Group Tutoring and Formative Asynchronous Peer Assessment using e-learning Technologies to Complement Staff Tutoring in Academic Writing

Björn Hedin
KTH School of Computer Science and Communication

Key words: academic writing, peer tutoring, formative peer assessment

Abstract:

Writing a master's thesis is a lonely task, which often takes longer than the nominal time. A previous study has shown that a way to counter this is to give tutoring in groups instead of individually. This paper describes and evaluates an attempt to complement individual staff tutoring with group tutoring, formative peer assessment and a self-study methodology course, using e-learning methods and without increasing the total teacher time invested for each student. The results clearly indicate the attempt was highly successful, with increased quality of the reports, decreased time for students to complete their work, and very satisfied students who believed these activities were roughly of equal use as individual tutoring even though they only consumed 15% of the total time invested by the teachers in the tutoring activities.

1 Introduction

Writing a master's thesis is the final part of five years of studies to get a master of science degree in media technology at KTH\(^1\). This, however, in many ways differ from the earlier education undertaken. Identifying the actual task at hand is now the responsibility of the student, there is no set deadline, the work has much higher demands on a scientific level, and there are no peers to ask for advice. The only formal support in this has traditionally been the tutor, who can find it difficult to have find to give feedback exactly when the students need it. Many students have found it difficult to work like this, and earlier evaluations at KTH have suggested this as a cause for lowering the quality of the students' work, increasing the time to complete their work and decreasing probability of ever completing the work [1].

An earlier study by at Lund University has shown that replacing individual tutoring with group tutoring was a way to decrease the time for students to complete their theses, increase student satisfaction and decrease teacher time per student, while maintaining the same level of quality [2]. Important factors were deadlines combined with group pressure and the possibility to discuss problems with peers instead of just with a supervisor. This paper, while inspired by this earlier study evaluates a different approach where the individual tutoring remains, but 10% of the total tutoring time allocated is instead transferred to supervising a peer tutoring and formative peer assessment activity, using distance e-learning methods and technology. Peer assessment, as defined by Topping et al. as "an arrangement for peers to consider the level, value, worth, quality or successfulness of the products or outcomes of learning of others

\(^{1}\) The Royal Institute of Technology, Stockholm, Sweden
of similar status"[3], has been found to yield outcomes for writing at least as good as teacher assessment and sometimes better [4]. 5% of the total tutoring time was also transferred to assessing a new self-study course in research methodology introduced at the start of the thesis works.

1.1 **Rationale for peer assessment**

It could be argued that the time associated with assessing peers takes valuable time from the "real task at hand", namely writing your own thesis. However, this is based on the assumption that the goal of the thesis writing is only the end product, and not the process. Biggs lists three main advantages of self assessment (SA) and peer assessment (PA).

1. SA and PA give the students first-hand, active involvement with the criteria for good learning.
2. Students learn how to select good evidence.
3. Judging whether a performance or product meets given criteria is vital for effective professional action.

These points are especially true for thesis work, where there is no "correct answer", and where the students really need to critically analyze what they do, what they should learn in the process, and how they should prove what they have learned.

1.2 **Rationale for formative assessment**

Biggs differs between many reasons for assessing students, the two most being:

1. "Formative assessment, the results of which are used for feedback during learning. Students and teachers both need to know how learning is proceeding. Feedback may operate both to improve the learning of individual students, and to improve teaching.
2. Summative assessment, the results of which are used to grade students at the end of a unit, or to accredit at the end of a programme."[5]

According to Rust [7] summative assessment is commonly used at universities, but formative assessment is less common. The summative assessment is also often done at the end of the course, when there is no time and little interest from the students to address the feedback of the assessment. Indeed, as pointed out by Biggs [5] "Error no longer is there to instruct, … error now signals punishment". Both Rust and Biggs further argues that it would be beneficial for the students to have more opportunities to learn from their mistakes by receiving feedback, that is formative assessment, during the course instead of just summative assessment after the course. I argue that the most important aspect of assessment is not awarding grades, but instead to provide feedback on how to bridge the gap between the students' current knowledge and beliefs and the knowledge and believes that are the target learning outcomes of the course or education, that is formative assessment.

