analysis of all 328 papers published at the conference; and some analysis of the authors.

The author analysis in publication P2 would certainly not qualify as bibliometrics, being nothing more than a tabulation of the countries of the first authors. In publication P5 the analysis extended to all authors. It tabulated the number of authors per paper (from 1 to 21) and the number of papers per author (from 1 to 14), and noted the achievements of a dozen of the most prolific authors.

The publication analysis showed another variation on what was becoming a standard pattern: teaching/learning techniques as the dominant theme, subject as the dominant scope, programming as the dominant context, and report as the dominant nature. However, it did have one feature of particular interest. As a study of ten instances of the conference, conducted before the ITiCSE analysis shown in Figure 3, it offered a longer timeline than any of our prior studies and suggested that the trends over three years (as, for example, in Figure 5) were more sustained. Figure 6, grouping the natures of these papers into just two, research papers (experiment/study/analysis) and other papers (report/position/proposal), shows a steady increase in the proportion of research papers that is interrupted only twice in the ten years.

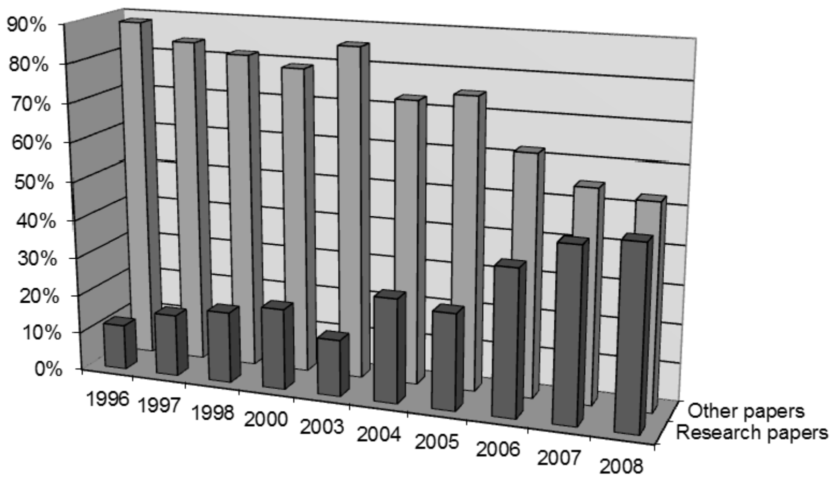


Figure 6. Increasing proportion of research papers over ten instances of ACE.

6.2 Extending beyond computing education, Publication P6

Publication P6 explicitly undertakes a comparison of two publication venues, the conference Koli Calling and the journal Informatics in Education, which have at times had a close relationship. The journal is not limited to computing education; it ranges over the many uses of computing in education, such as the development of technological innovations for education in a wide range of disciplines. Nevertheless, the journal has devoted two special issues to the conference – along with special issues for other conferences – so the conference and the journal are clearly not unrelated.

Koli Calling published 130 papers between 2001 and 2007, and Informatics in Education published 121 between its 2002 inception and 2007. These figures are highly comparable, adding to the similarity between the two venues.

It is no surprise that some of the journal papers have contexts not seen in the conference, contexts such as demography, medicine, science, and small business. However, the contexts not seen in the conference also include a number that appear specific to computing education, such as algorithms, human-computer interface, and operating systems. Conversely, there are contexts seen at the conference and not in the journal, but with the exception of study planning they can all be described as computing education contexts. Table 2 in publication P6 lists side by side the contexts found at each venue.

The themes are not greatly dissimilar between the venues. Teaching/learning techniques is the most prevalent theme in both, but in the journal online/distance delivery takes second place, followed by teaching/learning tools and assessment techniques. Ability/aptitude/understanding, which is often associated with research papers, ranks third at Koli Calling.

The venues display remarkable similarity in the natures of their papers. The combined proportions of experiment, study, and analysis papers are all but identical, giving totals of 37% and 38% of papers in the research grouping. Koli Calling has a slightly higher proportion of position/proposal papers than Informatics in Education, and it seems likely that this is because the conference invites discussion papers on work not yet undertaken.

The author analysis in this paper considered the proportions of authors who have contributed to just one paper, to two papers, and so on. The findings here appear to suggest that Koli Calling is more of a community than Informatics in Education. Nearly 80% of the journal's authors have contributed to just one paper, whereas only 65% of the conference authors have stopped at one. In the journal, the most prolific author has contributed to five papers, whereas the most prolific conference author has contributed to 13. By the same token, the journal authors come from a far greater number of countries than the conference authors. These findings are not surprising: a conference is by its nature more community-focused than a journal, especially a journal that sources some of its papers from a variety of conferences. Nevertheless, it is good to see that this sort of analysis is capable of establishing or confirming such findings.

6.3 Departure from computing education, Publication P9

Publication P9 reports the analysis of all six volumes of the journal Olympiads in Informatics. This was an interesting test for the classification system, as it was devised explicitly for computing education papers, and Olympiads in Informatics is not a computing education journal, nor even a journal on computers in education. The following paragraphs give an indication of how the system was adapted to analyse this very different corpus of papers.

Scope
Because the Olympiad competitions take place outside the realm of formal education, the defined scopes of subject, program/department, institution, and many institutions do not apply. However, a different form of scope was observed in the papers: many showed collaboration within a country, while a handful showed international collaboration. As the purpose of the scope dimension was to assess the breadth of collaboration, these differences were noted under that dimension.

Context
The context dimension refers to the subject matter being taught in the classroom in which the work is being carried out. The papers in this journal do not report on work in classroom education, so they do not have a context. Alternatively it could be argued that they all have the same context, programming. Under either interpretation, there is nothing to be gained by classifying the papers, as they will all fall into the same group.

However, once again a different form of context was evident in the papers: the country in which the work was done. The 101 papers in the six volumes come from 40 different countries, with eight papers arising from the international collaborations described under scope. This, therefore, is the context that was reported in this publication.

Topic (theme)
All of the papers effectively belong to a single theme in the original system, the theme of *competitions*. It was always acknowledged that the themes could be subdivided (for example, programming becoming game programming, graphics programming, introductory programming, etc), or indeed combined, and this journal required a subdivision of the competitions theme if there were to be any useful findings. With no prior guidelines as to what people were writing about in the Olympiads, the themes were derived directly from the data. The name 'topic' was revived for the dimension, as there were no classroom topics for it to be confused with. Only six topics were found in the full corpus: organisation of the events; tasks used in the events; grading of those tasks; preparation of students for an Olympiad; infrastructure for the events; and impact of the Olympiads outside their immediate sphere.

Nature
Preliminary analysis of the papers showed that they fell predominantly into just two categories of nature: report and analysis. One paper explicitly gathered data for the analysis, which would classify it in the study category, but I decided that a single paper in this category was not worth counting separately; instead I explicitly mentioned that this paper gathered its data expressly for the analysis.

With only two of the five categories represented, I chose to split the software reports from the other reports. In its second volume the journal had described itself as presenting research and development related to informatics competitions, and while development can have many interpretations, it tends to sit well as a description of new software.

This left the nature dimension with three categories: report, software, and analysis. While the analysis papers were too few (22%) to discern a clear trend over the six years, their numbers did appear to be increasing, and the paper encouraged this pattern.

The adaptation of the classification system to a markedly different set of papers showed that in addition to being suitable for classifying papers in computing education, it stands as a framework that can be adapted to produce informative results when classifying other types of publication. The dimensions continue to make sense, while their values need refining to classify different types of publication.

This particular use of the classification system has not been validated because there was only one classifier, and thus no measurement of inter-rater reliability. The paper is thus in the same position that publications P1 and P2 were in when they were published. However, the system used for those publications was subsequently validated, and it would seem reasonable to infer that this modified system could also be validated.

6.4 Bibliometrics and Lotka's law

Lotka's Law of author productivity (Nicholls 1989) is of particular interest to the subject of this thesis, because it describes a property of an established discipline. Arising from empirical observation, the law suggests that in a sufficiently large list of published papers within a discipline, 60% of authors will contribute to only one paper, 15% to two papers, 7% to three papers, and so on according to a specified power function in which the exponent is generally close to 2. Having explored the relationship for a number of distinct data sets, Nicholls suggest that single-paper authors are more likely to make up 71-81% of the total pool of authors.

Publication P9, discussed above, applied Lotka's Law to the authorship on the journal, and found a comfortable fit, as shown in Figure 7, with the power constant at exactly 2.0. The conclusion was that even after just six years of publication, the authorship fitted a pattern that was based on a sufficiently large list of papers within a discipline; and therefore that the authorship satisfied one posited criterion for a discipline.

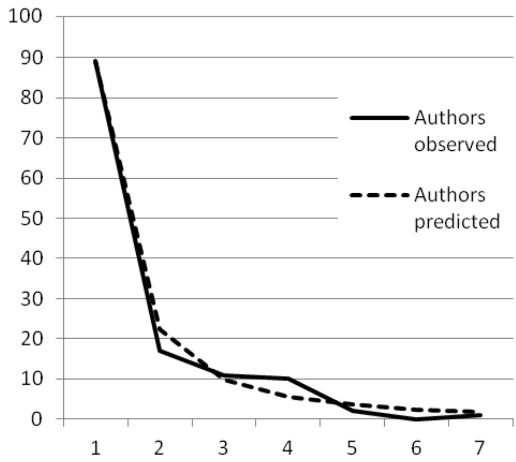


Figure 7. Authorship of Olympiads in Informatics, observed and predicted by Lotka's Law.

This prompted the question of how well the authorships of computing education publications would meet the same criterion. Only one analysis of this sort was carried out, for a bibliometric analysis of the ITiCSE conference that was never accepted for publication. Although ITiCSE had been running for 13 years, seven years longer than Olympiads in Informatics, its authorship pattern appears to be further from that expected of a discipline. First, the once-only authors make up 79.5% of all authors, very much at the high end of the range observed by Nicholls (1989). Second, to achieve even a passable match, the power constant needs to be about 2.65, which means a greater decline in the number of authors with more than a few papers. Third, even with such a high value for the power constant, the observed number of authors with six or more papers simply collapses in comparison with the prediction. Table 4 shows the observed and predicted numbers for ITiCSE, showing that, as a single venue, it does not appear to satisfy Lotka's Law.

