Koli Calling Comes of Age: an Analysis

Simon
School of Design, Communication, and Information Technology
The University of Newcastle
Newcastle, Australia
simon@newcastle.edu.au

Abstract

A detailed analysis of the full papers presented at the first six years of Koli Calling shows how the conference has matured over that time to become more research-oriented. While the first three years were dominated by papers proposing or reporting on classroom initiatives, the next three have seen a remarkable increase in the proportion of papers describing experimental and analytical research. This paper quantifies and explores that increase, and considers in addition the range of topics, contexts, and scopes of these six years of papers.

Keywords: Computing education research, literature review, classification.

1 Introduction

Koli Calling was launched in 2001 "to develop the exchange of relevant information between colleagues working within the same discipline" (Sutinen & Kuittinen 2002). In the manner typical of conferences, there was a list of seven indicative topics, and that list included Computer Science Education Research.

By the fourth year there was much talk of "CS Education and CS Education Research", and it was decided to categorise each paper as either *discussion* ("papers that present novel ideas, approaches, and systems for CS Education") or *research* ("papers in which these issues have been elaborated further in some rigid research setting") (Malmi 2004).

The fifth year elaborated on these definitions and added system and poster categories:

- research: submissions presenting a novel approach, method, tool, finding, interpretation, explanation, or other contribution in a solid scientific framework;
- system: papers describing methods or tools for learning or instruction in CS or a related field, presented in a constructive framework with references to related work;
- discussion papers and posters: typically describing original work in progress (Salakoski 2005).

Salakoski also noted that "over the years, [Koli Calling] has developed into its present form of a rigorous scientific meeting . . . Last year can be considered as the breakthrough of the conference" (Salakoski 2005).

At the 2006 conference the category definitions took on a slightly different hue:

- research: presenting novel results, methods, tools or interpretations that contribute to solid, theoretically anchored research;
- system: tools for learning or instruction in computing education, motivated by the didactic needs of teaching computing;
- discussion: presentation of novel ideas and prototypes;
- poster: work in progress (Berglund 2007).

In addition, in 2006 the conference name changed to reflect the increased research emphasis. Formerly known as the Finnish / Baltic Sea Conference on Computer Science Education, it now became the Baltic Sea Conference on Computing Education Research..

While Salakoski was quite certain that 2004 marked a watershed in the research emphasis of Koli Calling, the variation in the number and definition of categories of papers leads to three clear questions:

1. What is a research paper?
2. What proportion of papers presented at Koli Calling can be called research papers?
3. Has the proportion of research papers increased over the six years of Koli Calling?

I have recently devised a classification for computing education papers (Simon 2007), and now report on the application of this classification to the past six years of Koli Calling, to address these three questions and possibly to discover other features of interest.

The next section summarises some prior classification systems that have been applied to computer education research or to the publications that report on such research. The paper then proceeds to explain my own system, report the results of applying it to Koli Calling, compare those results with other categorisations of Koli Calling papers, and draw some conclusions.

2 Prior Classifications

The literature includes a number of systems for classifying papers in computing education. In Simon (2007), an analysis of recent papers at two Australasian conferences, these systems are discussed at some length, and attention is devoted to explaining why they were not suitable for the purpose of that analysis. Here they will be presented rather more briefly.

Perhaps the best known and most widely cited is Valentine (2004), who surveyed 20 years of SIGCSE Technical Symposium papers dealing with first-year Computer Science subjects, putting each paper into one of six categories:

- experimental: papers including any sort of scientific analysis;
- Marco Polo: descriptions of the application of a new curriculum, language, or course;
- philosophy: attempts to generate debate on philosophical grounds;
- tools: software tools developed to assist with aspects of teaching/learning or assessment;
- nifty: innovative, interesting ways to teach abstract concepts;
- John Henry: papers describing outrageously difficult ways of undertaking simple tasks.

While this categorisation distinguishes between research papers (the experimental category) and others (the remaining five categories), it does have some weaknesses, not least of which is the lack of clear guidelines as to how subsequent researchers might apply the scheme.

Fincher and Petre (2004) proposed ten subfields for computing education research, and Pears, Seidman, Eney, Kinnunen, and Malmi (2005) combined these into the four broader fields of

- studies in teaching, learning, and assessment,
- institutions and educational settings,
- problems and solutions, and
- computing education research as a discipline.

However, neither of these classifications appears to distinguish between research papers and others.