1.3 **Rationale for formative peer assessment**

Biggs argues that "Students need to take over the formative role themselves…", and further argues that "…self- and peer assessment are particularly helpful for this". As for peer assessment in general, Falchikov [8] lists a number of questions that could be used by students and teachers against peer assessment. Many are dealing with how to handle fairness, like students agreeing to award each other high grades, settling old scores, fear of reprisals,
existing friendships etc. However these problems arise due to using peer assessment as a basis for the actual grades the students will get in the end of the course, that is for summative assessment. If the assessment is instead formative, these problems disappear.

1.4 Theoretical framework

A theoretical ground for peer assessment can be found in the 'zone of proximal development', defined by Vygotski as "the distance between the actual development as determined by individual problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" [9]. Peers giving feedback to each other and sharing experiences can be seen as a form of scaffolding [10], which is an effective way to access the zone of proximal development. Vygotskij further stressed the importance of social interaction for learning to take place, and this is exactly what was lacking in the process for our student's academic writing.

An important factor in peer learning is also that it should be beneficial for all. This is addressed by Equity Theory, which states that all partners in a relationship should experience an equal ratio of benefit to effort expended [11]. If one student benefits other students more than the other students benefit her, there is an imbalance which can cause discontent. As we shall see later, this was one of the problems encountered in this study.

1.5 Descriptions of the reform

The changes in the procedures and formal requirements for writing master theses at KTH, hereafter named "the reform", consisted of two major parts: The introduction of a self-study methodology course and the introduction of a group activity. Since the economical framework had not changed, resources in the form of teacher-man-hours were moved from the individual supervision to the other two parts. Before the reform the teacher had 20 hours to supervise each student, after the reform this was reduced to 17 hours, 2 hours were allocated to leading the group activities and 1 hour was allocated to giving feedback on the self-study methodology course.

1.5.1 The self-study methodology course

This course could be commenced at any time by the student. The course consisted of five tasks with corresponding literature. Each task was assessed by a short report, which were handed in using a learning management system. The total expected work time for the student was three weeks. The tasks were

1. An introduction to research
2. Analyzing quality of reports
3. Formulating scientific problems
4. Questionnaires, interviews, selection, reliability, validity
5. Literature search, source criticism and references

1.5.2 The group activities

When the students registered for starting their thesis work, they were also put on a waiting list for group supervision. When about five students had signed up they were contacted by the group supervisor who is a teacher at the department, and a date for an initial physical meeting was set. At the initial meeting the students met and were introduced to each other, and presented what their work was going to be about.
The group had access to a private web based discussion forum where they could hand in their assignments and discuss their work, but also had access to a discussion forum where all students currently writing their theses had access. The group supervisor loosely followed the discussions in the group fora, and replied when necessary, but the main responsibility to answering questions and giving feedback laid on the group members.

The formal requirements were to complete five assignments related to the phase they ought to be at at certain times. The deadlines were decided at the first physical meeting. Then each member was required to formatively assess the work of all the other group members, and hand in their assessments in the discussion forum. The assignments progressed from the first assignment, where the task was to formulate suitable research questions and to give suggestions of literature to read, to the last assignment, where the task was to deliver an almost complete thesis.

1.6 Research questions

The research questions investigated in this paper were:

- To what extent has the quality of the theses improved?
- To what extent has lead time and active work time decreased, in terms of the time for students to complete their theses?
- What are the student's attitudes towards the individual supervision, the group supervision and the self-study research methodology course as parts of their thesis writing activities?
- To what extent has the four supportive methods contributed of individual supervision, company supervision the peer assessment and the self-study research methodology course contributed to the quality of the reports, formal content\(^2\), formulating research questions, choosing suitable research methodology and research methods, practical realization of the task, give structure to the day-to-day work and finally to act as a whip and carrot for progressing at a suitable pace?
- Do the students prefer virtual or face-to-face seminars?

2 Methodology

As methodology, action research was used in the meaning defined by Cohen and Manion: "a small-scale intervention in the functioning of the real world and a close examination of the effects of such an intervention"[12]. In this case the intervention was perhaps on a slightly larger scale, with the introduction of the formative peer assessment, group tutoring and self-study methodology course.