Table 4. Observed and predicted numbers of authors contributing to specified numbers of ITiCSE papers 1996-2008, using Lotka's law with C=79.5% and p=2.65.

Contributions	Authors observed	Authors predicted
1	891	891
2	142	142
3	48	48
4	22	23
5	12	13
6	1	8
7	1	5
8	2	4
9	1	3
10	0	2
11	0	2
12	1	1

On the other hand, a more recent analysis of the papers at ICER, the research conference, shows a much better fit, along with a lower proportion of single-paper authors (66.4%) and a lower power constant of 2.1. The numbers are shown in Table 5. This could be taken as suggesting, according to this single criterion, that the computing education research community, as represented by ICER, is more like a discipline than is the broader computing education community, as represented by ITiCSE.

Future work is expected to include similar analysis of the other recognised computing education publication venues.

Table 5. Observed and predicted numbers of authors contributing to specified numbers of ICER papers 2005-2014, using Lotka's law with C=66.4% and p=2.1.

Contributions	Authors observed	Authors predicted
1	174	174
2	36	41
3	20	17
4	11	9
5	4	6
6	6	4
7	3	3
8	1	2
9	2	2
10	1	1
11	0	1
12	1	1
13	1	1
14	1	1
15	1	1
16	0	1
17	0	0
18	1	0

7. Extending the scheme to examine research approaches

7.1 Research approaches of research papers in programming, Publication P7

Associate Professor Judy Sheard, who was a member of the AWPOCCELIPS project from its start to its end, was interested in delving more deeply into the research approaches of papers reporting on research in programming. This work was carried out by Sheard, Simon, one other member of AWPOCCELIPS, and one new member. We began with the 979 papers published at ICER, SIGCSE, ITiCSE, ACE, Koli Calling, and NACCQ in the four years 2005-2008. Most of these papers had been classified by the AWPOCCELIPS team, and the rest by Simon alone, so it was a simple matter to select just the papers with a context of programming, the prevalent context at each of the conferences, giving 349 papers; and then to select the papers with natures of experiment, study, or analysis, giving 164 papers reporting on research in programming contexts.

With these papers we explored the research method; specifically:

- whether their methodology is qualitative, quantitative, or mixed;
- what data gathering techniques they use (eg test, interview, survey);
- what data analysis techniques they use (eg phenomenography, statistical inference, etc);
- what aims they express;
- what outcomes they report.

This represented a new direction in the work. Rather than classifying complete corpuses of publications and reporting the findings, we were using prior classification to identify papers of interest, then examining those in a way that goes well beyond the classification system.

The bulk of this work was counting and reporting – for example, whether a project used exam results, surveys, interviews, validated inventory instruments, etc – which is substantially less subjective than the classifying. Unlike the classification using Simon's system, we did count multiple items in some categories. For example, in the 164 papers that we examined we identified 338 data-gathering methods. For the counting and reporting work we perceived no need for a formal assessment of inter-rater reliability. Rather, we jointly classified a small set of papers, then when we were confident that we were working similarly, partitioned the remainder of the papers and assessed them individually.

The other aspects of the work – aims and outcomes – would have been easier to list if they had always been clearly expressed in the papers; however, many authors do not clearly express the aims of their work, and some do not clearly express their outcomes, so a great deal of variation was found in our reports of these aspects. Feeling that it would be infeasible to train all four of us to the point where we would be able to report comparable aims and outcomes for the same paper, we chose instead to have a single person extract the aims and outcomes of all of the papers. While consistency is not guaranteed even with a single researcher, it is far more likely than with a group of researchers who are known not to be extracting the data consistently.

As the papers were selected on the basis of the nature and context dimensions, we also reported on their themes and their scopes.

Publication P7 reports on this work. Perhaps the most salient finding is a confirmation of the earlier observation that most papers presented at ICER are research papers in the context of programming. As shown in Table 6, 69% of the ICER papers over those four years were identified as being in the research grouping of nature and the context of programming. Next was ACE, with 18% of such papers, closely followed by ITiCSE and Koli Calling with 16%. This shows again how different ICER is from the other computing education conferences.

Table 6. Percentage of papers from each conference, 2005-2008, identified as experiment, study, or analysis papers in the context of programming.

Conference	2005	2006	2007	2008	Overall
ICER (N=59)	88%	85%	43%	63%	69%
ITiCSE (N=249)	13%	13%	21%	18%	16%
SIGCSE (N=415)	9%	7%	9%	18%	11%
ACE (N=99)	9%	31%	10%	22%	18%
Koli Calling (N=87)	16%	18%	18%	11%	16%
NACCQ (N=69)	11%	8%	7%	11%	9%

In terms of theme, teaching/learning techniques was well represented, as always, but even when combined with assessment techniques it was no match for the 40% of papers on the theme of ability/aptitude/understanding: in somewhat simplified terms, the challenge of working out why some students find it so difficult to learn programming.

In terms of scope, work based in single subjects was still the highest, at 54%, but multi-institutional work was impressively high at 26%. Many people conducting research based in programming courses do so in collaboration with researchers from other institutions.

Publication P7 reports too many other interesting findings to list them all here, but the work definitely served to present a clear picture of the sort of research being carried out in the context of programming and presented at these six computing education conferences.

7.2 The socially defined nature of research – again

All of the work discussed to date has applied the convenient definition of research as work falling into the experiment, study, or analysis category of Simon's system. It was acknowledged that research papers can certainly be found outside these three categories, but they were adopted as a pragmatic way of measuring the proportions of papers that can readily be seen as research, and comparing those proportions across time and venues.

Authors such as Holmboe et al (2001), Dale (2002), Valentine (2004), Fincher and Petre (2004), and Lister (2006), all suggest that there is a distinction between research and other papers, and either call for an increase in the proportion of research papers or observe that the increase is already under way. For the classification work that has been reported here, it was necessary to have some way of partitioning publications into research and other. Nevertheless, there should be no illusion that any single set of criteria can definitively accomplish this partitioning.

Holmboe et al (2001) refer to the criticism that computing education research is merely a way for teachers to write papers. Yet might there be some truth in that criticism? Are there circumstances in which all computing education publications, including practice papers, can be considered as research?

A number of projects involving classification of publications work on the premise that all published papers represent research in some form. Joy et al (2009) make the premise explicit: "our use of the term 'Computer Science Education research' is inclusive of the many practice-based papers that some might feel should not be considered under this label." Glass et al (2004) are not so explicit, but in choosing to examine the research approach of every sampled paper in each corpus they appear to be considering all of these papers as research papers; otherwise they might have chosen, like Randolph et al (2005) and Sheard et al (publication P7), not to analyse papers that they did not consider to be research.

The next phase of the classification work reported here has its basis in this convention that all published papers are in some sense research papers.

7.3 Theoretical foundations

Professor Lauri Malmi had long been interested in the nature of computing education research, and was a member of the team that explored the identification of a core literature for computing education research (Pears et al 2005). He and Professor Sheard spent some time working together on better ways of exploring both the methodologies used in CER and the theoretical frameworks, if any, used as the basis for the research.

Many of the definitions of research canvassed in chapter 2 include the notion that research should be based on theoretical foundations. Fincher and Petre (2004) observe that "CS education research . . . is theory-scarce" and devote a whole chapter to linking research to relevant theory. Holmboe et al (2001) argue that "the future work of CSE must have a stronger connection to the theoretical frameworks of education-related disciplines". Joy et al (2009) observe that "a large proportion of papers do not address issues of educational theory".

With this in mind, the questions to be explored by Malmi and Sheard were what proportion of computing education papers make appropriate use of relevant theories, and when they do, from what disciplines those theories come. These questions are somewhat related to those addressed by Vessey et al (2004), who explored the research processes evident in publications, so Malmi and Sheard began by examining the classification system devised by Vessey et al and adapting it as required to permit classification of computing education papers.

This gave rise to a system that came to be known as TMMCER (Theories, Methods, and Models in CER), although it was not called that in its first publication. The system classifies publications according to the following properties, each of which is explained more fully in publication P8.

- Theory/Model/Framework/Instrument. This dimension shows linkage to prior work by listing theories, models, frameworks, and or instruments (for example cognitive load theory, situated learning, threshold concepts) used or enhanced in the work.
- Technology/Tool. Any technology or tool (such as BlueJ, Jeliot, Peerwise) that is an integral aspect of the work being reported.
- Reference discipline. The originating discipline of a TMFI, technology, or tool.
- Research purpose (called research approach by Vessey et al). This dimension has three top-level values: descriptive, evaluative, and formative. Each is then further subdivided; for example, a paper describing a software tool would be classified as descriptive – technical system.
- Research framework. Recognised frameworks for research projects, such as action research, the Delphi technique, phenomenography, and constructive research (designing and building some artefact).
- Data source. The source of the data and how it was gathered. For example, naturally occurring data such as students' grades, data gathered specifically for the research, and reflection (the researcher's own experiences).
- Analysis method. What forms of analysis were used upon the data. These include interpretive classification (such as the application of Simon's system), conceptual analysis (breaking down concepts into their constituent parts), various levels of statistical analysis, and argumentation (for papers that show no other form of analysis).

I was fortunate to be able to attend a workshop at Aalto University (then Helsinki University of Technology) at which the team worked to refine the system, and I was invited to take part in the classification of a large body of papers. Like the work of Vessey et al, this work sought to classify research features of all papers in the corpus, and thus made the implicit assumption that all papers are research papers. Simon's system, on the other hand, partitions papers into a group that are clearly research (experiment, study, and analysis papers) and others (report and position/proposal papers).

