Empirical research in CSCW - a review of the ACM/CSCW conferences from 1998 to 2004

Jacques Wainer[1,2] **& Claudia Barsottini**[2]

[1]Institute of Computing
State University of Campinas (UNICAMP)
P.O.Box 6176
13083-970 Campinas - SP - BRAZIL
wainer@ic.unicamp.br

[2]Department of Health Informatics
Federal University of Sao Paulo (UNIFESP)
Rua Botucatu, 862
04023-062 Sao Paulo - SP - BRAZIL
{jacques.wainer,claudia.barsottini}@unifesp.br

Abstract

This paper reviews all the 169 full papers published in the ACM/CSCW conferences from 1998 to 2004. We classify the papers according to the type of empirical research they report. The classes are evaluation of groupware, description of work situations, hypothesis testing, bibliographic research, and papers with no empirical research. We show that the field is in constant change, that the proportions of papers in these different categories have been changing in these 6 years, and that for the evaluation papers, the results by Pinelle and Gutwin do not carry for the 1998-2004 period.

Keywords: *groupware, CSCW, systematic review, empirical research, bibliographic review, evaluation*

1. EMPIRICAL RESEARCH IN COMPUTER SCIENCE

We define empirical research, in opposition to analytical research and design work, as a a scientific inquiry that requires observation/measurement of "things in the real world". In a sense, empirical research in computer science is what makes it a part of "the other sciences," including here both the natural and the human sciences.

Some, if not most, of the research in computer science is **not** empirical, as evaluated in [22]. Tichy and collaborators [22] evaluated 400 articles published in 1993, 50 of them randomly selected papers published by ACM in 1993 and the rest systematically selected from a few journals in systems and software engineering, and classified the research in five categories (quoting [22] definitions):

- formal theory: articles whose main contributions are formally tractable propositions, e.g., lemmata and theorems and their proofs.

- design and modelling: systems, techniques, or models, whose claimed properties cannot be proven formally. Examples include software tools, performance prediction models, and complex hardware and software systems of all kinds.

- empirical work[1]: articles that collect, analyze, and interpret observations about known designs, systems, or models, or about abstract theories or subjects (as this paper does). The emphasis is on evaluation, not on new designs or models.

---
[1][22] definition of empirical work is much more limited that the one we use in this paper.

- hypothesis testing: articles that define hypotheses and describe experiments to test them.

- others: articles that do not fit any of the four categories above, e.g., surveys.

Tichy and collaborators found that for the random sample 12% of the articles are in the theory class, 70% are design and model, 2% are empirical, 2% are hypothesis testing 14% are classified as others.

A similar result was obtained by Glass and collaborators [8, 19] who analyzed 628 computer science papers from 1995 to 1999 and classified them regarding different dimensions. Of interest to this paper is the "research approach" dimension, whose possible values are descriptive research, formulative research or evaluative research. The work [8] finds that 9.88% of the reseach is descriptive, 10.98% is evaluative and 79.15% is formulative.

If one considers that Tichy's design and modeling class corresponds at least in part, with Glass's formulative approach to research, then one can conclude that from 70% to 80% of the published results in computer science from 1993 to 1999 are mainly the proposal of a *new* system, model, algorithm, process, taxonomy, framework (SMAPTF), and so on.

We will call research whose main approach is the proposal of a new SMAPTF a *design* research. We consider that the proposal of a new SMAPTF is not an empirical research, but a design research can contain some empirical component, usually as an *evaluation* of the new SMAPTF. In fact Tichy [22] finds that 60% of the design papers contain some form of evaluation. We will adapt [11] in defining evaluation of systems as:

- the process of ascertaining the merit and worth of an information system

We contrast evaluation with verification (the process of verifying that the system satisfy or matches the specification) and validation (the process of verifying that the system fits the intended usage). In general terms, verification matches the system to its specification, validation the system with what it was believes it should solve, and evaluation matches the system with its real users, in the real environment, and with the "real problems" it solves and creates. In this sense, both verification and validation are not necessarily empirical inquiry since they do not "measure" the system against reality, but against abstract constructs (specification and perceived problem) that are (hopefully) derived from reality. Evaluation, on the other hand, in necessarily an empirical activity.