Randolph, Bednarik, and Myller (2005) examined the full papers in the first four years of Koli Calling, performing a thorough analysis of the methodology of the 17 papers that involved research with human participants. The study pays scant attention to the remaining 42 papers, but does categorise them as

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

It might be tempting to conclude that the 17 papers represent research while the 42 do not, but this is clearly inappropriate, as research papers can certainly be found among, for example, meta-analyses and methodological papers.

Until now, therefore, it might appear that Valentine's 'taxonomy' is the only system that attempts to distinguish between research and non-research papers, and the basis of that distinction is not entirely clear.

Furthermore, all of these systems tend to linearise, to place upon a single axis properties that are in fact orthogonal. For example, Valentine's experimental category addresses the nature of the paper, while his tools category addresses the paper's subject matter. This is my main reason for choosing to devise a new classification, with four orthogonal dimensions, which I here apply to all 102 papers published in the proceedings of Koli Calling between 2001 and 2006.

3 A New Classification for Computing Education Papers

The new classification system categorises papers according to four distinct dimensions, which are orthogonal in the sense that each is independent of the others, and that a paper's categorisation in one dimension imposes no constraints or limitations on its categorisations in the others. The nature of a paper describes what sort of paper it is; its topic describes what it is about; its context describes the subject matter of the course in which it is based; and its scope is a measure of the breadth of the work in the computing education community.

The system was devised with the purpose of forming an overall picture of the papers presented in recent years at the two major Australasian computing education conferences. It categorised so as to summarise, so as in turn to facilitate an overview of the research. The development of the system is described in more detail in Simon (2007).

3.1 Nature

Many computing education researchers distinguish between practice and research papers. This new system of classification further divides research papers into two separate categories, *experiment* and *analysis*, introduces a *report* category that is reasonably congruent with practice papers as they are generally understood, and notes and explains an additional category, *position* papers.

According to prior definitions already canvassed, research can suggest "some rigid research setting" (Malmi 2004), "a solid scientific framework" (Salakoski 2005), "contribut[ing] to solid, theoretically anchored research" (Berglund 2007), or "including any sort of scientific analysis" (Valentine 2004). While these definitions suit the purposes for which they were intended, I have chosen to be rather more specific.

I define an *experiment* paper as one that reports on a clear and deliberate research experiment. The authors have set out to answer a particular question, devised a study (which might be as simple as a survey) to assist in that regard, carried out the study, gathered the data, and analysed it. An example paper in this category is *A multi-national, multi-institutional study of student-generated software designs* (Fincher, Petre, et al 2004).

An *analysis* paper is one whose authors have set out to answer a particular question, gathered existing data as appropriate, and analysed it. Unlike an experiment paper, it does not involve devising and conducting a study; instead it uses data already available, such as existing class results, the literature, or other sources. A typical analysis paper is *Progress Reports and Novices' Understanding of Program Code* (Mannila 2006).

A *report* is a paper describing something that has been tried or developed in an educational context. Many papers of this sort exhibit the minimal analysis of conducting a student survey to confirm the appearance of success; so long as it is clear that the principal intent of the paper is to report on the trial or the development, I still classify such papers as reports. A paper that falls clearly into this category is *Producing interactive web lectures with authorware* (Kerola 2003).

Finally, a *position* paper is one that elucidates the authors' thoughts on a matter, or perhaps sets out plans for future work, without having anything concrete to report upon. An example of a position paper is *Fibonacci Numbers Using Mutual Recursion* (Rubio & Pajak 2005).

By these definitions, I believe that both experiment and analysis papers would normally be regarded as research, that reports would normally be considered practice papers, and that position papers are in a field of their own.

It has been suggested that papers describing innovations in computing education constitute research, and that it is therefore unreasonable to exclude all reports from the body of work identified as research papers. This position can certainly be argued, but so can the position that any paper reporting on anything of interest to the computing education community constitutes research. The delineation in this system is inescapably subjective, but it does have the advantage of drawing a line between research and practice papers, and indeed position papers; and such a line is necessary if one is to have any chance of answering questions such as whether the proportion of research papers at Koli Calling has increased over the years.

Even so, more investigation is required into exactly what does constitute research, and into the validity of the research versus practice division. In the meantime, for the purposes of this paper it is probably safe to suggest that experiment and analysis papers are definitely research, and to leave open the question of whether some or all reports and position papers also qualify for that description.