The reconciliation of these apparently conflicting concepts lies in some of the definitions of Vessey et al, which have flowed through to the TMMCER work. Consider, for example, the three broad groups of research purpose. Evaluative work evaluates something, and therefore presumably collects and analyses data in the course of that evaluation. It might thus correspond with experiment, study, or analysis in Simon's system. Descriptive work might involve collection and analysis of data, but its principal purpose is clearly the description of a classroom event, a software system, or some other item. It is therefore likely to be classified as report under Simon's system. Formulative work, formulating and proposing a concept, a model, a process, or a set of standards, is likely to be classified as position/proposal in Simon's system, notwithstanding that some formulative papers would be generally recognised as embodying research.

It follows that papers classified as reports or position/proposal papers in Simon's system can nevertheless be assigned one or more research purposes in the TMMCER system. They can equally well be assigned other research properties; for example, a data source of reflection and an analysis method of argumentation. In general, papers with a descriptive research purpose, papers with no research framework or with a simple research framework, papers with no data source, or with the data source classified as reflection, and papers with an analysis method of argumentation, would all potentially fall into the nature categories of position/proposal or report.

The papers chosen for analysis were those published over seven years in the research conference ICER and the two principal computing education journals, Computer Science Education and ACM Transactions on Computing Education (formerly known as Journal of Educational Resources in Computing). These were chosen because they permit longer papers than many conferences, and are thus more likely to discuss the research aspects under consideration.

7.4 Publication P8

In publication P8, all seven dimensions of the new classification system are applied to the papers published in the first five years of ICER. The paper's findings are summarised briefly below, and are presented and discussed in more detail in the paper.

Theories, models, frameworks, instruments

Authors do not always explicitly indicate the theories, models, or frameworks on which their research is based, so it is not always easy to establish the presence of these foundations. In some 60% of the papers we found clear use of such theories or frameworks as cognitive load (Paas et al 2003), Bloom's taxonomy (Bloom et al 1956), the SOLO taxonomy (Biggs 2003), and threshold concepts (Meyer & Land 2005).

Technologies and tools

About 17% of the papers made specific use of particular software tools such as BlueJ (www.bluej.org), Jeliot (cs.joensuu.fi/jeliot), or Peerwise (peerwise.cs.auckland.ac.nz).

Reference discipline

If a paper is found to have used theories, models, frameworks, instruments, technologies, or tools, the reference disciplines are the disciplines from which those items originated. The principal reference disciplines found in these papers were education, psychology, and the broader discipline of computing (as opposed to computing education). A further six disciplines were represented by one paper each, showing that computing education research draws upon the work of many other disciplines.

Research purpose

We did not restrict each paper to a single research purpose, preferring instead to list all that we could confidently establish. Most papers (86%) included some form of evaluative research purpose, typically analysing empirical data for the evaluation. Formulative research purposes were evident in 36% of the papers, and descriptive in 22%. All of the papers with descriptive purposes had additional purposes: we found no purely descriptive papers.

Research framework

Many recognised research frameworks were applied in the ICER papers. The common ones were survey (39%), experiment (15%), constructive research (14%), and grounded theory (13%). Papers also applied phenomenography, phenomenology, and the Delphi method.

Data source

While naturally-occurring data such as students' artefacts and results offer much potential for educational research, only 8% of the papers relied solely on such data. The majority (79%) gathered data specifically for the purpose of the research, and 16% used data from both naturally-occurring and research-specific sources.

Analysis method

The papers displayed a good mix of qualitative and quantitative analysis methods. The qualitative analysis included classifications according to existing systems or systems developed in the research being reported; for the quantitative analysis, 11% of papers reported only descriptive statistics (mean, median, etc), but nearly 60% undertook some more detailed statistical analysis. In 17% of the papers the principal analysis method was argumentation, which might be considered somewhat less rigorous than the recognised qualitative and quantitative methods.

7.5 Publication P10

The pool of papers was expanded substantially for publication P10. ICER was still considered, but we also included the two principal computing education journals, Computer Science Education (CSE) and ACM Transactions on Computing Education (TOCE). As ICER had now been running for seven years, we analysed the papers from all three venues over that seven-year span. The expansion of the corpus necessitated a reduction in what we could report in a paper of fixed length, so we reported only the theories, frameworks, and models found in these papers and the reference disciplines from which they are drawn.

As with the classification of ICER alone in publication P8, nearly half of the papers examined did not build on identifiable theories, models, or frameworks (TMF); however, many of those are actually formulating new TMF, adding to the number of these that originate in computing education. Examples are Jadud's error quotient (Jadud 2006) and Hazzan's process-object duality theory of learning (Hazzan 2003). The proportion of papers with at least one TMF is over 50% in CSE and ICER, and in TOCE has climbed to over 50% during the period under consideration. Of the remaining papers, some undertake empirical analysis of data, but with no acknowledgement of a theoretical foundation for their work, and a few simply report on novel technical contributions or instructional methods.

The paper goes on to make recommendations for future work that will grow new theories and, where appropriate, combine and unify the many existing theories currently in use.

7.6 Simon's classification of the journals

In the course of the classification described in this chapter, all of the papers were classified according to Simon's system as well as the TMMCER system. For the purposes of this thesis, that classification has been extended to the end of 2014, giving a 10-year span for each journal. As with the classification of ACE papers in publication P5 (section 6.1), it was hoped that this span would be sufficient to show any trends.

Figure 8 shows the proportions of research and other papers for CSE, and Figure 9 shows the proportions for TOCE. Because the journals publish only about 15 papers each year, small numerical differences can appear substantial; to help smooth these differences, the figures plot each pair of years rather than each year.

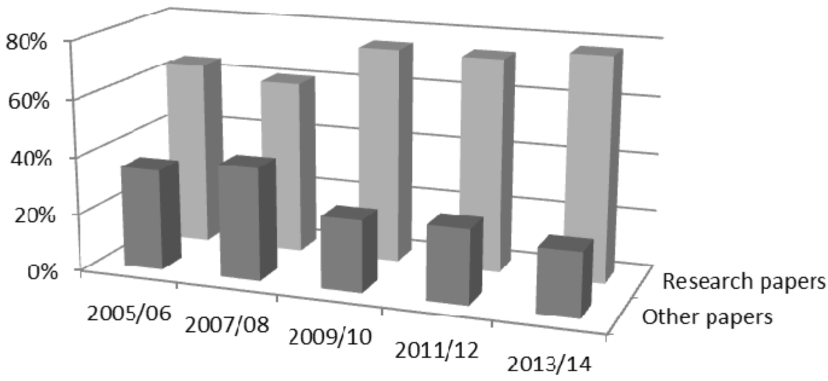


Figure 8. Proportions of research papers over ten years of the journal Computer Science Education.

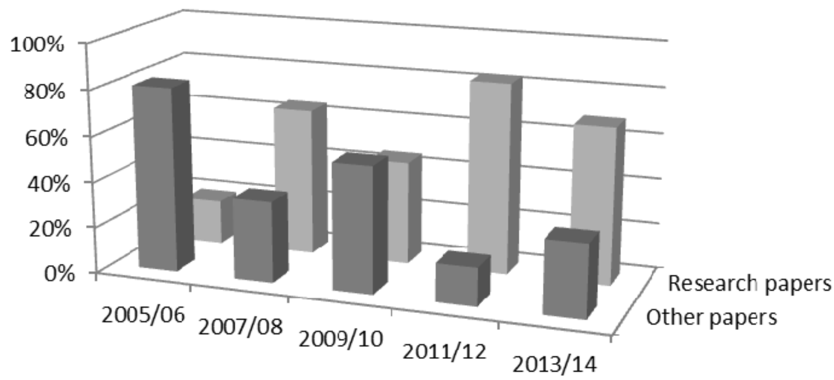


Figure 9. Proportions of research papers over ten years of the journal ACM Transactions on Computing Education.

Throughout the period being examined, the papers in CSE have comprised at least 50% research by the definition of convenience being applied in this thesis.

TOCE averaged about 40% research over the first six years of the ten, but in the following four years papers classified as report or position/proposal have become rare. Until 2008 this journal was called Journal on Educational Resources in Computing; while it included empirical research papers, its focus was on the presentation of new educational resources – tools. In 2009 the name changed to ACM Transactions on Computing Education, and the editorial for the first edition of the new journal explained the increasing importance of “a scholarly approach to teaching and learning” (Tenenberg & McCartney 2009). Editorials in subsequent issues offered guidance on how to undertake computing education research (Tenenberg & McCartney 2011a) and explicitly encouraged activities that can transform a tools paper into an empirical research paper (Tenenberg & McCartney 2011b). When classifying papers in these later years it was interesting to observe that many of those classified as reports were on the borderline with analysis. The definition of nature (as in the appendix) includes the observation that “even if there is some analysis, such as of student feedback, a paper will be classified as a report it if appears that its purpose is to describe what was done”. Many of the reports in these years did include analysis, but were ultimately classified as reports according to a subjective interpretation of whether their purpose was to present the system or the analysis.

7.7 Research questions RQ4 and RQ5 again

Research question RQ4 asks if publishing venues for computing education papers show an increase in the proportion of their papers that can be described as research, and RQ5 asks whether computing education has one or more research conferences and one or more research journals. In section 5.1 these questions were answered in the affirmative for computing education conferences; now they can be addressed for computing education journals.

ACM Transactions on Computing Education shows a clear increase in the proportion of research papers over the ten years studied. Especially given the small numbers, it is not so easy to discern a trend in this measure for Computer Science Education. However, as with ICER (chapter 5) such a trend can hardly be expected if the proportion is consistently high.

Looking just at the last few years, it is clear that papers classified as research make up well over half the papers published by each journal, and therefore that both can be called research journals.

Both RQ4 and RQ5 can now be considered as answered in the affirmative.