1.1. GROUPWARE AND CSCW

We will define *groupware research* as the proposal, implementation and evaluation of systems that allow/foster human-human collaboration. Groupware research is one of the forms of research within the broader area of Computer Supported Collaborative Work (CSCW).

In our view, groupware research, when started, was a design work oriented field - new systems were proposed, but few were put to use, to verify if they did accomplished what their authors intended. At a certain moment the groupware community started paying more attention to the evaluation of the groupware systems proposed, and less to the innovative and creative aspects of the design work in groupware.

But the CSCW area as a whole has other influences besides groupware research. Some of the most cited research in CSCW is not about groupware systems or even computers, but about social and group dynamics (for example Grudin's work [10]) or how people work together and coordinate among themselves using different ways (for example Suchman's [21]). These and other examples of research in CSCW are empirical, in the sense that they "look for things in the real world" but are not evaluations of groupware systems.

This work reviews empirical research published in the ACM CSCW conference, from 1998 to 2004. The goal is to understand the distribution of research emphasis in CSCW along these years, to forecast future development, if possible, and to map the techniques that have been used in the area regarding empirical research. We believe researchers will find it useful to know the new developments in empirical research in CSCW, in order to plan future research. Furthermore, in relation to evaluation or groupware, researchers will be able to compare their own research with others that used similar techniques published in the period from 1998 to 2004, and hopefully understand the strength and shortcomings of their own methodological choices.

This work started as an attempt to repeat the influential "Review of groupware evaluation" paper by Pinelle and Gutwin [16]. But reviewing some of the more recent work published in the ACM CSCW conference, we realized that groupware evaluation is only one of the forms of empirical research being done and published by the CSCW community. We then expanded our goal, to evaluate and classify all published empirical research published in that conference.

Section 2 discusses some of the work in groupware evaluations, with emphasis on Pinelle and Gutwin work. Section 3 presents our model of empirical research. Section 4 discusses our methods for selecting and evaluating the papers. Section 5 present the results, and section 6 discusses the results.

2. GROUPWARE EVALUATION

Groupware researchers have always believed in improved productivity and user satisfaction through collaboration. But it has been clear that evaluating such claims is difficult. As it has been pointed out by [3, 4, 14] among others, groupware evaluation has many sides since both the group dynamics and the environment in which the group is inserted have impact on the effect of using the groupware. The measure of the impact is also unclear - does one look for increase in productivity or quality in a specific task with the inclusion of the groupware technology [5], or does one look for increase in user satisfaction in performing the task, or does one look for changes in the group [18], or organizational culture brought forth by the use of the groupware? Finally, groupware evaluation does not necessarily imply in some measure of impact from the use of the system. Some groupware evaluations [17, 20, 2] among others, are about the usability of the system, not its impact in the work, workers and organizations [7].

2.1. THE PINELLE AND GUTWIN MODEL

In [16] Pinelle and Gutwin propose a groupware evaluation model based on system and research evaluation proposals made by other researchers and philosophers, mainly [13].

For space reasons we will summarize the Pinelle and Gutwin's model to only those dimensions that are relevant to this paper. The model classify evaluation according to the following dimensions (figure 1 in [16]):

evaluation type the possible values for this dimension are: *field experiment* performed in a non artificial (or naturalistic) environment but with manipulation of some dependent variable; *field study and case study* performed in a naturalistic environment but with minimal manipulation of the dependent variables; *laboratory experiment* performed in a controlled environment with manipulation of the dependent variable; *exploratory* performed in a controlled environment but with no manipulation; or *no evaluation*

qualitative vs quantitative if the evaluation is: *qualitative*, or *quantitative*, or *both*.

evaluation technique . The possible values for this dimension are: *observation, observation with videotape, interview, questionnaire, quantitative work measures, qualitative work measures, collected archival materials,* and *discussion*.

focus of evaluation . The possible values are: *patterns of system use, support for specific task, user interaction through the system, specific interface features,* *user satisfaction, end product*, and *efficiency in task performance*.

