

Meta-research on the groupware studies: a case study for groupware evaluation methodologies

TOSHIHIKO YAMAKAMI†

Groupware evaluation has been one of the key issues in groupware research literature. The difficulty of the groupware evaluation comes from the social factors, long-term ness, and the limitations of the intuitive design. Groupware evaluation is the one of the key challenges in groupware studies. It is time consuming, prone to be impacted by social aspects and dynamics and influenced by culture and hidden factors. The author proposes the meta-research concept, which tries to figure out the groupware methodology issues from the past literature. The author analyzes the recent five-year SIG notes of GN in 1998-2003 and the first one-year SIG notes of GW in 1993-1994, in order to pick up the applicable methodologies. The SIG note patterns, reference social networks, and evaluation transition patterns are studied to clarify the current state of art in the groupware evaluation.

The results show that the evaluation and questionnaire result report are increased. At the same time, they show that the quantitative evaluation and mutual reference do not show the difference between the early SIG publications and the recent ones. The author outlines the strategy for the groupware evaluation studies and describes the exploratory results from this meta-research.

1. Introduction

The groupware evaluation is one of the challenging issues in groupware research. Grudin¹⁾ pointed out as follows:

Laboratory evaluation was at its peak effectiveness when isolated PC use very much resembled typical laboratory situations. Methods to design and develop in the new circumstances of socially situated use are an area of intense focus today, and will remain a strategic issue for the future.

He listed up the difficulty of evaluation as the 6th issue in the eight challenges for eight challenges for developers of groupware²⁾:

Task analysis, design, and evaluation are much more difficult for multi-user applications than for single-user applications. ... Evaluation takes longer. ... The absence of definitive studies ensures that other researchers and developers will repeat costly mistakes. ... The almost insurmountable obstacles to meaningful, generalizable analysis and evaluation of groupware prevent us from learning from experience.

Groupware evaluation is a challenge that was recognized from the very first day of the groupware research. There are many literature on the groupware evaluation difficulties. The author tries to capture the long-term transition of the groupware evaluation methodologies in order to identify the current status of the evaluation-empowered groupware research. This paper is an attempt to track the groupware research evaluation methodologies using the SIG groupware publications as the meta-research materials.

2. Challenges

2.1 Groupware Evaluation

Groupware evaluation has several challenges to cope with from the birth of the research in 1980's:

- comparisons need identifications of a wide range of social contexts,
- collaboration needs multi-facet analysis, and
- collaboration results need a long term analysis with changing contexts.

These challenges make the simple comparison meaningless. The tasks and situations are difficult to control, especially over a long span of time. There is no conclusion on the methodology evolution because each experiment has the different contexts and backgrounds. It is mean-

† ACCESS, 2-8-16 Sarugaku-cho, Chiyoda-ku, Tokyo, JAPAN, e-mail:yam@access.co.jp

ingful that the scattered challenges are summarized from the long-span observation.

2.2 Related Work

Due to the comparison difficulties on the different background and context groupware research, the meta-research on the groupware research is rarely found.

2.3 Meta-Research

Meta-research is a field of research on research publications. The groupware SIG has its origin back to 1992. After a decade of the SIG activities, the SIG notes themselves could be a source of research how the groupware research was done. The research activity has the social, interactive, and organizational factors.

3. Preliminary Research

For the exploratory research, the author examined the technical notes from 1998 to 2003.

4. Method

4.1 Samples

The author analyzes 252 articles from GW-31 to GN-48. GW-31 took place in January 1999. GN-48 took place in May 2003. 252 articles includes invited talks and co-located meetings with other SIGs.

4.2 Analysis Targets

The author examined the following items:

- Existence of evaluation,
- Existence of quantitative evaluation,
- Existence of quantitative analysis materials, and
- Relations of references.