8. Computing education as a research discipline

8.1 Fensham's criteria

As explained in section 1.1, Fensham (2004) proposes a number of criteria for determining that science education has become a field of research. It is important to understand that these criteria were developed post hoc: Fensham asserted that science education was a field of research, and identified a number of features of science education that appeared to support the assertion. There is no single external definition of a field of research by which Fensham's criteria can be evaluated. Rather, the criteria are assessed in the same way as many other research endeavours: by the extent to which they are accepted by other researchers in the area. A number of publications have referred to and applied Fensham's criteria; for example, a report on discipline-based education research prepared for the US National Academy of Sciences (Singer et al 2012). On the other hand, we have found no publications that challenge the criteria. In that sense, they can be considered as accepted by the science education and related communities.

However, because the criteria are post hoc, they must be considered as descriptive rather than prescriptive. While an area of research can lay reasonable claim to being a field or a discipline if it satisfies all of the criteria, it is not necessarily excluded from that status if there are one or two that it fails to meet.

With that in mind, we shall now consider computing education through the lens of Fensham's criteria, grouped as in Fensham (2004).

8.2 Structural criteria

Academic recognition

The academic recognition criterion is satisfied if there are full professorial appointments explicitly in computing education. This criterion is potentially difficult to satisfy, as there are many professors whose research is almost entirely within computing education but whose appointments do not specify that field. Nevertheless there are many full professors as specified, particularly when considering the other names by which computing is known. Some examples follow.

- Sally Fincher, Professor of Computing Education, University of Kent, UK
- Quintin Cutts, Professor of Computer Science Education, University of Glasgow, UK
- Mordechai Ben-Ari, Professor of Computer Science Education, Weizman Institute of Science, Israel
- Carsten Schulte, Professor of Computer Science Education, Free University Berlin, Germany
- Heuiseok Lim, Professor of Computer Science Education, Korea University
- Ivan Kalaš, Professor of Informatics Education, Comenius University, Bratislava, Slovakia
- Said Hadjerrouit, Professor of Informatics Education, University of Agder, Norway

Another aspect of academic recognition mentioned by Fensham is the awarding of doctoral degrees in computing education. It follows almost automatically that where there are professors of computing education there will be students undertaking and completing postgraduate degrees in computing education, and indeed that is the case. To take one obvious example, this thesis is expected to lead to the award of a doctorate in computing education, one of at least a dozen produced at the same institution under the guidance of the same supervisor. And, of course, there are many other universities around the world where students are earning doctorates in computing education.

Research journals

There are journals that publish research papers in computing education but are not dedicated to computing education. These include the Journal of Engineering Education and the IEEE Transactions on Education. However, Computer Science Education and TOCE are both journals that seek, accept, and publish papers exclusively in computing education. The analysis reported in section 7.6 affirms that both are research journals. Therefore computing education does have research journals.

Professional associations

ACM is a professional association for members of the computing profession, and IEEE is a professional association for electrical and electronics engineers that has expanded to include computing professionals. Neither of these is exclusively for computing educators, but both have branches or divisions for computing educators. ACM's many special interest groups include SIGCSE, the Special Interest Group for Computer Science Education. IEEE has a Computer Society, a separate professional society under the umbrella of the larger organisation, and that society has education as one of its six program areas.

Both SIGCSE and the IEEE CS can be joined independently of their parent associations. Therefore SIGCSE is a professional association for computing education, and IEEE CS is a professional association of which a large part is devoted to computing education.

While both ACM and IEEE are based in the US, their membership is world-wide. There are also many national organisations for computing professionals, but it is not clear that any of these organisations has sub-organisations specific to computing education. However, SIGCSE also has the capacity to operate at a national level, and currently has chapters in Australasia, Turkey, India, and Spain.

It appears reasonable to conclude that computing education has professional associations. However, Fensham's examples make it clear that he is referring to professional associations for educational research, not just for education. There are no formal computing education research associations. The closest approaches are the research community within the associations mentioned above, and a mailing list that arose from a number of research workshops (see *Research training* below).

Research conferences

One of the goals of the analysis presented in this thesis was to distinguish between practice papers and research papers in computing education. While the distinction was never intended to be prescriptive, it has served as a reasonable way to partition large numbers of published papers.

Using this partitioning, it is clear that ICER is a computing education research conference; that Koli Calling and ACE show the potential to be largely computing education research conferences; and that SIGCSE and ITiCSE, while remaining true to their traditional role of reporting on practice experiences, are steadily increasing the proportion of their papers that can be considered as research.

Based on these observations, there are indeed computing education research conferences.

Research centres

There are two distinct meaning of the phase 'research centre'. One refers to centres funded by external grants and devoted entirely to research. It would be somewhat surprising if such research centres could be found in computing education, because their presence would suggest that research into education could usefully be divorced from the education itself.

The second meaning of research centre is an academic location at which much research is conducted by a number of academics and their postgraduate students. Fensham makes it clear that this is the type of research centre that he has in mind, and there are many centres of this sort in computing education, including, for example, Aalto University in Finland, Uppsala University in Sweden, the University of Kent in the UK, and Georgia Institute of Technology in the USA.

Research training

Computing education has hosted a remarkable venture in research training. Josh Tenenberg in the USA secured a grant from the National Science Foundation to train a small core of novice computing education researchers, form them into a community of practice, and seed them into the wider computing education research community (Fincher & Tenenberg 2006). To this end, Tenenberg facilitated an intensive workshop in the USA that was co-led by Sally Fincher and Marian Petre. A second grant led to a second workshop in the USA, with the same leaders. Further workshops flowed from these: for example, a third workshop of the same type in Australasia, conducted by Fincher and Petre and facilitated by Raymond Lister and Anthony Robins (Simon et al 2006); and workshops on a particular qualitative research method, phenomenography, conducted in Australia (Lister et al 2007) and the UK (Berglund et al 2008).

These initiatives together inducted more than 50 computing educators into computing education research. Research training is often thought of in terms of postgraduate students, but here we have the academics themselves being trained, and a number of the ‘students’ at these workshops have gone on to become leaders in computing education research. Furthermore, these educators, the community of practice established by the workshops, all subscribe to a CER mailing list on which they continue to discuss issues of interest to CER. This is the community that was alluded to above as a possible equivalent of a professional association for computing education researchers.

There is also clear evidence of the more traditional research training for postgraduate students. The first SIGCSE doctoral consortium was held in conjunction with the SIGCSE conference in 1998, and this association continued until the inception of ICER, which has hosted the consortiums since 2008. Koli Calling and ACE also host doctoral consortiums in computing education, although not on a strictly annual basis.

8.3 Intra-research criteria

Domain knowledge

Fensham calls this criterion ‘scientific knowledge’, but this is in the context of charting the emergence of science education as a research discipline. His explanation makes it clear that he is referring to knowledge of the domain (science in his case, computing in ours) that is integral to the education research. There must be facets of computing education research that are particular to computing, and that could not reasonably be carried out by general education researchers.

Satisfying this criterion is almost trivial. One of the major problems addressed by computing education research is the difficulty that some students have in learning computer programming. Our results have established that this is the dominant content not just of computing education papers but of computing education research papers. This difficulty is not a general learning difficulty; it is a difficulty specific to programming, and it would not be possible to conduct this research without an intimate knowledge of the subject matter. Therefore computing education research definitely requires domain knowledge.

Asking questions

The criterion of asking questions applies not just to domain-specific educational research but to all research. Even if the question is as simple as 'what will happen when I do this?', research is driven by questions.

Not all CER publications have made their questions explicit, but they are increasingly doing so. Most publications now include a section, close to the beginning, in which they clearly indicate the questions that the research is designed to answer.

Conceptual and theoretical development

With his third intra-research criterion Fensham calls for the emergence of models with some predictive or explanatory power. The CER discipline is still young, and is not inundated with such models; however, two examples should suffice to confirm their existence.

The Australasian BRACElet project conducted wide-ranging research into programming education by way of assessment items (both research-specific and naturally occurring) and the students' responses to those items. In one particular facet of the research (Lopez et al 2008), a clear hierarchy of programming-related tasks was established: the ability to trace code is a precursor to both reading code and writing code, and the ability to read code is also a precursor to writing code. On reflection this should not be surprising: in natural languages such as English and Finnish it is generally understood that one must be able to read before learning to write. However, the focus of programming education is very much on teaching students to write code, and the underlying assumption is that if in the process students are shown enough code, they will learn to read it. As well as establishing the hierarchy, BRACElet research showed convincingly that many students cannot read code, or at least that they cannot explain it; and it should therefore come as no surprise that they cannot write code.

In another project, on a somewhat smaller scale, Teague and Lister (2014a, 2014b, 2014c) have taken neo-Piagetian theory from general education and adapted it to computing education, devising what might be one of the best explanations to date of how students learn to program.

These are not the only examples, but they do serve to show that computing education research is producing models with the power to predict and explain the behaviour of students undergoing computing education.

Research methodologies

As we discovered when undertaking the analysis for publications P8 and P10, it is by no means clear what should be considered a research methodology. Fensham himself describes this criterion somewhat broadly, referring to “the invention, development or at least adaptation of methodologies, techniques, and instruments that have particular use for science education researchers as they endeavour to answer their specific research questions”. Accordingly, for the purpose of addressing Fensham’s research methodologies criterion we shall consider publications that use a TMFI (theory, model, framework, or instrument) with a reference discipline of CER; that is, papers relying on one of these items that has been developed within computing education research.

Of the papers that we examined, some 10% had TMFI with a reference discipline of CER. About half of these were classification systems, such as Simon’s system or the engagement taxonomy for algorithm visualisation (Naps et al 2002). Contributing student pedagogy has its origins in behavioural science (Collis & Moonen 2001) but in computing education it has taken on a new life (Hamer et al 2008). Particularly through the software product Peerwise (Denny et al 2008), it has spread to a large number of other education domains (https://peerwise.cs.auckland.ac.nz/). There are also theoretical constructions such as Jadud’s error quotient (Jadud 2006) and Hazzan’s process-object duality theory of learning (Hazzan 2003). Fincher et al (2011) explicitly note the need for the discipline to adapt and devise its own research methods, and propose two new research approaches called ‘my programming week’ and ‘emotional timelines’.