The model also includes dimensions that classify the groupware tool itself, and which define the place of the evaluation in the lifecycle of the system, among others.

Our first goal for this paper was to use the Pinelle and Gutwin model to evaluate the CSCW research for the period from 1998 to 2004. But in trying to repeat the measurements, we realized that there were no clear definitions for what the values of some of the dimensions really mean, in particular, the dimensions of evaluation technique and focus of evaluation.

2.2. PINELLE & GUTWIN'S RESULTS

[16] work analysed all papers published in the period from 1990 to 1998 in the ACM CSCW conference in which a new groupware system was proposed and possibly evaluated. The analysis resulted in 45 papers that were then classified using the Pinelle and Gutwin model.

The results of the analysis of the 45 papers, for the dimensions of evaluation type, qualitative/quantitative, evaluation technique, and evaluation focus dimensions, are:

- for evaluation type: 13 (28%) lab experiment, 12 (28%) field and case study, 1 (2 %) field experiment, and 9 (20%) of the research has no evaluation.

- for qualitative/quantitative dimension: 23 (72%) qualitative, 7 (22%) both, and 2 (6%) quantitative.

- evaluation technique: 24 (83%) observation, 10 (34%) observation with videotape, 12 (41%) interview, 9 (31 %) questionnaire, 9 (31%) quantitative work measures, and so on.

- evaluation focus: 16 (50%) patterns of system use, 15 (47%) support for specific task, 14 (44%) user interaction, 12 (38%) specific interface issues, 12 (38%) user satisfaction, 8 (25%) organizational impact, 5 (16%) end product, and 4 (13%) efficiency of task performance.

3. EMPIRICAL RESEARCH EVALUATION MODEL

We developed the following evaluation model for CSCW publications. The model started from the Pinelle/Gutwin, but it changed as we started evaluating the papers, once we realized that some papers were clearly a report of empirical research but were not groupware evaluation. As yet the model has no clear philosophical foundation although we feel that it derived from our

understanding of Tichy's work [22], specially its emphasis on *empirical* research.

The first dimension is **type of empirical research**, which may be:

design and evaluation of a groupware system. Design and evaluation (D&E) papers usually describe a groupware system in details, specially the functionalities that the authors believe are innovative and significant. Only a small part of the paper is devoted to the evaluation of the system

descriptive the description of a particular work environment/situation where collaboration (or communication in general) is a central aspect. The typical descriptive paper will describe (usually qualitatively) a work environment/situation where even without computer support, collaboration seems to be important. Another typical example of descriptive paper will describe the consequences of the introduction of a communication/collaboration system in organizations, or in peoples lives.

explanatory [2]. Explanatory papers describe experiments that support or falsify some hypothesis about collaboration. The typical explanatory paper will define a set of hypothesis regarding computer mediated collaboration, describe a set of experiments and the associated communication/collaboration tools used in the experiments, and discuss how the experiments confirm or disconfirm the hypothesis.

bibliographic . Bibliographic research systematically collects a set of scientific publications in CSCW, or related areas, and describes patterns of publication, choices taken by the researchers in the area, patterns of co-authorship, and so on.

not empirical . If the paper has no empirical component. Usually not empirical papers present a new groupware tool (but no evaluation of said tool), or a framework to build groupware tools, or a model for classifying groupware, or models for evaluating groupware.

If the publication is a design and evaluation, we further classify the research following a simplified Pinelle/Gutwin model. We chose the type of evaluation and the qualitative/quantitative as the most informative dimensions to classify groupware evaluation. In fact we considered also classifying the research in the methods and the focus of evaluation dimensions, but the lack of a

clear definition in [16] of what the values mean caused too much divergence between the two authors.