The identification of the existence of evaluation needs the consistent measure. When the major evaluation focused the system performance, the quantitative measures were easy to describe. In this research, the evaluation is taken in a broad view. When the author describes the section as evaluation or comparison to the other research, it is counted as evaluation. When the paper was an invited talk, it is common that the detailed evaluation was omitted and the general framework and direction were discussed. Over a long span of time, the invited talk tracking is sometimes difficult. Some of the invited talks described the detailed discussion of the quantitative discussion. In this exploratory analysis, the invited talk papers are included without any discrimination. Some of the papers are

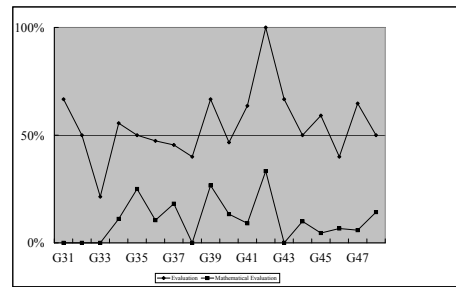


Fig. 1 Evaluation Ratio for SIG Groupware Notes in 1998-2003

presented in the joint workshops with the other SIG. It is difficult to identify the original SIG to which the papers were submit. The joint workshop papers are included without any discrimination.

4.3 Research Assumptions

There are following research assumptions in this research:

- The quantitative analysis increases over a long span of time,
- The quantitative analysis methodology is shared and diffused over a long span of time,
- The approach to the quantitative analysis increases over a long span of time, and
- The quantitative evaluation methodologies increases its significance in the research domains.

5. Results

The evaluation and mathematical evaluation ratio for SIG Groupware Notes in Information Processing Society of Japan (IPSJ) from 1998 to 2003 is shown in **Fig. 1**. In general, approximately half of the publications aimed at the evaluation. It means that there was special description of evaluation or comparisons in the publications. The change of the evaluation ratio over a long span of time is not clear from Fig. 1. The mathematical evaluation was done approximately 10 to 20 % of the publications. The change of the ratio over a long span of time is not clear from Fig. 1.

Each SIG Notes publication has a different coverage of the research. Therefore, the quantitative evaluation ratio is not stable.

In order to compare the long-term transition of the evaluation research, the similar analysis was done on the early days of the groupware

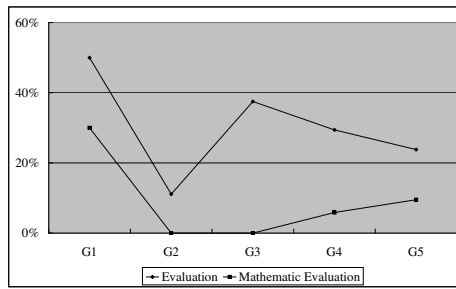


Fig. 2 Evaluation Ratio for SIG Groupware Notes in 1993-1994

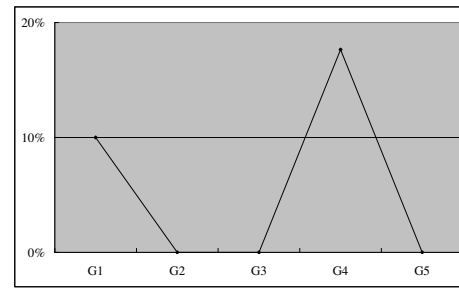


Fig. 4 Questionnaire Ratio for SIG Groupware Notes in 1993-1994

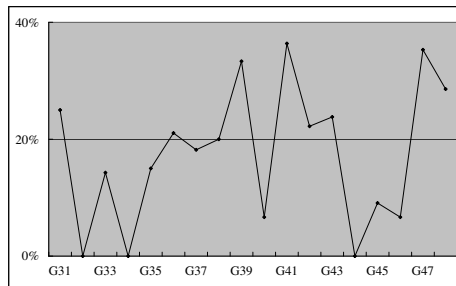


Fig. 3 Questionnaire Ratio for SIG Groupware Notes in 1998-2003

research during 1993 and 1994. The result is depicted in **Fig. 2**.