As well as making wide use of methodologies from other relevant disciplines, computing education research is definitely giving rise to its own research methodologies, frameworks, and instruments.

Progression

With his criterion of progression, Fensham refers to studies that build upon other studies to jointly progress understanding within the field. One of the strongest examples here comes from the dominant context of computer programming.

The ITiCSE conference is well known for its working groups. Group members typically start work well before the conference, obtaining ethics clearance from their institutions if required, gathering data, and perhaps beginning the analysis. At the conference they work intensively on analysis and writing, and shortly after the conference they submit their finished paper for review.

One working group at ITiCSE 2001 was ‘a multi-national, multi-institutional study of assessment of programming skills of first-year CS students’ (McCracken et al 2001). This group established convincingly that many students cannot write relatively simple computer programs at the end of their first programming course. Individual academics had long known this of their own classes, but this study showed that the problem was global, not restricted to isolated cases.

Three years later, the ‘Leeds working group’ (Lister et al 2004) built upon that study to show that many students at the end of their first programming course cannot read relatively simple code (in the sense of explaining what it does) or trace it (deduce its outputs from given inputs).

The BRACElet project (Lister et al 2006) delved far more deeply into these concerns, studying just what it is that students could and could not do, inviting other researchers to join in the investigation, and proposing ways that the problems might be addressed. When the project had officially ended, Lister and Edwards (2010) reported that it had led to the publication of “over 20 papers, with 28 academics as authors, from 20 tertiary institutions across seven countries.” However, that was not the end: work continues in the area, and many subsequent publications cite the project as their starting point.

There is clearly evidence of progression in computing education research.

Model publications

Fensham’s criterion of ‘model publications’ requires the existence of papers that other researchers will refer to when undertaking a particular type of study.

Tenenberg and Fincher (2007) introduced the idea of a ‘disciplinary commons’, a means by which teachers of computing (or of any other discipline) can meet to share their experiences and resources. While some would see this as teaching rather than research, the very act of sharing materials that are generally kept private tends to lead to ideas that will develop into research studies based upon the shared material. Adoption of the idea by other researchers (eg Morrison et al 2012) led to the production of a handbook (Fincher & Tenenberg 2011) that others will now follow when establishing disciplinary commons of their own.

Other papers are used as guides when embarking upon particular types of study (Fincher et al 2005 for multi-institutional, multi-national studies) or applying particular techniques (Publication 4 of this thesis for the Delphi technique).

Seminal publications

Fensham defines seminal publications as papers with important results that potentially change people’s direction. The ITiCSE working group of Pears et al (2005) explicitly addressed the foundation of a core literature of computing education research. They proposed that a core literature should comprise influential papers, which are widely recognised as having made a significant contribution to the field; seminal papers, which help to define new problems or otherwise catalyse research in a new area; and synthesis papers, which summarise and analyse work in particular areas of CER. Their concept of seminal papers appears to be in reasonable accord with that of Fensham.

In work that is now ten years old, Pears et al (2005) were able to identify eleven seminal papers from the preceding six years. It would be presumptuous of me to augment their list without discussion with them and others, but there are undoubtedly more papers since then that would be widely recognised as seminal.

8.4 Outcome criterion

Implications for practice

After six structural criteria and seven intra-research criteria, Fensham nominates a single outcome criterion: the research must have implications for practice, and practice must be seen to be changing as a consequence of the research.

Computing education research has many implications for practice; they are discussed in the final sections of many CER papers, and their general form is 'now that we have discovered this, how does it change what we should do in the classroom?' It is by no means guaranteed that the literature will provide evidence of the adoption of such findings: when educators adopt a new approach based upon the findings of research, they will not necessarily write papers to say what they have done. Nevertheless, there is ample evidence that the implications are actually being acted upon. At the detailed end of the scale, many papers explicitly acknowledge the influence of research. For example, "Building on research that identifies and addresses issues of women's underrepresentation in computing, this article describes promising practices in undergraduate research experiences that promote women's longterm interest in computer science and engineering" (Kim et al 2011).

At the broader end, there is evidence of widespread adoption of particular approaches either derived from research findings or leading to research findings. As explained by Guzdial (2013), *media computation* began not as a research project but as an idea that might help to address a number of problems such as academic misconduct and poor retention. From there, it developed into a teaching environment and system, textbooks, and a number of empirical research papers by instructors who had adopted the approach.

Being computer scientists, computing educators often build tools to address perceived needs, and much tool-based research has led to widespread adoption. For example, the web-based program visualisation tool Jeliot (Markkanen et al 1997), has led to numerous publications (eg Ma et al 2009, Pears & Rogalli 2011, Moreno et al 2014) indicating widespread adoption and extension.

Overall, there is clear evidence not just of implications for practice, but of those implications being acted upon.

8.5 Computing education research is a research discipline

Fensham listed 14 criteria by which it could be recognised that science education had emerged as a field of research. In this chapter it has been established that computing education satisfies 13 of those criteria, lacking only formal professional associations of computing education researchers. It is conceivable that no such association will ever be formed: that the informal associations encouraged by today's technology will be considered sufficient to nurture and maintain the necessary discipline of research.

Recalling that Fensham's criteria were gathered after the emergence of science education research, and are therefore descriptive rather than prescriptive, it does not seem to be stretching credibility too far to assert that, in meeting the other 13 criteria, computing education research has earned the right to be considered a research discipline of the sort that Fensham discussed. It is a young discipline, and some of the criteria have been met only quite recently; but nevertheless it has arrived.

It must be acknowledged again that Fensham's criteria are not the sole and agreed means of establishing what constitutes a research discipline. They are a convenient and well-reasoned checklist that nobody has actively disputed. But whatever other means might be used to establish the same thing, they are likely to have a great deal in common with Fensham's, and the credentials established in this thesis for computing education research are likely to lead to the same conclusion: computing education has emerged as a discipline of research.

9. Conclusion

This thesis summarises some eight years of work that has taken a number of unexpected turns, and has been intermingled with research on other projects. It has been illuminating to extract this work and describe its journey.

The journey began with what seemed a simple idea: to find a way of distinguishing practice papers from research papers in the computing education literature and measuring their relative proportions over time and across venues. Once it had become clear that no existing classification system would suffice, a new classification system had to be devised. Then more researchers had to be involved, to establish that the system actually made sense and was not just being applied arbitrarily by a lone researcher.

The involvement of other researchers opened the work to their questions, which went well beyond the initial question. First, for those papers that are classified as research, how have the researchers gone about that research, and what tools and approaches have they used? Second, moving to the alternative notion that all published papers are in some sense research, how have the researchers gone about their research? And third, how does all of this somewhat tedious classification help to form a picture of computing education research and its role in the body of computing education literature?

In the course of this work, Simon's system has been applied to papers from six computing education conferences and two computing education journals. These are by no means the only venues for the publication of papers in computing education. Joy et al (2009) survey 42 venues in which computing education papers have been published, with a view to helping authors select the most appropriate venue for their work. The venues include many journals and conferences on educational technology, which has some overlap with computing education, but is not the same; but they also include, for example, the ACM Conference on Information Technology Education (SIGITE), and the IEEE Frontiers in Education conference (FIE), both of which have substantial overlap with the conferences and journals examined in this thesis. The work presented here could not have been applied in a reasonable time to all of the venues where computing education papers are published; rather, it has analysed a substantial subset of the possible venues. The trends and patterns that have been found are fairly consistent across that subset, suggesting that they would also be found across the wider set of venues if they were investigated.

9.1 The research questions

Six research questions were posed in section 1.3 of this thesis. RQ1 was the overriding question and the next four were subsidiary questions posed to help answer the first. Therefore we shall consider those four before returning to RQ1.

> ***RQ2: How can it be determined whether a publication constitutes research?***

> ***RQ3: What other aspects of computing education publications would lead to an informative broad picture of complete corpuses of them?***

These questions were addressed through the development and application of Simon's system for classifying computing education publications. The system was created because no existing system appeared able to answer the questions.

There is no absolute definition that can be applied to every paper to determine whether it is a research paper; therefore a convenient approximation was required. The nature dimension of Simon's system offers such an approximation, although it comes closer to identifying empirical research than all research.

An answer to RQ2 is thus to apply the nature dimension of Simon's system, and consider as research all publications that are classified as analysis, study, or experiment. On a broad scale this can be considered a sufficient condition for describing a paper as research, although it has been acknowledged that it is by no means a necessary condition. Even accepting that publications can be meaningfully partitioned into 'research' and 'other', there are clearly position/proposal papers that should be classified as research. In the classification of the journals Computer Science Education and TOCE, reported in section 7.6, about half of the nine position/proposal papers discussed theories or methods of computing education research, and would thus appear to qualify as research papers.

The remaining three dimensions of Simon's system together comprise an answer to RQ3. These dimensions, together with nature, can be used to provide a broad picture of computing education (and other) publications, and the acceptance for publication of the papers presented in this thesis supports the argument that the picture has been found informative by reviewers and editors.

At the time this work was begun, a number of authors, such as Dale (2002) and Valentine (2004), suggested that there was a growth in the proportion of computing education papers that could be considered as research. RQ4 was devised to test this suggestion.

> ***RQ4: Do publishing venues for computing education papers show an increase in the proportion of their papers that can be described as research?***

Analysis of large numbers of the papers from six computing education conference and two computing education journals has answered this question with a resounding yes. The only venues that did not show clear upward trends were ICER and Computer Science Education, which already had high proportions of research papers.

To explicitly address two of Fensham's structural criteria, RQ5 was formulated as follows:

> ***RQ5: Does computing education have one or more research conferences and one or more research journals?***

While Fensham gave no measure by which conferences and journals could be classified as research conferences and research journals, the proportions of research papers at ICER, Computer Science Education, TOCE, and to a lesser extent at some other computing education conferences, suggest that this question must be answered in the affirmative.