If the publication is a descriptive, we further classify it whether it is quantitative or qualitative.

3.1. OBSERVATIONS

Classifying a research paper in one of the research types above can be problematic for two reasons - a particular paper may be in the fortier of two classes or it may be in two classes simultaneously. The two borders that are somewhat fuzzy are the descriptive/D&E and the explanatory/D&E. Let us discuss these borders.

A descriptive research describes a work situation/environment where collaboration happens with or without groupware tools. If the tool is well known, for example Lotus Notes or Internet Messenger, than we consider the research as descriptive - it is unlikely that one would consider the description of how IM is used by teenagers [9] a (qualitative) evaluation of IM. But if the groupware tool is less well known, if it was developed by the researchers themselves or in-house, and so on, it is less clear that the research is a description of a work environment and not a evaluation of the tool. In these cases, if the paper has a long description of the groupware tool itself, we considered it an design and evaluation paper. Otherwise we considered it a descriptive paper. The confusion between classifying a research as descriptive or as a D&E research paper were the most common source of disagreement between the two authors.

A second source of confusion is a groupware research paper (a paper that proposes and maybe evaluates a new groupware tool) in which the description of the current work practices serve as the requirement specification for the groupware tool. Another variation of this situation is possible when the evaluation of a groupware reveals unexpected modes of collaboration/work situations which are then described in details. In the set of research we evaluated, [12] is an example of the first case and [6] an example of the second.

The third source of confusion is the explanatory/D&E border. An explanatory paper tries to prove or disprove a clearly stated hypothesis, usually with very simple groupware support. Again, if the groupware is not so simple, it is unclear whether the research is about the hypothesis or about the groupware itself. If the groupware is simple and little space in the paper is dedicated to describing it (and its novel features), we considered the paper as explanatory, otherwise we considered it a D&E research.

Finally, a single research may be in more than one class. A particular research [23] is both a descriptive and an explanatory research - the descriptive part lists a set of observed phenomena which are tested in an experiment.

A scenario we did not encounter, but find it possible and more likely in the future, is the research that is both

---

[2]The terms "descriptive" and "explanatory" were proposed in [15] for the categorization of *qualitative* research. We use the terms in a broader sense to refer to any empirical research.

descriptive and evaluative. Descriptive research usually ends with a "implications for design" section, in which a groupware system that would either mitigate the problems of non-collaboration or enhance the modes of collaboration is proposed. The next logical step is to implement the groupware and evaluate it. If these two steps are reported in a single publication, that research would be both descriptive and evaluative.

Finally, some decisions regarding the class not-empirical need to be made explicit. We decided to classify as not-empirical research that proposed a framework, and illustrated its use by describing a groupware tool using the framework. We considered that such publication is an *illustration* of how the framework can be put to use, and not an *evaluation*. That is a debatable decision - one could argue that that is indeed an evaluation, in particular, a qualitative, introspection-based evaluation of the framework.

4. METHODS

We selected, and read all the papers published in the 1998, 2000, 2002 and 2004 ACM CSCW conferences, but we removed from our sample the papers with 4 pages or less, which, we believe, corresponded to the short paper (or ongoing work) category.

All papers were independently evaluated by both of us. If there were discrepancies in the evaluation of any item, the paper was re-evaluated by the two researchers, and then discussed. If there was no convergence after the second evaluation and the following discussion, the evaluation of one of us (JW) was used as the final one.

5. RESULTS

Table 1 report the number of papers in each conference that was evaluated. Table 2 is the main result of this work. It classifies all papers in its type of empirical research. In order to reduce the size of the bibliography we use a incomplete reference - each paper is described as pXXX-YY, where XXX is the page where the paper starts and YY is the year of the conference. The interested reader can easily obtain the full reference for the papers in the ACM's site, in particular the Table of Contents for the CSCW page [1]. Table 3 lists the papers that are classified into more than one type.

Table 4 lists the evaluation type for the papers that were classified as design and evaluation. Table 5 list the qualitative/quantitative dimension both for the D&E and for the descriptive researches.