The evaluation is a part of the research methodologies. Each publication has the different research stage. In order to identify the demands for evaluation methodologies, the author takes another view on the past literature. Between the research incubation and the stable research result evaluation, there could be the intermediate stage. The author focuses to identify a measure to capture such an intermediate stage. In this paper, the author takes a questionnaire report in the publications. From the literature survey, it is noted that the many publications covered the questionnaire results. The author performed an analysis on the ratio of the publications with questionnaire results. This is one of the indications that each publication tries to head for the groupware evaluation. The ratio of the publications with questionnaire results from 1998 to 2003 is shown in **Fig. 3**.

In order to identify the long-term evolution, the similar analysis is done in the publication during 1993 to 1994 and the result is shown in **Fig. 4**.

There are fluctuations among SIG Notes publications. However, the ratio of the publications with questionnaire results are significantly increased in these comparisons.

6. Discussions

6.1 Changes of Evaluation Methodologies over Time

In order to identify the evaluation diffusion in the groupware research, the author focuses on the references in the past literature. For comparison, the author executes the similar research on the early stage of the SIG groupware in early 1990's. The result is shown in **Fig. 5**. In the early research, the number of the reference papers was limited. On the contrary, the research community was tight and small, therefore, the groupware-focused papers had the tendency to refer the SIG-groupware Notes frequently. From the reference ratio, these two conflicting factors were balanced. There is no significant increase of the SIG-groupware Notes references. The most common references are observed in the references to the same author's previous publications. It shows the natural research evolution in each research project. The occurrences of the evaluation methodology diffusion are rarely observed. It should be noted that this analysis is based on the SIG-groupware publications only. When the journal papers and workshop papers are included in the research evolution studies, the results may come to the different conclusions. It is for further studies. From the exploratory studies on the mutual references, the author's impress is that the research diffusion in the groupware research is not common in the past literature. There are a small number of excellent quantitative eval-

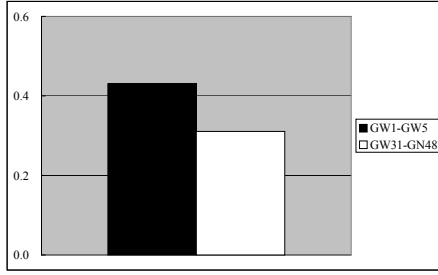


Fig. 5 Mutual Reference Ratio

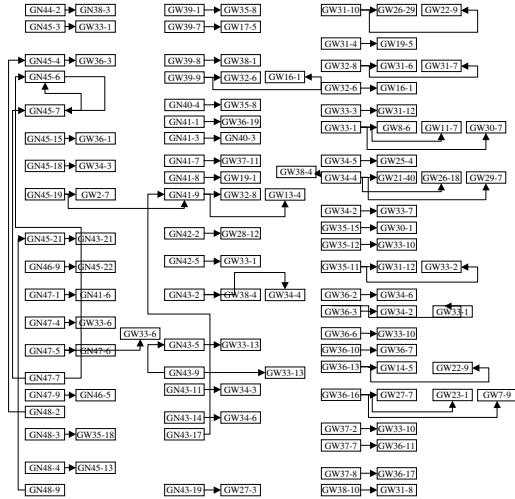


Fig. 6 Reference Topology Graph

uation publications including a new proposal of the quantitative measures. The visible evidence from the impacts from these publications are not clear in this study.

The reference relationship diagram is shown in **Fig. 6**. The author calls it as *research sociogram*. It depicts the relationship between research publications.