Several methods were used for data collection, as described below.

2.1 Lead-time and active work time

The lead-time, meaning the time it takes for a student to complete their theses, was measured using two different methods, in order to increase inner validity by triangulation. The first method was based on the time interval between when the students registered for the course and the time when the thesis was reported as completed. The second method was based on the

\(^2\) In the meaning of making correct references, using proper language etc.
students own estimations of the time used. The students also estimated the actual work time, meaning how many months of full-time work they had used for their theses. This difference is important in some cases, especially when the students are taking other courses or working in parallel. These data were collected using the questionnaires described below.

2.1.1 Lead-time based on registration date

The routines for registering and reporting the students were changed in connection with the reform, making it more difficult to compare the time from starting to completing the work. Before the reform, the students first found a suitable task, after which the registration took place. After that, the work was expected to take 20 weeks of effective work time.

After the reform, the registration of the students can take place a long time before the work on the individual task starts, since the self-study methodology course now was a part of the thesis work. Even though it was recommended that it was completed before the individual task was started, this was not formally required and was indeed not always the case. This meant that some students registered for the course one year before the individual task was started, while others registered just like before, that is when the individual task was set. Since the methodology course was expected to require about three weeks of work, the total time expected to be spent on the individual task was reduced to 17 weeks, in order to maintain the total expected workload of 20 weeks.

In order to compensate for the problems of comparing lead-time before and after the reform, the definition of start time was changed from registration date to the date when the supervision groups were formed. This event occurred when the individual task was set, which corresponds to the date the students were registered in the old system. However, since a group is formed only when there are enough students to form a group, some students in the group have started their individual tasks some weeks before this event. To compensate for this, the time interval between when the previous and the current group started their supervision was noted \((G_n - G_{n-1})\). The assumption was then made that the students in the new group signed up for group supervision evenly distributed over time, which gives that the average delay between registering for, and starting, the group supervision was \((G_n - G_{n-1})/2\), which should be added to the total time for getting a fair measure of the lead-time.

The formulas used to calculate the lead-time, and the corresponding definitions are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Formulas for calculating lead time.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(TB = A - R)</td>
</tr>
<tr>
<td>TB = Lead-time before the reform</td>
</tr>
<tr>
<td>A = Date of reporting the final, approved report</td>
</tr>
<tr>
<td>G_n = Date for forming the student's supervision group</td>
</tr>
<tr>
<td>M = Expected time for completing the methodology course</td>
</tr>
</tbody>
</table>
2.1.2 Lead-time and active work time based on the students' own estimations

The students gave their own estimations of the time to complete their theses, using both calendar time and effective working time. Even though this is more subjective than measuring registration dates, it is a good help for increasing the inner validity by triangulation, and for clarifying some extreme values caused by students who missed registering their activities.

2.2 Report quality

The very concept of quality is in itself difficult to define [15], and even more so in education [16]. Determining the quality of the report is furthermore a very subjective activity. In this case however, the examiner has remained the same person both before and after the reform, which decreases the reliability problems of having different people comparing their subjective views of report quality. This eliminates problems with inter-judge reliability, but the intra-judge reliability problem remains [17]. A complex scheme of judging report quality was developed, using a form with several different criteria to consider, but in the end lack of time restricted the evaluation of this very important aspect to an interview with the examiner where his subjective view of the change in quality was determined.

2.3 Questionnaire

To collect the students' views, a web based questionnaire was used. All students who had been assigned to a supervision group in the new structure and who had finished or were expected to finish their reports within one month, were contacted using email and SMS. All students who answered were given a lottery ticket. After two days, a reminder was sent by email to those who hadn't answered the questionnaire. After four days, the questionnaire closed and the analysis of the results commenced. Two students had forgotten the passwords required to complete the questionnaire, and are therefore not included in the population. A total of 35 out of the 64 students answered, giving a reply frequency of 55%.

To increase the validity of the questionnaire, the first version was read by the examiner, the teacher in charge of the group supervision and the teacher in charge of the methodology course. Their opinions were incorporated into the questionnaire which was again read by the same people, after which the third and final version was presented to the students.