I have been asked why the system distinguishes between experiment and analysis papers when both are combined to form what I am calling research papers. The answer to this is that the system was developed not to identify research papers but to analyse a corpus of papers and see what emerged. When the analysis was being performed, these two categories seemed quite distinct (and easily distinguished), and were both therefore incorporated into the system. It happens that together they comprise a more widely recognised category, but this does not diminish the distinction between them.

3.2 Topic

The topic dimension describes what a paper is actually about, and its membership emerges from a content analysis of the corpus being studied. Even so, the list of categories appears to be reasonably stable: the list that emerged for Koli Calling, shown in Table 1, is all but identical to the list that emerged from the prior study of two Australasian conferences (Simon 2007).

Most of the topics should be reasonably self-explanatory, but a few of them might require a word of explanation.

With regard to both *assessment* and *teaching/learning*, the *tools* topic is used for a paper reporting on the development of a new tool, or perhaps a novel use of an existing tool, while papers that report on a reasonably expected use of an existing tool will come under the *techniques* category. *Do students SQLify? Improving learning outcomes with peer review and enhanced computer assisted assessment of querying skills* (de Raadt et al 2006) reports on the development of a new tool for use in assessment, and so is categorised under *assessment tools*; whereas *Automatic grading of graphical user interface programs exploiting Jemmy* (Surakka et al 2005) reports on the use of existing tools to perform assessment tasks, and so is categorised under *assessment techniques.*

While *teaching/learning techniques* concerns ways of teaching and learning (for example, *Learning programming by programming: a case study* (Hassinen & Mäyrä 2006)), *teaching/learning* concerns the act of teaching and/or the act of learning (for example, *Survival of students with different learning preferences* (Bednarik & Fränti 2004)).

The *research* topic does not indicate whether a paper is a research paper – the nature dimension does that. In the topic dimension, *research* denotes papers that are essentially *about* research – as, for example, Randolph et

Table 1: topics covered in six years of Koli Calling

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

al (2005).

3.3 Context

Some readers might be surprised not to see topics such as first-year programming, capstone projects, group work, and so on. After careful consideration I have decided that these are very seldom the topic of a paper; rather, they are the *context* in which the work was done and the paper written. This therefore forms a new dimension, which I call context. A paper's context will most often be a subject area of some sort, such as programming, computer systems, theory of computation, etc; but not all computing education papers are set in the context of particular subjects, so further categories, such as literature, have been added to better reflect the observations.

As with topic, the values in this list will vary according to the corpus of work being analysed. Table 2 lists the contexts found while analysing the six years of Koli Calling papers. Again, most of these values should be fairly self-explanatory. A broad-based context describes a paper that is set in no particular subject area, either because it covers multiple subjects (*Contextual computing studies in Tanzania* (Sutinen et al 2002)) or because it is very general and mentions no particular subjects (*Evaluation of faculty workload for various methods in computer science education* (Kurhila 2002)). Many capstone projects entail group work; a paper dealing with such a project will be categorised according to whether the emphasis is on the project subject itself (*Moral conflicts perceived by instructors of a project course* (Vartiainen 2005)) or on the groups undertaking the project (*A pilot study concerning power in CS student project groups* (Wiggberg 2006)).

3.4 Scope

The final dimension helps describe the breadth of the work on which the paper is based. The narrowest focus is a single *subject*; some papers report on a range of subjects within a *department/program*; others might focus on the whole *institution*, and others again on *many institutions* (where 'many' can be as few as two).

Not all papers have an identifiable scope. *Explanograms: low overhead multi-media learning resources* (Pears & Olsson 2004) explains a tool that can be used to help with teaching, but the explanation is not based on a particular subject, on many institutions, or on something between. For this reason, the scope dimension includes a *not applicable* category.

The principal value of this dimension is that it can be seen as one possible measure of collaboration within the computing education community: while work in a single subject can be the work of an individual or a small teaching team, work across multiple institutions necessarily reflects significant community involvement. This is why the dimension is retained in the system despite the fact that it cannot be usefully applied to all papers.

Table 2: contexts found in six years of Koli Calling

broad-based	literature
capstone project	logic
compilers	mathematics
data structures	programming
database	software engineering
group work	study planning
hardware/architecture	theory of computation
information systems	work experience
introduction to IT	writing

4 Classifying Koli Calling

Having established the classification system described in section 3, it seemed reasonable to apply it to other computing education conferences, both to facilitate an overview of the papers at those conferences and (eventually) to permit some sort of comparison of the major computing education conferences worldwide.