The principal research question was

> ***RQ1: Has CER emerged as a research discipline?***

Following the development of Simon's system and its application to large numbers of computing education papers from many venues, and following the answers established for research questions RQ2 – RQ5, this question was addressed using the criteria proposed by Fensham (2004) for a research discipline. While the professional associations criterion was not formally met, it was proposed that the presence of the other 13 leads to the conclusion that computing education research can justify calling itself a discipline of research. Like other discipline-based education research ventures, it exists to address the specific problems of educating students in its domain, and thus to establish new approaches to teaching and learning that will enhance the achievements of its students, and, indeed, the satisfaction of its students and of its teachers.

The final research question arose during the conduct of the research:

> ***RQ6: What theories, methods, and approaches are used in computing education research papers, and in what disciplines do they originate?***

This question was answered, along with related questions, by the work reported in chapter 7 and in publications P7, P8, and P10.

9.2 Significance of this work

The work presented in this thesis reports on the emergence of computing education as a research discipline. What does this mean for the members of the discipline? What does it mean to be part of a research discipline?

Fensham (2004) observes that establishing the identity of science education as a research discipline permits science education to stand comparison with the other sciences as a research discipline. This applies equally to computing education research. The confirmed existence of the discipline lends weight to computing education researchers who are told by other computing researchers that what they do is not research. While this thesis has entertained the proposition that all computing education papers constitute research, we must be careful when pressing claims in this regard. It is likely that computing researchers who denigrate computing education research do so principally on the basis of the many practice papers that are published in the field. This thesis has confirmed the growth of rigorous research in computing education; if computing education researchers are to use this finding to press their claim for being researchers, they should take care not to conflate their research with the many teaching practice papers that are still being published.

Practice papers are not research in the sense in which it is used in the bulk of this thesis, and they are not research in the eyes of our colleagues who consider computing education an easy way to produce publications on the basis just of one's teaching experience. However, it must be emphasised again that these papers fill a valuable role in the computing education community. While calling for an expansion of computing education research, Dale (2002), Valentine (2004), Fincher and Petre (2004), and others acknowledge the value of the swap-meet that takes place at some conferences, and Goldweber (Goldweber et al 2004) explicitly argues that research papers must not take over completely from practice papers.

The balance between research and other papers will continue to be determined by the computing education community itself. Calls for journal and conference papers make it reasonably clear what types of submission they seek, and these guidelines are generally embodied in the instructions to reviewers and in the acceptance decisions made by the chairs or editors. If the community feels at some point that the balance is not right, editors and chairs will be able to adjust the balance by more explicit guidelines for authors regarding the types of submission that are most likely to be accepted at their venues, and by instructions to reviewers regarding how to implement those guidelines. This is what happened when ICER was proposed and created, and when JERIC became TOCE and clearly changed its emphasis.

Fensham (2004) notes further that establishing the identity of science education as a research discipline enhances its authority to argue with other scientists "about the many aspects of what is appropriate in science education". This also clearly applies to computing education. Many decisions about how to teach are made on the basis of "folk pedagogy: traditional pedagogy as . . . embodied in local custom or lore" (Lister 2008). It is the responsibility of computing education researchers to either validate or debunk folk pedagogic practices, and to propose and validate new practices to replace those that it debunks: to bring as much rigour to their explorations of computing education as other computing researchers apply to their non-educational fields of research.

Once established as a research discipline, computing education must of course continue to develop. Tedre (2015) observes that "characterizing a field through its fundamental questions [gives] an intuitive glimpse of what the field is about", while leaving open "many paradigmatic elements, such as choice of proper methods, standards for choosing between competing solutions, rules for interpreting results, and conventions for resolving disputes." The analysis presented in this thesis shows the dominance of programming courses as the contexts for computing education research, and suggests that two closely related fundamental questions of computing education research are:

> ***How do people learn to program?***
>
> ***Why do many people find it so difficult to learn to program?***

These questions certainly merit the attention that they are given by computing education researchers. At the same time, there are innumerable other aspects of computing education that deserve to be explored. A classification system presented by Kinnunen et al (2010) highlights areas that appear to receive little or no research attention, and that might therefore be ripe for investigation. As a research discipline, computing education should consider what its fundamental questions might be, and consequently what issues require research.

The depiction of computing education research presented in this thesis has the potential to form the basis for related work by other researchers, work that will take the concepts used here and enhance and adapt them as required. The definition of research used in the early chapters of this thesis is not intended to stand through time as authoritative; rather, it was devised for this particular investigation and has been successfully applied to address the questions being asked. If it finds subsequent use, it will be because it is again considered both convenient and suitable to address further questions. Research is not, and should never be, a static concept: it must grow and adapt to its surroundings, and this thesis forms part of a recognised stage in, and contribution to, the advancement of empirical computing education research.

There are at least two ways in which this work can contribute to the continuing change in the nature of publications in computing education. It can be applied in some aspects of research training, to help describe what kinds of paper researchers might aim to write, or what kinds of paper can legitimately be included in a doctoral degree based upon publications. It can equally well be used in seeking support for further academic recognition, demonstrating to computing researchers that computing education publications are not all experience reports, and that computing educators can and do carry out a great deal of empirical research.

In addition, if conference chairs, journal editors, and their reviewers pay heed to the message presented here, they can more consciously decide what types of paper that their conferences and journals will seek to publish.

Beyond computing education, this work can stand as a guide to other discipline-based education researchers. Education research in other disciplines has a number of things in common with computing education research. In most disciplines where they are found, education publications include practice papers. This is good for the disciplines, and should continue. However, just as with computing education, other disciplines must benefit from rigorous research addressing their particular problems and issues. Furthermore, it seems likely that education publications in those disciplines suffer from the same dismissive perceptions as those in computing: that they are not really research, and are just an easy way for teachers to get publications. Work of the type applied in this thesis can be applied in other disciplines, and the findings can be used for the same purpose: to make it clear that discipline-based education publications can indeed be rigorous research, and, when it matters, to establish guidelines for distinguishing the research from the experience reporting.

The work presented here also illustrates the application of Fensham's criteria to establish the emergence of a discipline-based education research discipline. As such, it can be used as a guide for researchers in other disciplines wishing to follow the same path – and, in particular, to assure them that the criteria are necessarily descriptive rather than prescriptive, and can therefore be applied with a modicum of flexibility.

9.3 Recommendations to computing educators

The story told by this thesis is of a gradual progression from the exchange of educational experiences in swap meets to empirical research in computing education. It is interesting to be informed of innovations and developments in the classroom, but it is far more useful to be presented with evidence of their effectiveness. One message to educators, then, is to pay attention to the research. If an approach is confirmed to be valuable, and is appropriate to your circumstances, use it. That's what the research is for.

Another message is to innovate – but to do so in a research setting. If you have promising ideas, implement them, but at the same time conduct research to evaluate them, and publish that research – even if the evaluation fails to establish that the innovation was effective.

A third message is to be prepared to look outside computing education for relevant research and for ideas that might lead to innovation. While an increasing number of ideas for computing education research are coming from computing education research, there is an abundance of educational research ideas in education itself, in other discipline-based education research fields, in psychology, in the social sciences; computing education research is multidisciplinary, and a failure to recognise this could lead researchers to overlook many excellent directions for future research.

9.4 Classifying this thesis

How would this thesis itself be described according to the classification schemes applied in the work?

According to Simon's system it would be classified as follows:

- nature: analysis (and thus research)
- context: literature
- scope: not applicable

- theme: research

And according to TMMCER:

- theory/model/framework: Simon's system; this would be counted as a TMF not in the paper where it was first described, but in subsequent publications
- technology/tool: none
- reference discipline: computing education for Simon's system
- research purpose: evaluative – positivist and descriptive – other
- research framework: for some of the work, the Delphi technique, survey
- data source: literature
- analysis method: interpretive classification, descriptive statistics

9.5 Future work

Simon's scheme for classifying (computing) education papers has served its initial purpose, but it does not end its life in the pages of this thesis. It has been independently applied by other researchers in computing education (Rowan & Dehlinger 2013, Falkner et al 2014) and there is a suggestion that it might be used in the construction of a more comprehensive analysis in engineering education (Fayyaz & Jesiek 2012). Time permitting, it will also be applied to subsequent corpuses of publications in the conferences and journals already classified, to see if further changes can be detected since the classifications reported here.

Work also continues on the TMMCER project, with findings still emerging from our analysis of the current data.

References

Azevedo, J (1997). *Mapping reality*. State University of New York Press, New York.

Banerjee, M, M Capozzoli, L McSweeney, D Sinha (1999). Beyond kappa: a review of interrater agreement measures. *Canadian Journal of Statistics* 27:1, 3-23.

Behrens, H, P Luksh (2006). A bibliometric study in crystallography. *Acta Crystallographica* B62, 993-1001.

Berglund, A, I Box, A Eckerdal, R Lister, A Pears (2008). Learning educational research methods through collaborative research: the PhICER initiative. *Tenth Australasian Computing Education Conference (ACE 2008)*, 35-42.

Biggs, JB (2003). *Teaching for quality learning at university: what the student does*. Open University Press, Maidenhead.

Bikner-Ahsbahs, A, S Prediger (2006). Diversity of theories in mathematics education – how can we deal with it? *ZDM* 38:1, 52-57.

Blaxter, L, C Hughes, M Tight (2001). *How to research*. Open University Press, Maidenhead.

Bloom, BS, BB Mesia, DR Krathwohl, (1956). *Taxonomy of Educational Objectives*. David McKay Co Inc, New York.

Borrego, M (2007). Development of engineering education as a rigorous discipline: a study of the publication patterns of four coalitions. *Journal of Engineering Education* 96:1, 5-18.

Bouma, GD, R Ling (2004). *The research process*. Oxford University Press, South Melbourne.

Brew, A (2001). *The nature of research*. Routledge, Milton Park.

Chalmers, AF (1982). *What is this thing called science?* University of Queensland Press, St Lucia.

Chung, KH, HS Pak, RAK Cox (1992). Patterns of research output in the accounting literature: a study of the bibliometric distributions. *Abacus* 28:2, 168-185.