The questionnaire was semi-structured [13], containing mainly quantitative questions, but also open questions. It was divided into different categories, each starting with a number of multiple choice questions and finishing with one or two open questions with free text answers. This was done in order to increase the reply frequency, allowing less motivated students to answer the questionnaire quickly using the multiple choice questions, while allowing the more motivated students to elaborate their answers in the open questions.

2.3.1 Questionnaire analysis

A number of factors were examined using the questionnaire, to determine how important the supportive methods were considered for each factor by the students. The factors were quality of the reports, formal content, formulating research questions, choosing suitable research methodology and research methods, practical realization of the task, give structure to the day-to-day work and finally to act as a whip and carrot for progressing at a suitable pace. The supportive methods were individual supervision, group supervision, company supervision and
methodology course. The results for each factor are discussed under the corresponding header in section 3.6.

Two types of diagrams are used for each factor. The first diagram shows the distribution of answers for each factor, normalized to a percentage. The second diagram shows the pairwise comparison of the effects of the four supportive methods for each factor for each student, where one factor is considered either more important, of equal importance or less important than another factor. As opposed to the first type of diagram, this gives the intra-relationships for the supportive methods for each student, meaning that for each factor, it ranks the four methods for each student.

Finally, a crude measure of ranking of the methods was made, comparing the percentage who answered with one of the two positive answers for each method/factor with each other.

### 2.4 Sources of error

One reliability problem is that the student groups have changed. Before the change, all students had studied one out of two specializations. After the change, a new group was added, which contributed to 65% of all theses. This group generally had higher grades before commencing higher education, which might mean they should be better qualified to write their theses, which could affect both the lead-time and the report quality.

Similarly, the group of supervisors has also changed which might affect the quality of the individual supervision either in a positive or negative way.

Neither of the error sources above should affect the students' subjective views of the importance of the four supportive methods compared to each other. However, the total population of 64 students, and the reply rate of 55% makes it difficult to draw statistically verified conclusions from the questionnaire, but the results in some cases gives clear indications.

### 3 Results

#### 3.1 Lead-time

The two methods used for measuring lead-time to some degree filled different purposes. The measurement of registration dates compared to completion dates could be used to compare how the reform has affected the throughput in relative terms. The students' estimations of calendar time and effective working time could be used firstly for validating the first measurement method, but also to interpret extreme results.

##### 3.1.1 Lead-time based on registration dates

Using the formulas described in section 2.1.1, the median time for completing a thesis before the reform was 8.6 months, and the average was 10.2 months. After the reform the median time was 7.3 months and the average time was 7.6 months. All these values can still increase if some of the students believed to have discontinued their work decide to complete the same. Since slightly more students who started their work during the old system have discontinued their work than students starting during the new system, and since completed theses started during the older students per definition will have taken longer time to complete it is likely that the difference will increase rather than decrease.
3.1.2 Lead-time and effective work time based on the students' estimations

Using the same adjustments for added time due to administration and the new methodology course as described in section 2.1.1, the students' estimations of the median calendar time for completing their theses was 7.8 months, which is close to the 7.3 months measured in with the registration dates, which indicates a high inner validity. The median time for actual full-time work was 5.1 months which can be compared to the nominal 5 months, which indicates the amount of work is according to plan.

3.2 Completion rate

Of all 78 students who had started the group supervision, 87% had completed their work and another 9% were considered to be in phase with their work, mainly students who recently had started their work. 6% were considered to, at least temporarily, have discontinued their work.

The corresponding figures for the 96 students starting their work before the reform, 92% had completed their work and 8% discontinued their work.

3.3 Student attitudes to group activities

The students believed the idea of group activities, like the group tutoring and formative peer assessment, was very good as seen in Figure 1. In the questionnaire, 91% of the students choose one of the two positive alternatives, and no-one gave the worst of the four possible alternatives.

![Figure 1. Students' answers to the question "what is your opinion of the idea of the group activities as a mandatory part of the master thesis course ?" (n=35)](image)

The main problems according to the free text answers in the questionnaire were

- Not all students handed in their assignments and comments of other's assignments on time.
- The group members were at different stages already when the group was created.
- Many comments made by peers were about details and not considered very constructive.
- The group members often worked in very different fields.
- The final assignment, where the students were require to read through and give comments on all other group members near-completed reports required too much time.