This classification system has been applied by a single researcher to all 102 full papers published in the proceedings of Koli Calling 2001-2006, with consideration where appropriate to trends over that time. 'Full papers' is here taken to mean all papers published in the proceedings other than those labelled as keynote, invited, demo, or poster.

As mentioned before, the system was not devised with any particular agenda other than to categorise, summarise, and view; but in the case of Koli Calling, it did seem to offer a means of assessing the suggestion that the conference had become more research oriented, so the analysis was carried out with that suggestion in mind.

4.1 Nature

Taking the definition of nature in section 3.1 as an answer to what constitutes a research paper, what proportion of Koli Calling papers are research papers, that is, papers whose nature is either experiment or analysis?

Table 3 shows the number and proportion of papers that fall into each of the categories of nature. Combining the experiment and analysis categories, we see that about 35% of all full papers are classified as research papers. This is comparable with the recent analysis of the major computing education conferences in Australian and New Zealand (Simon 2007), in which 22% of the papers were categorised as experiment and 13% as analysis, for the same 35% overall proportion of research papers.

The third research question of this paper concerns a possible increase of the proportion of research papers

Table 3: natures of all full papers

	count	proportion
experiment	14	14%
analysis	21	20%
report	48	47%
position	19	19%

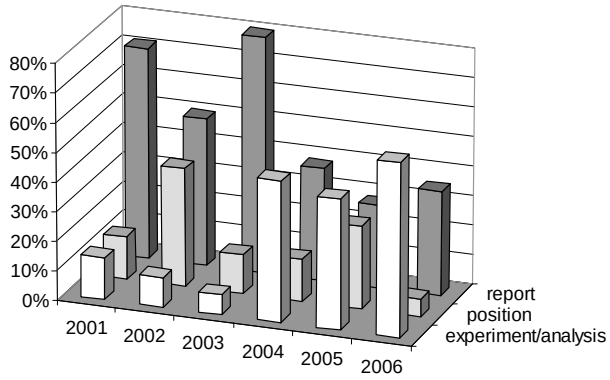

Figure 1: Proportion of papers by nature and year

over time. Figure 1 represents the data graphically, combining experiment and analysis papers into a single research grouping, which is plotted alongside position and report papers for each year of the conference.

The results here are little short of astonishing. For the first three years of the conference, research papers made up respectively 14%, 10%, and 7% of the proceedings; the next year they surged to 47%, a level that was all but maintained with 44% in 2005 and surpassed with 59% in 2006.

Salakoski (2005) did not have this classification system to hand when he described 2004 as a year of breakthrough, but this analysis clearly supports his evaluation of the 2004 and 2005 conferences, and indicates that the pattern has continued at least to 2006.

4.2 Topic

Figure 2 shows the number of papers in each topic over the six years of Koli Calling. Given the 'swap meet' nature of computing education conferences, it is not surprising to see a large number of papers on teaching/learning techniques (28%), teaching/learning tools (20%), assessment techniques (8%), and assessment tools (8%). A solid 8% of papers on ability/aptitude is more or less guaranteed by the perennial problem of programming aptitude, while the strong representation of papers on distance/online delivery (11%) might reflect a particular facet of Finnish education, as it is almost twice the 6% found in the earlier study of Australian and New Zealand papers (Simon 2007).

4.3 Context

Figure 3 illustrates the number of papers in each of the observed contexts. The predominance of work set in the context of programming subjects (35%) is to be expected, perhaps because so many computing education subjects entail programming, and almost certainly because of the constant attempts we make to address the particular difficulties faced by students in learning to program. Programming aside, the only standout context is broad-based, at 23%, and that is because this category does not represent an actual context but encompasses papers that cover more than one context and more abstract papers that are devoid of context.

4.4 Scope

Over the six years of the conference, some 53% of the papers have reported on work concerning a single subject; 11% on work spread across a program or department; 9% on work concerning multiple institutions; and 27% on work with no identifiable scope.

As mentioned in section 3.4, I contend that work conducted across multiple institutions tends to indicate a stronger involvement with the computing education community than work dealing with single subjects. In the prior study using this classification system (Simon 2007) I observed some correlation between nature and scope: essentially, multi-institutional papers are more likely to be research papers than reports or position papers. Having noted the 2004 surge in research papers at Koli Calling, I was curious as to whether the scopes of those research papers bore out this apparent correlation. Did the increase in experiment and analysis papers correspond with a broadening of scope?

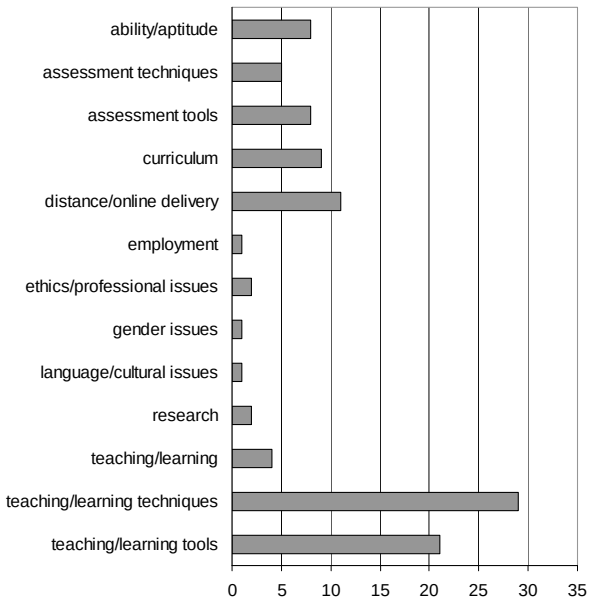

Figure 2: number of papers in each topic

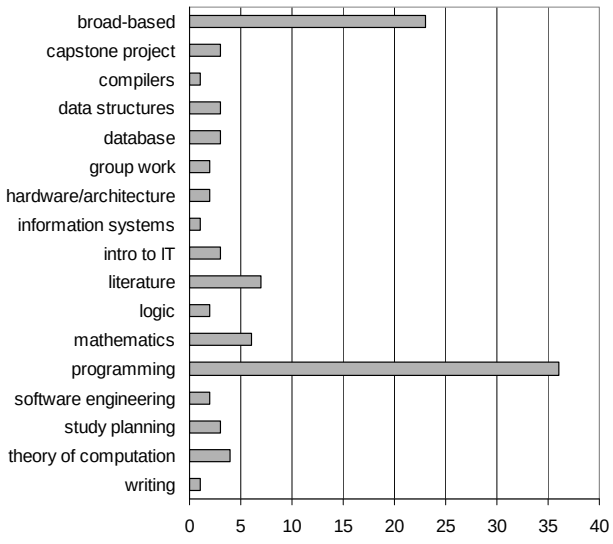

Figure 3: number of papers in each context

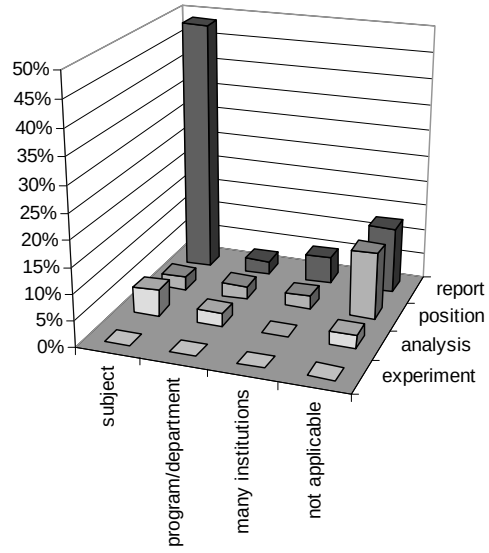

Figure 4: scope and nature of papers, 2001-2003

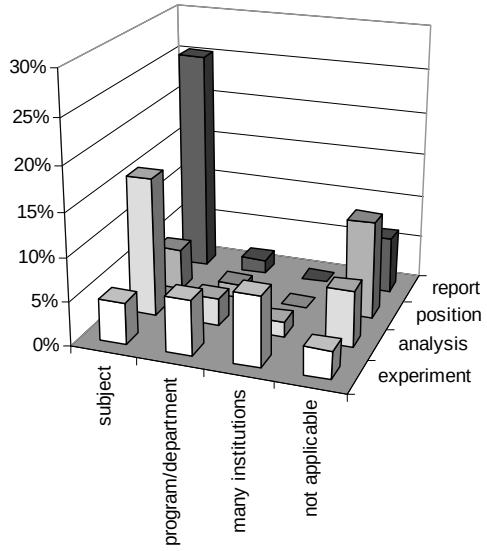

Figure 5: scope and nature of papers, 2004-2006

Figure 4 shows a two-dimensional analysis of papers by nature and scope for the first three years of the conference, and figure 5 shows the same analysis for the next three years. While both figures are clearly dominated by reports based on single subjects, this dominance dropped from nearly 50% in the first three years to about 25% in the next three.

Having already observed the surge in experiment and analysis papers in the three years from 2004, we now see from figures 4 and 5 that many of the new research papers are also based in single subjects (5% of the papers in these three years are experiment papers in the subject scope and 16% are analysis papers in the subject scope), but the 9% that are research papers based across the program/department (6% experiment and 3% analysis) and the 10% that are research papers based across multiple institutions (8% experiment and 2% analysis) show that the surge in research papers does indeed correspond with an increase in papers of broader scope.

5 Comparison with Prior Work

How do these findings tally with those of Randolph et al (2005) in their study of what was essentially a single major aspect of the papers from 2001 to 2004? Unfortunately, this is not as straightforward a question as it seems, because the differences between the systems are too great.

Randolph found that 17 papers involved research with human participants while 42 did not. Over the same period I find 14 experiment and analysis papers and 46 report and position papers. While I cannot explain the minor difference in the total, I believe that these findings are reasonably congruent.

As the main thrust of Randolph's categorisation concerns methodology, his system counts methodological cases rather than papers, and identifies 74 cases in the 59 papers. It seems, though, that most of the multi-case papers are in the human research area, as the remaining 42 papers appear to give rise to only 44 cases. Therefore it might seem reasonable to compare my counts of non-research papers with Randolph's counts of non-human-research cases – if I could work out what to compare with what.

Randolph's 'literature reviews, meta-analyses' might well fall into my analysis category; 'program descriptions without anecdotal evidence', 'program descriptions with anecdotal evidence', and 'technical investigations' might generally correspond to my reports; and 'theoretical, methodological or philosophical papers' might correspond to my position papers. But these putative correspondences are rather too vague to be of use, and after a brief look at the numbers I have concluded that no meaningful comparison is possible.

As an aside, Randolph noted the region of first author's affiliation for each paper, finding that about 90% of all papers came from Finland. Extending this analysis to the following two years and examining it year by year, I note a steady internationalisation of the conference (figure 6); the number of papers whose first authors are from Finland has dropped from 100% in 2001 to 53% in 2006.

It is interesting that there are consistently fewer papers from Baltic Sea countries other than Finland than there are from the rest of Europe. I leave the conference organisers to speculate on the reason for this apparent

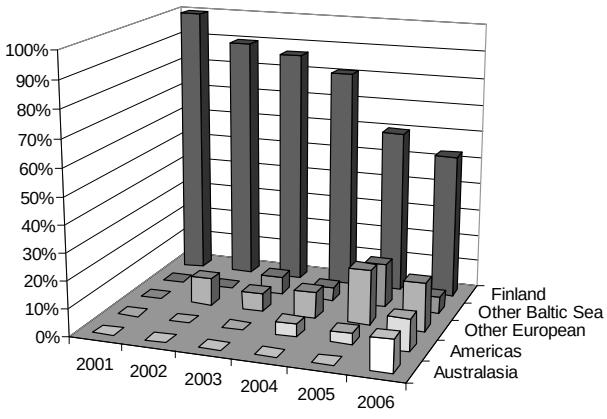

Figure 6: internationalisation of Koli Calling, as shown by affiliation of first author

lack of interest from Finland's close neighbours.

In addition to Randolph's study, it might be worth looking for some congruence between this paper's nature classification and the research/system/discussion distinction introduced for the more recent Koli Calling conferences. One might expect that most of the papers classified in the proceedings as research papers would fall into my experiment and analysis categories, and that most of the papers classified in the proceedings as technical investigations would fall into my report category.

Figure 7 presents the cross-tabulation of the nature classification with what I shall call the proceedings category. We must bear in mind that these categories were introduced only in 2004 and were added to in 2005, so there are many Koli Calling papers that have no proceedings category at all.

Nevertheless we see that of the papers examined, most of those whose proceedings category is research are indeed experiment (28%) or analysis (45%) papers, and that most of those whose proceedings category is system are indeed reports (88%). It is also no surprise that position papers make up a solid 35% of the papers categorised in the proceedings as discussion papers.

What might be more puzzling is that the papers categorised in the proceedings as discussion papers include 23% that I categorise as experiment papers and 15% that I categorise as analysis papers. This is explained by the fact that discussion papers are solicited somewhat differently from research or system papers; they are clearly intended to be reports on work in progress, and to be shorter than research and system papers. Therefore this difference is not a shortcoming of Koli Calling's discussion paper category or of my classification system; rather, it is a recognition that work in progress can fall into any of the four nature categories.

6 Conclusions

This classification system for computing education papers incorporates a reasonably firm definition for the notion of a research paper, which can fall into one of two categories. An experiment paper is one that reports on a research question, a study to address that question, and an analysis of the results of the study, while an analysis paper is one that reports on a research question and an analysis of existing data to address the question.

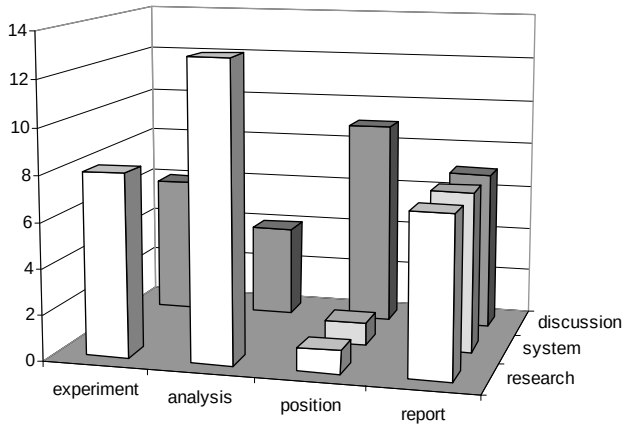

Figure 7: nature vs proceedings category

It is pleasing to report that some 35% of the full papers presented at Koli Calling fall into one or other of the research categories, and that the proportion displayed a remarkable increase in 2004 and has sustained that increase since then.

This increase in the proportion of research papers could be due either to an increased proportion of research submissions or to an increased inclination to accept research submissions over others. The two published Koli Calling acceptance rates (Malmi 2004, Salakoski 2005) are quite high, so it seems unlikely that the change is a result of the paper selection process; one must conclude that the conference really has seen a dramatic increase in the proportion of research papers being submitted.

6.1 Research Papers versus Practice Papers

How can the computing education community benefit from this work? While many academics appear to be aware of a distinction between practice and research papers, a clear explanation of the distinction might help them to target their work to the nature that they prefer. I believe that the concrete definitions presented here will be of help in this regard.

In addition, I believe that a clear understanding of this classification system will assist conference chairs to better distinguish the natures of submitted papers, which might help if the chairs wish to lead the conference in a particular direction, such as towards a higher proportion of research papers.

While this paper might lean slightly towards the position that research papers are in some sense better than practice papers, that is not its intention. The primary intention was to present and apply a tool that can be used to summarise, and therefore to better view, a large corpus of papers in computing education research; and a secondary intention was to see if the resulting view provided any support for earlier assertions that Koli Calling had developed into a more research-oriented conference.

6.2 Future Directions

To date there has been no study of inter-rater reliability in the use of this classification system. Such a study is expected to take place at a workshop planned for early 2008. In addition, it is planned to analyse other major computing education conferences over a comparable timeline, with a view to comparing and contrasting those conferences.

It is clear that there is also scope for further investigation into the apparent division of papers into research and practice; into the validity of labelling experiment and analysis papers as research and labelling reports and position papers as 'not research'; and into such value-laden areas as a desirable balance between research and practice papers.

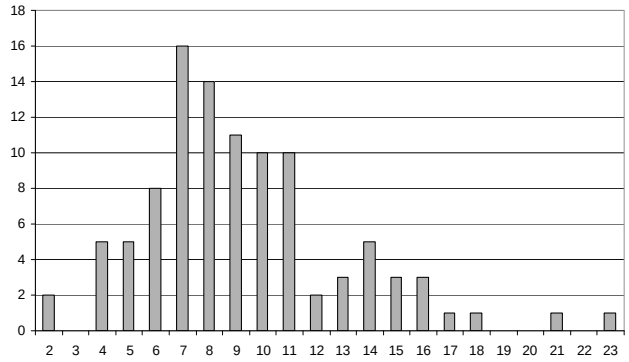

Figure 8: number of words in the titles of papers

6.3 Postscript

As a minor distraction from the analysis of the papers I also recorded the number of words in the title of each. The results are presented in figure 8. Are submissions with seven-word titles more likely than other submissions to be accepted to Koli Calling? Readers are invited to draw their own conclusions.

Acknowledgements

This paper has improved as result of feedback from the Koli Calling referees, discussion with many people at the conference, and subsequent discussion with Leslie Schwartzman. I am grateful to all of these people.

References

Bednarik, R. & Fränti, P. (2004). Survival of students with different learning preferences. *Proc. 4th Finnish / Baltic Sea Conference on Computer Science Education*, 121-125.

Berglund, A. (2007). Foreword to *Proc. 6th Baltic Sea Conference on Computing Education Research*, iii.

de Raadt. M., Dekeyser, S., & Lee, T.Y. (2006). Do students SQLify? Improving learning outcomes with peer review and enhanced computer assisted assessment of querying skills. *Proc. 6th Baltic Sea Conference on Computing Education Research*, 101-108.

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

Fincher, S., Petre, M., Tenenberg, J., Blaha, K., Bouvier, D., Chen, T-Y., Chinn, D., Cooper, S., Eckerdal, A., Johnson, H., McCartney, R., Monge, A., Moström, J.E., Powers, K., Ratcliffe, M., Robins, A., Sanders, D., Schwartzman, L., Simon, B., Stoker, C., Elliott Tew, A., & VanDeGrift, T. (2004). A multi-national, multi-institutional study of student-generated software designs. *Proc. 4th Finnish / Baltic Sea Conference on Computer Science Education*, 20-27.

Hassinen, M. & Mäyrä, H. (2006). Learning programming by programming: a case study. *Proc. 6th Baltic Sea Conference on Computing Education Research*, 117-119.

Kerola, T. (2003). Producing interactive web lectures with authorware. *Proc. 3rd Finnish / Baltic Sea Conference on Computer Science Education*, 18-21.

Malmi, L. (2004). Foreword to *Proc. 4th Finnish / Baltic Sea Conference on Computer Science Education*, iii.

Mannila, L. (2006). Progress Reports and Novices' Understanding of Program Code. *Proc. 6th Baltic Sea Conference on Computing Education Research*, 27-31.

Pears, A., & Olsson, H. (2004). Explanograms: low overhead multi-media learning resources. *Proc. 4th Finnish / Baltic Sea Conference on Computer Science Education*, 67-74.

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

Randolph, J., Bednarik, R., & Myller, N. (2005). A methodological review of the articles published in the proceedings of Koli Calling 2001-2004. *Proc. 5th Finnish / Baltic Sea Conference on Computer Science Education*, 103-109.

Rubio, M., & Pajak B. (2005). Fibonacci Numbers Using Mutual Recursion. *Proc. 5th Finnish / Baltic Sea Conference on Computer Science Education*, 174-177.

Salakoski, T. (2005). Foreword to *Proc. 5th Finnish / Baltic Sea Conference on Computer Science Education*, iii.

Simon (2007). A classification of recent Australasian computing education publications. *Computer Science Education* **17**(3) 155-169.

Surakka, S., Auvinen, J., & Ihantola, P. (2005). Automatic grading of graphical user interface programs exploiting Jemmy. *Proc. 5th Finnish / Baltic Sea Conference on Computer Science Education*, 49-56.

Sutinen, E., & Kuittinen, M. (2002). Foreword to *Proc. 1st Annual Finnish / Baltic Sea Conference on Computer Science Education*, ii.

Sutinen, E., Vesisenaho, M., & Virnes, M. (2002). Contextual computing studies in Tanzania. *Proc. 1st Annual Finnish / Baltic Sea Conference on Computer Science Education*, 84-88.

Valentine, D. (2004). CS Educational Research: A Meta-Analysis of SIGCSE Technical Symposium Proceedings. *Proc. 35th SIGCSE Technical Symposium on Computer Science Education, ACM SIGCSE Bulletin*, **36**(1) 255-259.

Vartiainen, T. (2005). Moral conflicts perceived by instructors of a project course. *Proc. 5th Finnish / Baltic Sea Conference on Computer Science Education*, 25-32.

Wiggberg, M. (2006). A pilot study concerning power in CS student project groups. *Proc. 6th Baltic Sea Conference on Computing Education Research*, 132-135.