Tables 6 lists the total for each type of empirical research, for each year. Tables 7 and 8 list the totals for

Year	papers selected (more than 4 pages)
1998	40
2000	36
2002	39
2004	55
total	169

Table 1. Number of CSCW papers accepted and selected for evaluation

each subclass of the D&E and descriptive classes of empirical research, for each year.

6. DISCUSSION

The first important aspect, regarding the empirical research types, is the almost monotonic decline of the not-empirical research, the relative stability of the design and evaluation research (on around a third of the papers), the rapid grow and possible stability of the descriptive research, and the monotonic increase in explanatory research. Figure 1 plots the evolution of these research types. Figure 1 shows that the area is in constant change in the period.

If we take Pinelle and Gutwin [16] data regarding evaluation type and the qualitative/quantitative dimension for the design and evaluation research, then the differences between the period 1990-1998 and the period 1998-2004 periods are significant: chi-squared = 9.5745, df = 2, p-value = 0.008335 for the evaluation type and chi-squared = 8.0109, df = 2, p-value = 0.01822 for the qualitative/quantitative/both dimension. Thus, the proportions of the different evaluation techniques and whether researchers used a qualitative or quantitative approach are significantly different in the period 1990-1998 from the period 1998-2004. There was a very significant increase in field experiment as an evaluation technique, that is, the use of manipulations in a naturalistic environment. There was also a strong increase in quantitative evaluations, which changed from 6% of the evaluations in the 1990-1998 period to 31% in the 1998-2004 period.

Regarding descriptive research, the qualitative methodology, that is the use of ethnographic methods to understand work practices and work environments, is still dominant (table 8). Although there was a trend to increase the use of mixed methods, the 2004 conference interrupted the trend, and thus it is still unclear the future directions of the descriptive research.

Bibliographic research, like this one, has its place in any scientific domain, but of course it has to be a minority form of research - only when there is enough non-bibliographic research published it is worth performing this kind of research.

Finally, we are very excited with the steady growth

Research type	year	references
design and evaluation	1998	p1-98 p11-98 p109-98 p149-98 p159-98 p197-98 p217-98 p265-98 p275-98 p285-98 p345-98 p363-98
(total 51)	2000	p1-00 p11-00 p89-00 p97-00 p135-00 p155-00 p163-00 p241-00 p251-00 p261-00 p289-00 p299-00
	2002	p96-02 p106-02 p116-02 p136-02 p146-02 p156-02 p276-02 p296-02 p314-02 p324-02 p354-02
	2004	p1-04 p29-04 p39-04 p49-04 p82-04 p152-04 p172-04 p192-04 p232-04 p242-04 p274-04 p314-04 p467-04 p477-04 p533-04 p554-04
descriptive	1998	p29-98 p39-98 p49-98 p315-98p355-98 p393-98
(total 51)	2000	p31-00 p49-00 p79-00 p127-00 p201-00 p269-00 p309-00 p319-00 p329-00 p339-00
	2002	p1-02 p11-02 p21-02 p126-02 p166-02 p176-02 p186-02 p196-02 p216-02 p226-02 p266-02 p286-02 p306-02 p334-02 p344-02 p372-02
	2004	p63-04 p72-04 p102-04 p125-04 p132-04 p142-04 p192-04 p202-04p222-04 p294-04 p304-04 p350-04 p360-04 p378-04 p388-04 p396-04 p419-04 p497-04 p564-04
explanatory	2000	p21-00 p241-00
(total 17)	2002	p31-02 p41-02 p49-02 p206-02 p216-02 p226-02
	2004	p182-04 p212-04 p252-04 p284-04 p332-04 p429-04 p487-04 p604-04 p614-04
bibliographic	2004	p582-04
not empirical	1998	p19-98 p59-98 p69-98 p99-98 p119-98 p129-98 p139-98 p169-98 p179-98 p189-98 p207-98 p227-98 p237-98 p247-98 p257-98 p295-98 p305-98 p325-98 p335-98 p383-98
(total 51)	2000	p41-00 p59-00 p69-00 p107-00 p117-00 p145-00 p171-00 p181-00 p191-00 p211-00 p221-00 p231-00 p279-00
	2002	p58-02 p68-02 p77-02 p87-02 p236-02 p246-02 p256-02 p364-02
	2004	p11-04 p92-04 p112-04 p162-04 p262-04 p409-04 p437-04 p447-04 p457-04 p523-04
could not evaluate		p89-98 p373-98

Table 2. References for the type of empirical research. See text for the explanation of the format of the reference.

Reference	Types of empirical research	
p241-00	D&E	explanatory
p216-02	descriptive	explanatory
p226-02	descriptive	explanatory
p192-04	D&E	descriptive

Table 3. References classified in more than one type of empirical research.

evaluation type	year	references
lab experiment	1998	p11-98 p109-98 p149-98 p197-98 p217-98 p265-98 p275-98 p345-98
(total 30)	2000	p1-00 p11-00 p97-00 p135-00 p155-00 p163-00 p241-00 p251-00 p289-00 p299-00
	2002	p96-02 p106-02 p116-02 p324-02 p354-02
	2004	p1-04 p172-04 p232-04 p242-04 p274-04 p467-04 p477-04
field experiment	1998	p285-98 p363-98
(total 16)	2002	p116-02 p136-02 p146-02 p156-02 p296-02 p314-02
	2004	p29-04 p39-04 p82-04 p152-04 p192-04 p232-04 p533-04
field study	1998	p1-98 p11-98 p149-98 p159-98 p275-98
(total 11)	2000	p89-00 p261-00
	2002	p276-02
	2004	p49-04 p314-04 p554-04

Table 4. References for the evaluation type.

quant/qual	year	references	
		design and evaluation	descriptive
quantitative	1998	p1-98 p149-98 p159-98 p275-98 p345-98 p363-98	p355-98
	2000	p1-00 p11-00 p135-00 p163-00 p241-00 p251-00 p299-00	p127-00 p309-00 p319-00
	2002		p176-02 p286-02 p334-02
	2004	p1-04 p172-04 p274-04	p294-04 p497-04
total		16	9
qualitative	1998	p109-98 p197-98 p217-98 p285-98	p29-98 p39-98 p49-98 p315-98 p393-98
	2000	p89-00 p155-00 p261-00 p289-00	p31-00 p49-00 p79-00 p201-00 p269-00
	2002	p96-02 p106-02 p136-02 p156-02 p296-02 p324-02 p354-02	p21-02 p166-02 p186-02 p196-02 p226-02 p266-02 p344-02 p372-02
	2004	p39-04 p82-04 p152-04 p192-04 p467-04 p533-04	p63-04 p72-04 p125-04 p132-04 p142-04 p202-04 p222-04 p304-04 p350-04 p378-04 p388-04 p396-04 p419-04 p564-04
total		22	33
both	1998	p11-98 p265-98	
	2000	p97-00	p329-00 p339-00
	2002	p116-02 p146-02 p276-02	p1-02 p11-02 p126-02 p216-02 p306-02
	2004	p29-04 p49-04 p232-04 p242-04 p314-04 p477-04	p102-04 p360-04
total		12	9

Table 5. References for the quantitative/qualitative dimension.

Research type	1998	2000	2002	2004	total
design and evaluation	12 (30%)	12 (33%)	11 (28%)	16 (30%)	51
descriptive	6 (15%)	10 (28%)	16 (41%)	19 (35%)	51
explanatory	0 (0%)	2 (6%)	6 (15%)	9 (17%)	17
bibliographic	0 (0%)	0 (0%)	0 (0%)	1 (2%)	1
not empirical	20 (50%)	13 (36%)	8 (21%)	10 (19%)	51
could not evaluate	2 (5%)	0 (0%)	0 (0%)	0 (0%)	2
total	40	36	39	54	

Table 6. Totals for the type of empirical research. Percentages refer to the totals for each year. Each column does not necessarily adds to the column total because of the papers that are classified in multiple types.

Evaluation type	1998	2000	2002	2004	total
lab experiment	8 (67%)	10 (83%)	5 (45%)	7 (44%)	30 (59%)
field experiment	2 (17%)	0 (0%)	6 (55%)	7 (44%)	15 (29%)
field study	5 (42%)	2 (17%)	1 (9%)	3 (19%)	11 (22%)
quantitative	6 (50%)	7 (58%)	0 (0%)	3 (19%)	16 (31%)
qualitative	4 (33%)	4 (33%)	8 (73%)	6 (38%)	24 (47%)
both	2 (17%)	1 (8%)	3 (27%)	6 (38%)	13 (25%)
total D&E	12	12	11	16	51

Table 7. Totals for the evaluation type and qualitative/quantitative dimension for the design and evaluation research. Percentages refer to the totals for each year.

of the explanatory research. We feel that this "hypothesis testing" kind of research is a novelty in the Computer Science area, and it is gratifying that CSCW which, to the authors knowledge, was the first field to accept and value descriptive work, is now increasingly accepting this form of scientific inquiry. If we trace the history of CSCW as starting with the development of new, creative, and exciting groupware systems, the explanatory research places the field in a more mature position, in some way approaching the other natural sciences. Creativity, in natural sciences, comes in the proposal of new theories and hypothesis, and progress comes from systematically testing the theories and hypothesis with well planned, precise, and rigorous experiments.

7. CONCLUSIONS

This paper analysed all full papers accepted in the ACM/CSCW from 1998 to 2004 and classified them into 6 classes according to the type of empirical research they report. The results show a constant decrease in the number of papers that does not report an empirical result, a constant proportion of papers that propose a new groupware system and evaluate it (in some form), and an increasing number of papers that describe a work/collaboration situation and papers that test hypothesis by using experiments. There was only one example of bibliographic research. For evaluations, the most common type of evaluations are, in decreasing frequency, lab experiments, field experiment, and field/case study. Evaluations are mostly only qualitative, followed by only quantitative, followed by both. Descriptive research is mainly qualitative, and then in equal frequency quantitative and both.

As future work, we think it would be interesting to further divide the not-empirical category into some subclasses. At a first approximation, the not-empirical class can be further divided into papers that propose/describe a groupware but have no evaluation of the system, papers that propose frameworks to develop groupware, papers that present algorithms (more frequently, algorithms for synchronous editing), and taxonomy/model proposal papers.

An important contribution would be to extend this research into all CSCW conferences and journals. Of course, it would not be possible to select and analyze all papers in these conferences, and thus a random selection would be more appropriate. This wider view of the field would allow not only to see larger, international trends, but would also allow one to find the "characteristics" of the different conferences and journals. Finally, for the design and evaluation papers, it would be interesting to classify them into some of Pinelle and Gutwin's other di-

mensions.

This work has some limitations. The first one is that, although the ACM/CSCW conference is considered an important conference, it cannot be taken as the standard for publications in the CSCW area. The second limitation is that the authors developed the model in section 3 during the review process itself. As we mentioned, we started with the goal of repeating Pinelle and Gutwin's analysis for the 1998-2004 period. During the review process, we realized that groupware evaluation is only one of the types of empirical research possible. We proposed the class of exploratory research only after encountering a few examples of that form of research, and realizing that although they included some sort of groupware support, they did not fit well our working definition of groupware evaluation, because the the groupware used was not original or "interesting". Thus, the review criteria was developed as we reviewed the papers. Systematic reviews of the literature are usually done using a criteria defined beforehand. We do not know if the way this review was performed adds a systematic bias to the results.

REFERENCES

[1] ACM. TOC: Computer Supported Cooperative Work. http://portal.acm.org/toc.cfm?id=SERIES296, 2007.

[2] P. Antunes, M. Borges, J. Pino, and L. Carriço. Analyzing groupware design by means of usability results. *Proceedings of the Ninth International Conference on Computer Supported Cooperative Work in Design*, 2005.

[3] R. M. Araujo, F. M. Santoro, and M. Borges. The CSCW Lab for groupware evaluation. In *Groupware: Design, Implementation and Use: 8th International Workshop, CRIWG 2002*, number 2440 in Lecture Notes in Computer Science, pages 222–231. Springer, 2002.

[4] R. M. Araujo, F. M. Santoro, and M.R.S. Borges. The CSCW Lab ontology for groupware evaluation. In *Proceedings of the 8th International Conference on Computer Supported Cooperative Work in Design*, pages 148– 153. IEEE Press, 2004.

[5] R. Baeza-Yates and J. A. Pino. A first step to formally evaluate collaborative work. In *GROUP '97: Proceedings of the international ACM SIGGROUP conference on Supporting group work*, pages 56–60, New York, NY, USA, 1997. ACM Press.

[6] Andrew L. Cohen, Debra Cash, and Michael J. Muller. Designing to support adversarial collaboration. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 31–39, New York, NY, USA, 2000. ACM Press.

[7] K. Eason and W. Olphert. Early evaluation of the organizational implications of CSCW systems. In P. Thomas, editor, *CSCW Requirements and Evaluation*, pages 75–89. Springer, 1996.

[8] R. L. Glass, V. Ramesh, and I. Vessey. An analysis of research in computing disciplines. *Commun. ACM*, 47(6):89–94, 2004.

[9] Rebecca E. Grinter and Leysia Palen. Instant messaging in teen life. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 21–30, New York, NY, USA, 2002. ACM Press.

[10] J. Grudin. Groupware and social dynamics: eight challenges for developers. *Commun. ACM*, 37(1):92–105, 1994.

[11] E. G. Guba and Y. S. Lincon. *Effective evaluation*. Jossey-Bass, 1981.

[12] Lena Mamykina and Catherine G. Wolf. Evolution of contact point: a case study of a help desk and its users. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 41–48, New York, NY, USA, 2000. ACM Press.

[13] J.E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In R.M. Baecker, J. Grudin, W.A.S. Buxton, and S. Greenberg, editors, *Readings in Human-Computer Interaction: Toward the Year 2000*, pages 152 – 169. Morgan Kaufman Publishers, 1995.

[14] Dennis C. Neale, John M Carroll, and Mary Beth Rosson. Evaluating computer-supported cooperative work: models and frameworks. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 112–121, New York, NY, USA, 2004. ACM Press.

[15] W. Orlikowski and J. Baroudi. Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, 2(1):1–28, 1991.

[16] D. Pinelle and C. Gutwin. A review of groupware evaluations. In *Proceedings of WET ICE 2000*, pages 86–91. IEEE Computer Society, 2000.

[17] D. Pinelle, C. Gutwin, and S. Greenberg. Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(4):281–311, 2003.

[18] A. Pinsonnealult and K. L. Kraemer. The impact of technological support on groups: An assessment of the empirical research. *Decision Support Systems*, 5(2):197–216, 1989.

[19] V. Ramesh, R.L. Glass, and I. Vessey. Research in computer science: An empirical study. *Journal of Systems and Software*, 70(1-2):165–176, 2004.

[20] M.P. Steves, E. Morse, C. Gutwin, and S. Greenberg. A comparison of usage evaluation and inspection methods for assessing groupware usability. *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, pages 125–134, 2001.

[21] L.A. Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1987.

[22] W. F. Tichy, P. Lukowicz, L. Prechelt, and E. A. Heinz. Experimental evaluation in computer science: A quantitative study. *Journal of Systems and Software*, 28(1):9–18, 1995.

[23] Steve Whittaker, Quentin Jones, and Loren Terveen. Contact management: identifying contacts to support long-term communication. In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 216–225, New York, NY, USA, 2002. ACM Press.