In general, the collaboration process brings the social relationship diagram into a closely related status. Such a result is reported in³⁾. Over a long span of time, the evolution of the collaboration tends to form an active core group in which each member has strong and active relations and a satellite group in which one of the core group members has active relations with other satellite members. From the exploratory study on this *research sociogram*, it shows the

early stage of the isolated research interaction. There is no evidence to identify any core group activities in the research sociogram. The research publications cover a wide range of groupware research domains. Therefore, the accuracy of the research sociogram on the groupware research collaboration status is not clear in this early stage of the meta-research. From the graph viewpoint, Fig. 6 shows a shallow binding and isolated characteristics. The relationship chain is not long and generally isolated. From this exploratory analysis, the research sociogram shows a very sparse graph structure. The major structure is an isolated serial structure denoting the single project evolution. It needs further studies to identify any social structures using this type of research socio-gram analysis.

6.2 Quantitative Analysis on the evaluation evolution over a span of time

In **Table 1**, the contingency table with the categorical variable, GW1-GW5(1993-1994) and GW32-GN48(1998-2003) is tested using χ^2 test. The null hypothesis is that there is no relationship between the category variable and the frequency of the evaluation, quantitative evaluation, questionnaire report and the number of references.

The results show that the null hypotheses on the evaluation and questionnaire report are false. At the same time, the null hypotheses on the quantitative evaluation and reference existence are true. The category variable is the time of the publications, whether it is in the early publication (during 1993-1994) or in the recent publication (during 1998-2003). This shows that the existence of the evaluation sections shows the increase over a span of time. The existence of the questionnaire report shows the same trend. However, the mutual reference existence in the SIG does not show the difference over a decade. The existence of quantitative evaluation shows no statistically significant

Table 1 χ^2 test on the null hypothesis among GW1-GW5 and GW32-GN48

	χ^2	level of significance
evaluation	10.6226	0.0011
quantitative evaluation	0.0027	0.9585
questionnaire report	12.0283	0.0005
reference	0.1787	0.6725

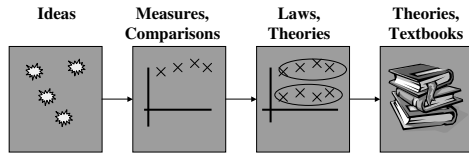


Fig. 7 A 4-stage Model of Research Evolution

cant difference. This shows that the evaluation efforts are increased and that the quantitative evaluation is not a common practice in this research domain. There is no strong evidence between the mutual influences among the research publications in the SIG from the reference existence viewpoints. This needs further considerations as follows:

- Influence of the journals and symposia,
- Influence of the external proceedings in IEEE or ACM,
- Reference convention,
- Accuracy of the use of the reference as the meta-research material.

These issues are for further studies.

6.3 Lessons learned in the meta-research

The literature analysis needs careful design due to the wide variety of the description styles in the publications. In the preliminary analysis, the author used the simple evaluation mark analysis to mark the evaluation criteria. It added the difficulty in the long-term publication tracking. This method is difficult to verify in the iterated analysis. In the second experiment, the author takes the skeleton analysis to describe the topic and the story lines of each publication, which helps the pair-wise comparison when the publication looked the borderline case in the quantitative analysis.

6.4 Implications for the research direction of groupware evaluation

The four stage model proposed by Olson et al⁴⁾ is outlined in **Fig. 7**. It should be noted that

It should be noted that there are innovative research based on insightful ideas. When there is a dominating methodology for evaluation, it could be time for the research domain to be

closed. It is important to incubate both of the innovative ideas and quantitative evaluation methodologies. The straightforward application of the established quantitative methodologies is sometimes dangerous because it neglects the group work's in-depth social analysis and dynamism over a long span of time. The methodologies to identify the groupware specific characteristics are still to be explored. Examples include:

- Time-scale evaluation like time zone analysis, day of week analysis, and group work time rhythm analysis,
- Work culture analysis,
- Trust and Norm development in systems,
- Group dynamics analysis, and
- Small group analysis.