Collis, B, Moonen, J (2001). *Flexible learning in a digital world: experiences and expectations*. Kogan Page, London.

Creswell, JW (2003). *Research design*. Sage Publications, Thousand Oaks.

CSC (2013). *Computer Science Curricula 2013*. ACM, New York, USA; IEEE Computer Society, Los Alamitos, USA.

Dale, N (2002). Increasing interest in CS Ed research. *SIGCSE Bulletin* 34:2, 16-17.

Denny, P, A Luxton-Reilly, J Hamer (2008). The PeerWise system of student contributed assessment questions. *Tenth Australasian Computing Education Conference (ACE 2008)*, 69-74.

Falkner, K, R Vivian, N Falkner (2014). The Australian digital technologies curriculum: challenge and opportunity. *16th Australasian Computing Education Conference (ACE 2014)*, 3-12.

Fawcett, J (1999). *Theory and research*. F.A. Davis Company, Philadelphia.

Fayyaz, F, BK Jesiek (2012). Work in progress: mesoscopic analysis of engineering education scholarship in electrical and computer engineering, 2002-2011. *IEEE Conference on Frontiers in Education*.

Fensham, PF (2004). *Defining an identity – the evolution of science education as a field of research*. Kluwer Academic Publishers, Dordrecht.

Fincher, S, R Lister, T Clear, A Robins, J Tenenberg, M Petre (2005). Multi-institutional, multi-national studies in CSEd Research: some design considerations and trade-offs. *First International Computing Education Research Workshop (ICER 2005)*, 111-121.

Fincher, S, M Petre (2004). *Computer science education research*. Routledge Falmer, London.

Fincher, S, J Tenenberg (2006). Using theory to inform capacity-building: bootstrapping communities of practice in computer science education research. *Journal of Engineering Education* 95:4, 265-278.

Fincher, S, J Tenenberg (2011). *A commons leader's vade mecum*. University of Kent Press, Canterbury.

Fincher, S, J Tenenberg, A Robins (2011). Research design: necessary bricolage. *Seventh International Computing Education Research Workshop (ICER 2011)*, 27-32.

Fleiss, JL (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76:5, 378-382.

Glass, RL (1992). A comparative analysis of the topic areas of computer science, software engineering, and information systems. *Journal of Systems Software* 19, 277-289.

Glass, RL, V Ramesh, I Vessey (2004). An analysis of research in computing disciplines. *Communications of the ACM* 47:6, 89-94.

Glass, RL, I Vessey, V Ramesh (2002). Research in software engineering: an analysis of the literature. *Information and Software Technology* 44, 491-506.

Goldweber, M, M Clark, S Fincher, A Pears (2004). The relationship between CS education research and the SIGCSE community. *SIGCSE Bulletin* 36:3, 228-229.

Guzdial, M (2013). Exploring hypotheses about media computation. *Ninth International Computing Education Research Workshop (ICER 2013)*, 19-26.

Hamer, J, Q Cutts, J Jackova, A Luxton-Reilly, R McCartney, H Purchase, C Riedesel, M Saeli, K Sanders, J Sheard (2008). Contributing student pedagogy. *SIGCSE Bulletin* 40:4, 194-212.

Hazzan, O (2003). How students attempt to reduce abstraction in the learning of computer science. *Computer Science Education* 13:2, 95-122.

HERDC (2015). 2015 Higher Education Research Data Collection: Specifications for the collection of 2014 data; Australian Government. http://docs.education.gov.au/system/files/doc/other/final2015herdcspecifications_1.pdf

Heron, PRL, DE Meltzer (2005). The future of physics education research: intellectual challenges and practical concerns. *American Journal of Physics* 73, 390-394.

Holmboe, C, L McIver, C George (2001). Research agenda for computer science education. *13th Workshop of the Psychology of Programming Interest Group (PPIG 13)*, 207-223.

Jadud, M (2006). Methods and tools for exploring novice compilation behavior. *Second International Computing Education Research Workshop (ICER 2006)*, 73-84.

Joy, M, J Sinclair, S Sun, J Sitthiworachart, J López-González (2009). Categorising computer science education research. *Education and Information Technologies* 14, 105-126.

Kervin, L, W Vialle, J Herrington, T Okely (2006). *Research for educators*. Thompson, South Melbourne.

Kim, KA, AJ Fann, KO Misa-Escalante (2011). Engaging women in computer science and engineering: promising practices for promoting gender equity in undergraduate research experiences. *ACM Transactions on Computing Education* 11:2, article 8.

Kinnunen, P, V Meisalo, L Malmi (2010). Have we missed something? Identifying missing types of research in computing education. *Sixth International Computing Education Research Workshop (ICER 2010)*, 13-21.

Lister, R (2006). Call me Ishmael: Charles Dickens meets Moby Book. *SIGCSE Bulletin* 38:2, 11-13.

Lister, R (2007). The Randolph thesis: CSEd research at the crossroads. *SIGCSE Bulletin* 39:4, 16-18.

Lister, R (2008). After the gold rush: toward sustainable scholarship in computing. *Tenth Australasian Computing Education Conference (ACE 2008)* 3-17.

Lister, R, ES Adams, S Fitzgerald, W Fone, J Hamer, M Lindholm, R McCartney, JE Moström, K Sanders, O Seppälä, B Simon, L Thomas (2004). A multi-national study of reading and tracing skills in novice programmers. *SIGCSE Bulletin* 36:4, 119-150.

Lister, R, A Berglund, I Box, C Cope, A Pears, C Avram, M Bower, A Carbone, B Davey, M de Raadt, B Doyle, S Fitzgerald, L Mannila, C Kutay, M Peltomäki, J Sheard, Simon, K Sutton, D Traynor, J Tutty, A Venables (2007). Differing ways that computing academics understand teaching. *Ninth Australasian Computing Education Conference (ACE 2007)*, 97-106.

Lister, R, J Edwards (2010). *Teaching novice computer programmers: bringing the scholarly approach to Australia*. Official report to the Australian Learning and Teaching Council. http://www.olt.gov.au/resource-teaching-novice-computer-programmers

Lister, R, B Simon, E Thompson, JL Whalley, C Prasad (2006). Not seeing the forest for the trees: novice programmers and the SOLO taxonomy. *11th Conference on Innovation and Technology in Computer Science Education (ITiCSE 2006)*, 118-122.

Lopez, M, J Whalley, P Robbins, R Lister (2008). Relationships between reading, tracing and writing skills in introductory programming. *Fourth International Computing Education Research Workshop (ICER 2008)*, 101-111.

Ma, L, J Ferguson, M Roper, I Ross, M Wood (2009). Improving the mental models held by novice programmers using cognitive conflict and Jeliot visualisations. *14th Conference on Innovation and Technology in Computer Science Education (ITiCSE 2009)*, 166-170.

Malmi, L (2004). Foreword to *Proceedings of the Fourth Finnish/Baltic Sea Conference on Computer Science Education*, iii.

Markkanen, J, E Sutinen, T Teräsvirta, J Tarhio, E Rautama (1997). Animation of algorithms with Eliot and Jeliot (demonstration). *Supplemental proceedings of the Conference on Integrating Technology into Computer Science Education (ITiCSE-WGR'97)*, 152.

McCracken, M, Y Ben-David Kolikant, V Almstrum, C Laxer, D Diaz, L Thomas, M Guzdial, I Utting, D Hagan, T Wilusz (2001). A multi-national, multi-institutional study of assessment of programming skills of first-year CS students. *SIGCSE Bulletin* 33:4, 125-140.

McDermott, L, E Redish (1999). Resource letter: PER-1: Physics education research. *American Journal of Physics* 67:9, 755-767.

Meyer, JH, R Land (2005). Threshold concepts and troublesome knowledge (2): epistemological considerations and a conceptual framework for teaching and learning. *Higher Education* 49, 373-388.

Moreno, A, E Sutinen, M Joy (2014). Defining and evaluating conflictive animations for programming education: the case of Jeliot ConAn. *45th ACM Technical Symposium on Computer Science Education (SIGCSE 2014)*, 629-634.

Morrison, B, L Ni, M Guzdial (2012). Adapting the disciplinary commons model for high school teachers: improving recruitment, creating community. *Eighth International Computing Education Research Workshop (ICER 2012)*, 47-54.

Naps, T, G Rößling, V Almstrum, W Dann, R Fleischer, C Hundhausen, A Korhonen, L Malmi, M McNally, S Rodger, A Velázquez-Iturbide (2002). Exploring the role of visualization and engagement in computer science education. *SIGCSE Bulletin* 35:2, 131-152.

Nicholls, PT (1989). Bibliometric modeling processes and the empirical validity of Lotka's law. *Journal of the American Society for Information Science* 40:6, 379-385.

Niss, M (2007). The concept and role of theory in mathematics education. *Relating practice and research in mathematics education. Fourth Nordic Conference on Research in Mathematics Education (NORMA 05)*, 97-110.

Okasha, S (2002). *Philosophy of science*. Oxford University Press, Oxford.

Osorio, NL, MA Osorio (2002). Engineering education in Europe and the USA. *Science & Technology Libraries* 23:1, 49-70.

Paas, F, A Renkl, J Sweller (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist* 38, 1-4.

Pascarella, ET, PT Terenzini (1991). *How college affects students: findings and insights from twenty years of research*. Jossey-Bass, San Francisco.

Patra, SK, S Mishra (2006). A bibliometric study of bioinformatics literature. *Scientometrics* 67:3, 477-489.

Pears, A, M Rogalli (2011). mJeliot: a tool for enhanced interactivity in programming instruction. *11th Koli Calling International Conference on Computing Education Research (Koli Calling 2011)*, 16-22.

Pears, A, S Seidman, C Eney, P Kinnunen, L Malmi (2005). Constructing a core literature for computing education research. *SIGCSE Bulletin* 37:4, 152-161.