One goal of the group activities was to broaden the students' perspectives, making them confront the problems and ideas other students worked on in their theses. Due to this, a large portion of the group activities were directed to reading fellow students report drafts and to
assess them formatively. To examine the students' attitudes to this, the question "to what extent was it useful\(^3\) to read and comment other students' work?" was asked. The students' answers, as seen in Figure 2, shows that the students believed the activity was very useful, where 76% responded with one of the two positive alternatives and no-one believed the activity to be of no use at all.

![Figure 2](image)

**Figure 2.** Students' answers to the question "to what extent was it useful to read and comment other students' work?" (n=34)

### 3.4 Student attitudes to the distance learning aspect

One important question was whether the group activities should be traditional seminars, or like in this study virtual. This is especially interesting since these students are not traditional distance learning students, rather having studied in a traditional environment for four years. Two questions in the questionnaire were asked in order to examine this aspect.

![Figure 3](image)

**Figure 3.** Students' answers to the question "to what extent was it positive from a learning- and practical point of view to have virtual seminars instead of traditional seminars?" (n=32)

As shown in Figure 3 69% of the students responded with one of the two positive alternatives regarding whether the virtual seminars were a positive experience from a learning and practical point of view compared to traditional seminars. On the question whether they preferred virtual or traditional seminars the difference was even greater, where 76% preferred virtual seminars as shown in Figure 4. One reason for this was the increased remote mobility, where the students were not required to be at a specific location at a specific time REF, something which is difficult when the work takes place at a company in the same town, in another town or even in another country.

\(^3\) The Swedish word "lärorik" has a meaning corresponding roughly to "useful", "instructive", "a good learning activity".
3.5 Self-study methodology course

The students also believed the idea of a self-study methodology course was very good as illustrated in Figure 5, where 91% responded with one of the two positive alternatives.

3.6 Examined factors

The students' subjective views of the help the four supportive methods gave for a number of factors were determined using the questionnaire described in section 2.3 and analyzed using the methods described in section 2.3.1. The supportive methods were individual supervision, group supervision, company supervision and a methodology course. The factors examined were quality of the reports, formal content, formulating research questions, choosing suitable research methodology and research methods, practical realization of the task, give structure to the day-to-day work and finally to act as a whip and carrot for progressing at a suitable pace. The results for each factor are discussed under the corresponding header below. For a description of the diagrams used, see section 2.3.1.

3.6.1 Report quality

The question "to what extent do you believe that $X$ affected the quality of your final report in a positive way", where $X$ was substituted for the four supportive methods used. No attempt was made to further define the complex concept of "quality", and instead the students' personal interpretations of the concept was used. The results are shown in Figure 6.
Figure 6. Students' answers to the question "to what extent do you believe that X affected the quality of your final report in a positive way?" n(group supervision)=35, n(individual supervision)=35, n(company supervision)=28, n(methodology course)=35

Comparing the methods to each other gives that 68% gave the methodology course a positive judgment followed by the group supervision with 61%, the individual supervision at 60% and finally the company supervision at 57%.

3.6.2 Formal content

The question "to what extent do you believe that X contributed to help you with the formal content [like references]", where X was substituted for the four supportive methods used. The results are shown in Figure 7.

Figure 7. Students' answers to the question "to what extent do you believe that X contributed to help you with the formal content [like references]?" n(group supervision)=35, n(individual supervision)=35, n(company supervision)=32, n(methodology course)=35

Comparing the methods to each other gives that 74% gave the methodology course a positive judgment followed by the individual supervision with 37%, the group supervision at 31% and finally the company supervision at 14%.

3.6.3 Formulating research questions

The question "to what extent do you believe that X contributed to help you with formulating your research questions", where X was substituted for the four supportive methods used. The results are shown in Figure 8.
Figure 8. Students' answers to the question "to what extent do you believe that X contributed to help you with formulating your research questions"? n(group supervision)=35, n(individual supervision)=35, n(company supervision)=29, n(methodology course)=35

Comparing the methods to each other gives that 77% gave the company supervision a positive judgment followed by the individual supervision with 57%, the group supervision at 37% and finally the methodology course at 34%.

3.6.4 Choice of research methods

The question "to what extent do you believe that X contributed to help you with choosing your research methods", where X was substituted for the four supportive methods used. The results are shown in Figure 9.

Figure 9. Students' answers to the question "to what extent do you believe that X contributed to help you choosing research methods." n(group supervision)=35, n(individual supervision)=35, n(company supervision)=31, n(methodology course)=35

Comparing the methods to each other gives that 60% gave the individual supervision a positive judgment followed by the company supervision with 52%, the methodology course at 43% and finally the group supervision at 23%.

3.6.5 Realization of the task

The question "to what extent do you believe that X contributed to help you the realization of your task", where X was substituted for the four supportive methods used. The results are shown in Figure 10.
Comparing the methods to each other gives that 77% gave the company supervision a positive judgment followed by the individual supervision with 57%, the group supervision course at 34% and finally the methodology course at 31%.

3.6.6 Structuring the work

The question "to what extent did X fill the function of structuring your work ", where X was substituted for the three supervision methods used. The methodology course was not included since it was the interaction with the group and the supervisers that was examined here. The results are shown in Figure 11.

Comparing the methods to each other gives that 74% gave the group supervision a positive judgment followed by the company supervision with 50% and the individual supervision at 34%.

3.6.7 Whip and carrot

The question "to what extent did X fill the function as whip and carrot to make your work progress?", where X was substituted for the three supervision methods used. Again, the methodology course was not included since it was the interaction with the group and the supervisors that was examined here. The results are shown in Figure 12.
Comparing the methods to each other gives that 77% gave the group supervision a positive judgment followed by the company supervision with 71% and the individual supervision at 51%.

4 Discussion and conclusion.

Returning to the research questions formulated in section 1.6, the first one was to what extent the quality of the theses had improved. According to the examiner, the quality had indeed improved considerably, but that can also have other explanations, such as the change in student groups. However, the improvement was quite general, even among the students selected from the same groups as before. Looking at the students' own opinions on the effects on quality by the different support methods, there was no great difference between the methods. Since the methodology course and group activities were considered more useful than individual and corporate supervision by roughly one third of the students, and equally useful by another third, this indicates the group activities and methodology course indeed had a positive effect on the quality of many of the reports.

As for the lead-time, there was an improvement where the mean lead-time dropped from 8.6 months to 7.3 months. The students' own opinions were also that they spent roughly the nominal 5 months working on their theses.

The students' general opinions of the reform was very positive, where 91% were positive to the group activities and 91% were also positive to the methodology course. A further 76% also believed it was useful to have to read and comment other students' work.

The effects of the reform on the studied factors showed different method had advantages in different areas. As for learning how to write the formal content correctly, the methodology course was clearly best. Formulating the research questions, choice of research methodology and practical realization of the task were still best handled by the individual supervisor and company supervisor, but for structuring the work and acting as whip and carrot, the group supervision was the best activity.

Finally, the question of whether the students preferred face-to-face or virtual seminars a clear majority of 69% preferred the virtual seminars.

The results clearly indicate the reform has been successful. Given the very limited teacher time invested in the group supervision and the methodology course a result where the group supervision and the methodology course even approached the usefulness of the individual supervision would have been a success, but since their usefulness actually exceeded that of the individual supervision in several areas the reform must be considered highly successful.
The choice of a e-learning approach for students used to face-to-face learning was also successful, which makes similar reforms possible in a wide number of different fields.

There is room for improvement however. Some students took the peer assessment more seriously than others and many groups reported difficulties for many members in handing in the assignments on time. This gave cause for discontent, just like predicted by Equity theory. One possible way for approaching this problem could be for a teacher to do summative assessments of the students' formative assessments, and to involve teachers more in the peer assessment activities. In this way, the benefits of formative peer assessment would remain, but possible imbalance caused by different contributions by different peers would be compensated by a reward in the summative assessment.

References:

Author:
Björn, Hedin, M.Sc.
KTH School of Computer Science and Communication, Department of Media Technology and Graphic Arts
Lindstedtsvägen 5, SE-10044, Stockholm, Sweden
bjornh@kth.se