For the group work evaluation, the methodologies to identify the group factors in a small group configuration are still to be explored. For example, it is difficult to obtain convincing results when the work experiment is done in the small group closely related to the research. In such a configuration, it is important to put emphasis on the exploratory evaluation. For example, interview, or 2-stage questionnaires to identify the specific factors, and combination evaluation using both of traffic analysis and questionnaires. The single questionnaire is difficult to construct the quantitative discussions. The 2-stage questionnaire consists of the first stage free form questionnaire and the second stage in-depth score-based questionnaire. The score-based questionnaire can build a step towards the future quantitative analysis. When the quantitative verification is difficult, it is important to identify the mental model analysis, stage analysis, and evolution analysis. The basic materials like input/output data sets to compare the results are not mature in this research field.

6.5 Evaluation as Meta Research

The research publication itself can provide the implications for the collaboration. The research publication is a set of social activities rather than a set of independent activities. Each publication can impact another. Several research activities can form a cluster of research in a certain research domain. The series of the publication can form a research development time line. The ideas and methodologies

can propagate over a set of publications. From the meta-research of the research, the most apparent examples include the chronic development of a research project. Second, the research projects can collaborate. The propagation of methodologies is not apparent. For the meta-research, the implications are as follows. At first, the author examined the evaluation criteria for each research publication. After some endeavors, the author recognized that the description evaluation criteria are needed. For example, the definitions of quantitative research needed the criteria to identify the relationship to the evaluation. There are three major cases for detailed analysis. The first case includes formula definitions and formal description of mathematical models. The relations of such mathematical descriptions and evaluation needed the categorization criteria. The second case includes the traffic analysis. A wide range of quantitative measures of systems was described. This needs the in-depth categorization. The third case includes the factor analysis. This is a completely mathematical result, however, it does not directly lead to the evaluation. For clarity, in this research, the qualitative results from interviews are not counted as quantitative evaluation. When there was a long-term operation result, it is counted as evaluation.

7. Conclusions

The groupware and network services exploit a new research domain empowered the advanced networks. The research needs exploration in the methodologies because it needs a complicated analysis in the social issues in the technology-augmented environment. The design intuition cannot directly apply to the collective behavior in the group works. In addition, the group work tends to have a long life cycle during which a wide range of social factors evolves. This outreach in the space scale, time scale and social scale in the group work makes the group work research more challenging. In this research, the author takes a new view on the group work research itself when the Japanese group work research in the SIG-groupware marks its first decade in the information processing society of Japan. In order to explore the evaluation research direc-

tions in Japanese groupware research, the author performed exploratory analysis on the past literature in the SIG-groupware. The evaluation positions in each publication in the SIG-groupware were analyzed in the long-term transition. In addition, the publication relationship is analyzed from the reference sections. The results show that the evaluation and questionnaire result report are increased. At the same time, they show that the quantitative evaluation and mutual reference do not show the difference between the early SIG publications and the recent ones. The author outlines the strategy for the groupware evaluation studies and describes the exploratory results from this meta-research. The implications for the groupware evaluation research are presented. There are several insightful research projects towards quantitative groupware evaluation. It is also to be noted that there are still many papers with questionnaires to be evaluated in some quantitative measures. The bird-eye view of the Japanese groupware evaluation methodologies is described in order to facilitate the further exploration of the Japanese groupware evaluation methodologies.

References

- 1) J. Grudin, "The end of exile," *ACM Computing Surveys*, vol. 28, no. 4, pp. 135, December 1996.
- 2) J. Grudin, "Groupware and social dynamics: eight challenges for developers," *CACM*, vol. 37, no. 1, pp. 92–105, 1994.
- 3) T. Yamakami, "Information flow analysis: An approach to evaluate groupware adoption patterns," *Trans. IPSJ*, vol. 36, no. 10, pp. 2511–2519, October 1995.
- 4) J. S. Olson, S. K. Card, T. K. Landauer, G. M. Olson, T. Malone, and J. Leggett, "Computer-supported co-operative work: Research issues for the 90s," *Behaviour and Information Technology*, vol. 12, no. 2, pp. 115–129, March 1993.