Pears, A, S Seidman, L Malmi, L Mannila, E Adams, J Bennedsen, M Devlin, J Paterson (2007). A survey of literature on the teaching of introductory programming. *SIGCSE Bulletin* 39:4, 204-223.

Powell, C (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing* 41:4, 376-382.

RAE (2005). Research assessment exercise guidance to panels http://www.rae.ac.uk/pubs/2005/01/rae0105.doc

Ramesh, V, R Glass, I Vessey (2003). Research in computer science: an empirical study. *Sprouts: Working Papers on Information Systems* 3:7, 165-176.

Randolph, J (2007a). *Computer science at the crossroads: a methodological review of the computer science education research: 2000-2005*. (Doctoral dissertation), Utah State University. http://archive.org/details/randolph_dissertation

Randolph, JJ (2007b). What's the difference, still? A follow up methodological review of the distance education research. *Informatics in Education* 6:1, 179-188.

Randolph, J, R Bednarik, N Myller (2005). A methodological review of the articles published in the proceedings of Koli Calling 2001-2004. *Fifth Finnish / Baltic Sea Conference on Computer Science Education (Koli Calling 2005)*, 103-109.

Randolph, JJ, G Julnes, R Bednarik, E Sutinen (2007). A comparison of the methodological quality of articles in computer science education journals and conference proceedings. *Computer Science Education* 17:4, 263-274.

Rowan, M, J Dehlinger (2013). A systematic literature review on using mobile computing as a learning intervention. *18th Conference on Innovation and Technology in Computer Science Education (ITiCSE 2013)*, 339.

Shavelson, RJ, L Towne (eds)(2002). *Scientific research in education*. National Academy Press, Washington DC.

Simon, S Fincher, A Robins, B Baker, I Box, Q Cutts, M de Raadt, P Haden, J Hamer, M Hamilton, R Lister, M Petre, K Sutton, D Tolhurst, J Tutty (2006). Predictors of success in a first programming course. *Eighth Australasian Computing Education Conference (ACE 2006)*, 189-196.

Singer, SR, NR Nielsen, HA Schweingruber (eds)(2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. National Academy Press, Washington DC.

Slavin, RE (2007). *Educational research in an age of accountability*. Pearson, Boston.

Sriraman, B, L English (2005). Theories of mathematics education: a global survey of theoretical frameworks/trends in mathematics education research. *ZDM* 37:6, 450-456.

Teague, D, R Lister (2014a). Longitudinal think aloud study of a novice programmer. *16th Australasian Computing Education Conference (ACE 2014)*, 41-50.

Teague, D, R Lister (2014b). Manifestations of preoperational reasoning on similar programming tasks. *16th Australasian Computing Education Conference (ACE 2014)*, 65-74.

Teague, D, R Lister (2014c). Programming: reading, writing and reversing. *19th Conference on Innovation and Technology in Computer Science Education (ITiCSE 2014)*, 285-290.

Tedre, M (2015). *The science of computing: shaping a discipline*. CRC Press, Boca Raton.

Tenenberg, J, S Fincher (2007). Opening the door of the computer science classroom: the disciplinary commons. *38th ACM Technical Symposium on Computer Science Education (SIGCSE 2007)*, 514-518.

Tenenberg, J, R McCartney (2009). Introducing the ACM Transactions on Computing Education (editorial). *ACM Transactions on Computing Education* 9:1, article 1.

Tenenberg, J, R McCartney (2011a). Computational tools for computing education (editorial). *ACM Transactions on Computing Education* 11:1, article 1.

Tenenberg, J, R McCartney (2011b). Entry points for computing education research (editorial). *ACM Transactions on Computing Education* 11:4, article 1.

Tsai, C-C, ML Wen (2005). Research and trends in science education from 1998 to 2002: a content analysis of publication in selected journals. *International Journal of Science Education* 27:1, 3-14.

Valentine, D (2004). CS educational research: a meta-analysis of SIGCSE Technical Symposium proceedings. *SIGCSE Bulletin* 36:1, 255-259.

Vessey, I, V Ramesh, RL Glass (2002). Research in information systems: an empirical study of diversity in the discipline and its journals. *Journal of Management Information Systems* 19:2, 129-174.

Vessey, I, V Ramesh, RL Glass (2004). A unified classification system for research in the computing disciplines. *Information and Software Technology* 47, 245-255.

Wankat, PC (1999). An analysis of the articles in the Journal of Engineering Education. *Journal of Engineering Education* 88:1, 37-42.

Wankat, P (2004). Analysis of the first ten years of the Journal of Engineering Education. *Journal of Engineering Education* 93:1, 13-21.

Wegner, P (1976). Research paradigms in computer science. *Second International Conference on Software Engineering*, 322-330.

Williams, B, P Neto (2012). Tracking engineering education research and development – contributions from bibliometric analysis. *International Journal of Engineering Pedagogy* 2:2, 37-44.

Yetim, M, M Turoff (2004). Structuring communication processes and enhancing public discourse: the Delphi method revisited. *Ninth International Working Conference on the Language-Action Perspective on Communication Modeling (LAP 2004)*, 235-251.

Appendix: Simon's system of classification for computing education papers

Nature – a fixed dimension

experiment	A paper describing a scientific-style experiment, with control group and experimental group, and independent and dependent variables.
study	A paper describing a study that has gathered data for the explicit purpose of the research, and subsequently analysed the data.
analysis	A paper describing analysis of existing data, such as student answers to examination questions, or papers found in the literature.
report	A report on something that has been done, such as the introduction of a novel technique in the classroom; often called an experience report. Even if there is some analysis, such as of student feedback, a paper will be classified as a report it if appears that its purpose is to describe what was done.
position/proposal	A paper expressing an opinion or proposing future work.

Scope – a fixed dimension

subject	The work has taken place – the data, if any, have been collected – within a single subject/course/unit/paper at a single institution.
program/department	The work has taken place within two or more subjects in the same department or degree program.
institution	The work has taken place across two or more departments or degree programs at a single institution.
many institutions	The work has taken place across two or more institutions. Note that the focus is on where the work has taken place: where the data, if any, have been collected. If data are collected at a single institution, the contribution of an author from another institution does not make it a multi-institutional paper; that would require data collection (or equivalent) from more than one institution.
not applicable	Some work cannot be tied to any of the scopes mentioned above. An example would be a literature review: while it might have been conducted by a team of authors from around the world, it did not necessitate their involvement, and could have been conducted by a single author.

Context – a dynamic dimension, to be augmented as required; the values listed here indicate the contexts that have been identified in the papers classified to date. Most of the contexts should be self-explanatory, as they will be the names of subjects/courses/units/papers found in computing degrees. Other contexts are explained in the table.

artificial intelligence	
broad-based	This is the context given to most publications reporting on work that is not set in any particular classroom subject. Examples would include a report on the development of a tool that is for use across many different subjects, or a paper that discusses the curriculum for a complete degree.
capstone project	
communication skills	
compilers	
cryptography	
data mining	
database	
data structures	
eBusiness/eCommerce	
ethics/professionalism	
formal methods	
graphics	
groupwork	This context is used for papers that might report on work in other contexts, but that focus on the groupwork aspect, not on the subject matter taught.
hardware/architecture	
human-computer interaction	
information systems	
introduction to IT	
literature	This context is used for all literature reviews and surveys, even if they focus on a particular subject area such as programming.
logic	
mathematics	
multimedia	
networks	
operating systems	
postgraduate/research	This context refers not to work being done by postgraduate students as part of their research, but to courses taught to postgraduate research students, such as courses in research ethics or academic writing.
programming languages	This context is for courses in programming languages, which will, for example, study the syntax and semantics of different programming paradigms.
programming	

robotics	
school/outreach	This context captures the many papers discussing school outreach activities or other work conducted in school education. Such papers could have been listed under broad-based, but there are enough of them to warrant a count of their own.
security	
software engineering	
study planning	This is another context that could be encompassed in broad-based, but that has enough papers to seem to warrant its own category.
systems analysis	
webpage development	
work experience	

Theme – a dynamic dimension, to be augmented as required; this table lists the themes that have been identified in the papers classified to date; the list has proved robust over the course of the study, requiring little augmentation. Brief explanations are provided for the themes that are not necessarily self-evident.

ability/aptitude/understanding	Students' ability and/or aptitude for computing, and understandings of students or of teachers
accessibility	Accommodations made for students with particular accessibility requirements, such as visual impairment
assessment techniques	Techniques used in assessment, including the use of specific software tools that have already been introduced
assessment tools	Presentation of new or enhanced software tools for use in assessment
attitudes	Students' (and potentially teachers') attitudes to the computing field or specific courses
cheating & plagiarism	
competitions	
credit for prior learning	Assessment of exemptions from certain study requirements based upon prior study
curriculum	
educational technology	Tools, typically but not exclusively hardware, as used in computing education
ethics/professional issues	
gender issues	
language/culture issues	
online/distance delivery	
recruitment, progression, pathways	
research	This theme is not for research papers – the nature dimension makes that identification; it is for papers that are *about* research
teachers	

teaching/learning techniques	Techniques used in teaching, including the use of specific software tools that have already been introduced
teaching/learning theories & models	A clear focus not on how we teach, not on the tools we use to teach, but on the relationship between teaching/learning and the underlying theories and models of those processes
teaching/learning tools	Presentation of new or enhanced software tools for use in teaching
tutors, demonstrators, mentors	

Aalto-DD 150/2015

The past 10-15 years have seen a clear change in the nature of publications in computing education. From a field dominated by experience reports, descriptions of classroom practice, there has been a steady increase in the proportion of publications reporting on rigorous educational research. This thesis charts the progress of the change through a detailed analysis of the papers published in a number of computing education conferences and journals. It also examines the research approaches and methods used in these papers. Based on this analysis and a number of other features, it concludes that computing education is now established as a discipline of research, alongside the other research disciplines that are found within the broad academic area of computing. Finally, it considers the significance and ramifications of this finding.

ISBN 978-952-60-6415-4 (printed)
ISBN 978-952-60-6416-1 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University

Department of Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS