A model for online learner support based on selecting appropriate peer tutors

Open Universiteit

www.ou.nl

A LEARNER SUPPORT MODEL BASED ON PEER TUTOR SELECTION – EXPERIMENTAL RESULTS

Van Rosmalen, P., Sloep, P.B., Brouns, F., Kester, L., Berlanga, A., Bitter, M. and Koper, R. (All authors are employed at the Open University of the Netherlands).

Address for correspondence: Peter van Rosmalen, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands. Email: Peter.vanRosmalen@ou.nl.

Abstract

Tutors have only limited time to support students. In this paper, we discuss a model that addresses the question of how to help students answer content-related questions. A small group of students is created, which consists of the student who asked the question and peers who should be able to answer it. Criteria used to compose the group are the content of the question in relation to the knowledge and skills of the peers. The model supports the collaboration with text fragments selected from the study materials. We will introduce the model and briefly discuss the results of the calibration and a simulation of the model. Finally, we will discuss the outcome of an experiment with two groups of approximately 50 students, who used the model for a period of 8 weeks. The results indicate that the students positively value the model and that it is possible to solve a substantial number of their questions.

Keywords

elearning, Latent Semantic Analysis, lifelong learners, peer support, tutor workload

Introduction

In modern learning settings, students typically spend a significant amount of time learning online. The advent of the knowledge economy and the individualization of our society are two leading factors that underpin their increasing demand for flexibility: students want to be able to study at the place, time and pace of their own choosing (logistic flexibility); also, students are unwilling to submit themselves to pre-planned, rigid programmes, but want their prior competences honoured and their specific study plans catered for (subject matter flexibility). However, as in traditional, on-site settings, students will have questions on where to start, how to proceed, how to understand and apply the available study material, or they will want to have their contributions assessed. In this paper, we will concentrate on one element of this challenge, to wit, answering questions related to the content studied. For a tutor, this is considered a time-consuming and disruptive task (De Vries *et al.*, 2005). Yet, learning may improve if learners can ask questions and receive timely and relevant feedback (Howell, 2003). To address this issue, we have developed a model of how best to organize this with the help of peer-tutoring and we have instantiated this in software so as to be able empirically to test the model.

Our model seeks to solve content-related questions by involving peers (peer tutoring) in answering them. To that end, we identify appropriate and available students as well as documents, and bring these together in a so-called *ad hoc, transient* community. Such a community is *ad hoc* in that its only purpose is to solve a particular question; it is transient in that it vanishes the moment the question has been solved. The model distinguishes (Table 1) six main steps of which the second step depends on a language technology called Latent Semantic Analysis (LSA). In the following section we will introduce the current implementation; next we will quickly go into the results of the calibration and a simulation of a prototype of the model. The main focus of the present article is an exploration of the results of a field experiment with two groups of approximately 50 students, who use the model in the context of a Learning Network on the topic of 'Internet Basics'.

For a proper perspective on the present article, bear in mind that it is part of a series. Kester *et al.* (2007) examine in details the theoretical aspects of learning in communities. Van Rosmalen *et al.* (2006) discuss the use of LSA, Van Rosmalen *et al.* (in press) describe the technological aspects of the model and discuss how to tweak its parameters. Finally, see Koper *et al.* (2005) for the wider context of our model: a Learning Network, i.e. a self-organised, distributed system of lifelong learners, Activity Nodes and competences, designed to facilitate learning. The present article completes the design and development cycle of the model.

Table 1: The main steps of the model.

Pre-condition	A Learning Network with a set of Activity Nodes and a set of users with their profiles indicating their progress with regard to the Activity Nodes
Main steps	1. *Anne* poses a question.
	2. The *system* determines:
	a. the most relevant text fragments;
	b. the appropriate Activity Nodes;
	c. the most suitable students.
	3. The *system* sets up a wiki with the question, the text fragments and guidelines.
	4. The selected *students* receive an invitation to assist.
	5. *Anne and the peer-students* discuss and formulate an answer in the wiki.
	6. If answered (or after a given period of time) *Anne* closes the discussion and rates the answer.
Post-condition	The answer is stored.

Model implementation

A prototypical software application has been developed to test the model. Through the virtual learning environment Moodle (http://www.moodle.org), the students are exposed to a Learning Network, its Activity Nodes and a question module (AskCQ) (Fig. 1) that organises and structures the question answering process.

Fig 1 Part of the interface of the AskCQ module with (1) 'Uw vragen' (your questions) with a link to pose new questions and an overview over the posed questions: 'Status' (Klaar = ready; Bezig = busy), 'Uw vraag' (your question), 'De Antwoordwiki' (a link to the wiki), 'Vraag afronden' (End a question: Afgerond = rated; or a link to the rating); (2) 'Uw antwoorden' (your answers) with an overview of answer given: 'Status' (Klaar = ready; Bezig = busy), 'De vraag' (the question), 'De Antwoord-Wiki' (a link to the wiki).
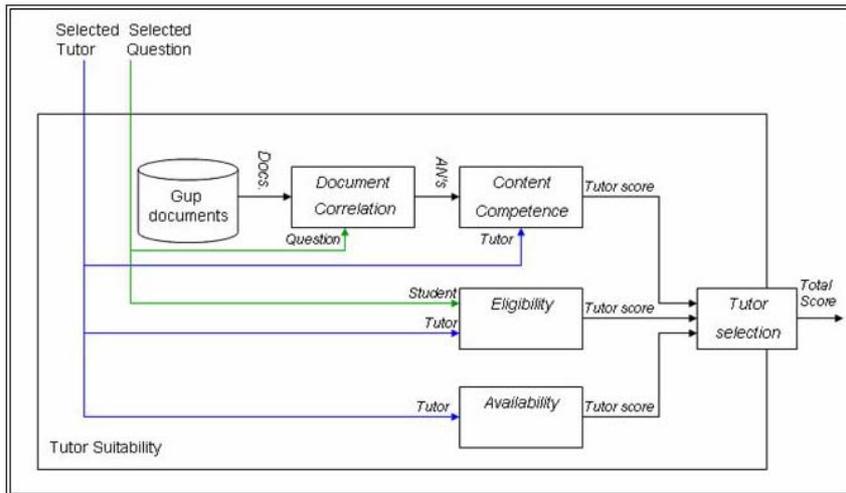
Fig 2 The design interface. 'Gup documents' to build the corpus and to set the LSA parameters; 'Document correlation' to set the document matching method; 'Content competency', 'Eligibility', 'Availability' and 'Tutor selection' to set the peer selection parameters.

The model covers three phases. In the design-phase, its operating context is defined. All texts of the Learning Network are captured and put into a corpus for processing; also, all parameters, the LSA and the peer selection parameters, are set. The designer has a specific interface (Fig. 2) at his disposition to create the corpus and to experiment with and set the LSA and peer selection parameters (see the next phase for details). In principle, the design phase is operative only when launching and updating the Learning Network.

The question-phase starts whenever a student poses a question (e.g. "when I register for a particular chat room, does my registration allow me to use several pseudonyms?"). First, GTP (General Text Parser; Giles *et al.*, 2001), an LSA implementation, maps the question into the collection of text fragments in the corpus. The GTP module returns correlations between the question and text fragments. Text fragments always come from a unique Activity Node. Hence, a high correlation between the question and some text fragment also implies a high correlation between the question and a specific Activity Node. Thus having identified the Activity Nodes that are relevant for the question, ATL (A Tutor Locator; De Jong *et al.*, 2007) selects 2 peer-students most suited to give an answer. The selection is based on their competency on those Activity Nodes that GTP found to correlate highest with the question. This is called the peer's *content competency*. Actually, ATL uses a weighted sum of four criteria. Besides content competency*,* it brings into the mix *tutor competency*, *availability* and *eligibility* (Van Rosmalen *et al.*, in press).

Finally, in the answer phase a wiki is created that includes the question and the three text fragments that in the question phase GTP identified as correlating best with the question (Fig. 3). The wiki is populated with the 2 peer-students who have been invited and have agreed to help. Peers and the question-asking student discuss possible answers and, hopefully, arrive at a satisfactory one. Parenthetically, we chose to work with three documents as a compromise between too few to be helpful and too many to be all read by the supporting peers.

Fig 3 Part of the WIKI: (1) "Vraag" (the question); (2) Three hyperlinks to the selected text fragments 'tekst 1', 'tekst 2' and 'tekst 3' (3) wiki-interface Toon (Show), Bewerk (Edit), Links (Links) and Geschiedenis (History).

Calibration and a first simulation

To ensure the model's viability, we calibrated the LSA-parameters, and simulated and tested two of its key aspects. First, we checked how good LSA was at identifying the topic of a question (i.e. to which Activity Nodes a question belongs) and at selecting text fragments useful for answering the question. Second, we checked if the peer selection criteria met our expectations. The domain of the Learning Network we used is 'Internet Basics', a collection of texts, links and tasks that aim to instigate a basic understanding of the Internet (Janssen *et al.*, 2007). It contains 11 Activity Nodes (content modules), each of which introduces a different aspect of the Internet. The Activity Nodes consist of an introduction, exercises, references to external web pages for further study, and an assessment.

For the simulation, we formulated a set of 16 test questions, each related to exactly one Activity Node. The prototype identified the correct Activity Node for 12 out of the 16 questions (75%). Moreover, two of the authors of the Learning Network in question evaluated the fit of the 3 text fragments proposed by the system. They indicated that for 7 of the questions, one or more of the text fragments were indeed useful in answering those questions. The authors also indicated that 5 of the 16 questions posed were beyond the scope of the contents of the Activity Nodes studied. Taking this into account, the score is 7 questions with useful text fragments out of a total of 11 (about 60%, for details, see Van Rosmalen, 2006).

To test the validity of the peer selection criteria, we created 5 student accounts (Table 2) and assigned a set of test values to the parameters of the peer selection formula. Here we will only illustrate how they worked, for details see Van Rosmalen *et al.,* (in press). By way of test, student L_1 twice 'asked' one of the 16 questions mentioned above. The question was related to Activity Node 2. The first time the student asked the question, the peer-student with the highest rank was selected, as expected. Furthermore, the results of the test showed, that we could balance the selection of peers by taking availability and eligibility into account. For the first question the value of eligibility favoured student L_2 over student L_3, i.e., it prioritized the selection of a peer-student in the same study-phase. (Note: L_2 and L_3 have content competency 1 and availability 0.5. However, only L_2 and L_1 finished Activity Node 1, therefore L_2 has a higher eligibility than L_1). However, if we pose the same question again a compensation mechanism becomes operative due to the decreased availability of Student L_2. (Note: Because of his being chosen to answer the first question the availability of L_2 will become 0).

Table 2. Content competency of student L_1 - L_5 for Activity Node 1 and Activity Node 2, and their availability score.

	L_1	L_2	L_3	L_4	L_5

Content competency Activity Node 1	1 (= successfully completed)	1	0.3	0 (= not started)	0
Content competency Activity Node 2	0.3 (= in progress)	1	1	0	0
Availability (at the start)	0.5 (= moderately available)	0.5	0.5	0.5	0.5

Materials and methods

Having designed and developed a model that fitted our goals and having calibrated it in order for it to function optimally, we proceeded to carry out an experiment. To that end, we made available for 8 weeks a course in the Learning Network on Internet Basics, the same course that was used to calibrate the model. The course is a free (no tuition) course. No credits were given for its completion. The number of study hours is about 22. 111 Students were recruited from both students and staff in our organisation. One of them withdrew after one week. All 110 remaining students were treated as novices when starting the course. For the experiment, the students were divided at random in two groups. Looking at the group characteristics in detail, the groups were very similar with regard to age, gender, computer experience and previous education. Computer experience varied from having little (38) to much (42) and only few students having very little (3) or very much (4) experience (23 non-responders). Males and females were present in almost equal numbers (46 versus 49, 15 non-responders). The modal age was between 46 and 55 (32), trailing off to 25 between the ages of 56 and 65, and 18 between the ages of 36 and 45, with 12 under 35 (5 of which under 25) and 5 over 65 (18 non-responders). Participants educated at the tertiary level (58) vastly outnumbered participants having had up to secondary education (25) or primary education only (3) (24 non-responders or differently educated).

In the experimental group, we used the following weights for the parameters: 1.0 *content competency*, 0.5 *availability* and 0.5 *eligibility*. The model prescribes that the content competency should always be positive. Therefore to avoid a "cold-start" problem, i.e. having no suitable students at the beginning, if necessary the system automatically switched to the settings of the control group. In the control group, we only ensured that the questions would be divided evenly between the students. This meant 1.0 *availability* and all other weights zero (for a detailed description of the parameters and criteria see Van Rosmalen *et al.*, in press). For each question we invited the two students with the highest scores. If within 48 hours no one responded the two next best students were invited.

All students received general instructions on the use of the Learning Network and a specific instruction on how to use the AskCQ-module for all their content-related questions. They received a global explanation of the objectives of the experiment and the suggestion to use the AskCQ-module at least twice, if appropriate. In addition, three times during the 8 weeks the Learning Network was running, a newsletter was distributed with information regarding the course and also a notice on the AskCQ-module. Staff-tutors did not assist answering content-related questions at all. Their role was limited to rating the results of each question-answer pair after completion of the experiment.

Finally, of the total of 110 students, 78 students were active: 40 in the experimental group and 38 in the control group; the remaining 32 students showed no or very limited activity. This means they did not complete any of the 11 Activity Nodes. The average numbers of Activity Nodes completed was for each group 6. The total number of students that successfully completed the course was 25 for the experimental group and 24 for the control group.

Hypotheses

We tested three related hypotheses:
- The model helps to solve a substantial number of the content-related questions posed by students, without invoking any staff support (hypothesis *A*). A substantial number is about 50%, as in our view this is the minimum percentage sufficient to justify the investment needed to develop this

kind of system. In specific cases though, e.g. if discussions between students have a very high priority, the percentage may of course be set lower.

- Ad hoc transient communities for which peers were selected on the basis of a combination of the four criteria proposed (i.e. including the LSA-based content competency) outperform communities whose peers were selected with workload balancing in mind only (hypothesis *B*). This hypothesis concerns the effectiveness of the LSA-based peer selection system.
- Text fragments with which the wikis were seeded helped the peer-tutors to answer the questions asked by their peers (hypothesis *C*). This hypothesis is about the usefulness of seeding the wikis with text fragments identified with LSA.

Hypothesis A is tested directly, simply by counting the number of questions solved successfully. Whether a question is solved successfully, is assessed by the student who posed the question and by two expert tutors. Hypothesis *B* has been tested by administering different peer-student selection approaches to an experimental and a control group of students; group membership was decided by random selection. Peer tutors in the experimental group were selected on the basis of content competency, determined with LSA, availability and eligibility; peer tutors in the control group were selected on the basis of availability only. Dependent variables measured are the number of invitations accepted, the time to answer a question, the number of questions answered, and the quality of the answers given. Hypothesis C has been tested in a limited sense only. The wikis of both the experimental and control group were seeded with text fragments in order to avoid compounding the effect of the way in which the groups have been composed with the availability of text fragments. Therefore, the effect of the availability of text fragments on, say, the quality of the answers given could not be tested. In order still to assess their usefulness we looked at the sources the students used to answer the question (did they use the fragments at all?) and at the extent to which they perceived the text fragments as useful.

Data collection

We collected four types of data. The former two, logging data and student ratings, have been collected during the experiment; the latter two, staff-tutor ratings and evaluation data, have been collected after the experiment:

- Logging data. The progress data of the students, i.e. for each question: question time; invitation-accepted time; number of invitations accepted; answer-accepted time; value of the tutor suitability and the value of the underlying criteria; the main Activity Node of the question.

- Student ratings. For each question, the students that accept the invitation each rate their own peer-tutor suitability; the question poser indicates the main Activity Node of the question and rates the answer received.

- Staff-tutor ratings. At the end of the course, two staff-tutors rated the answers of all closed questions-answer pairs; for these pairs they also specified the Activity Node from which the question derived in their opinion.

- Evaluation data. At the end of the course the students received a questionnaire on the usability aspects of the system. The questionnaire included among other things questions on the usefulness of the supplied text fragments; on how students supplied their answers (prior knowledge, an Activity Node or another source); on whether students perceived other positive outputs such as getting to know each other or understanding the learning material better; and on whether students appreciated the overall approach and use of the system.

Results

Hypothesis A: Questions solved

During the 8 weeks of the course, a total of 101 questions were posed, 59 in the experimental group and 42 in the control group (see Table 4, rows 1a-1c for more details). According to the question

posers, the number of successfully answered questions was 42 (71%) for the experimental group and 19 (45%) for the control group (Table 3, row 1). However, these figures do not suffice to conclude that hypothesis A has been confirmed.

First of all, the student judgments should be in line with the staff-tutors' rating. After all, the students could have been satisfied too easily. The two staff-tutors rated all questions, including the questions started but not yet rated by the students (Table 3, rows 2-4). The overall agreement between the tutors on solved versus not-solved questions (i.e. "solved" = rating 4 and 5 and "not solved" = rating 1 and 2) is high: 83% (62 out of 75) or 73% (64 out of 88 if we also include the rating 3 "solved/not solved"). If we combine the judgment of the students and the tutors, by counting a question as solved if at least two of the three ratings are 4 or above, the number of questions solved is approximately the same as the number indicated by the students (Table 3, row 5). So student opinion do not differ much from expert (staff) opinions.

Second, irrespective of an overall agreement between students and staff, there should only be very few 'false-positives'. A false-positive is an answer that according to the student is right but actually is wrong. Too many false-positives are a threat to the quality of education. Based on the ratings, we identified 8 questions that required further analysis. Careful reading of each of the questions showed that none of the answers was a genuine false-positive:

- 5 questions were irrelevant and/or closed by the question poser before the peers could help.
- 2 answers received a rating of 3 by both staff-tutors. In both cases, however, one could easily argue that the question was fairly well answered.
- Finally, one remaining question was not articulated well, making it difficult to judge whether the answer was adequate or not.

Table 3. Question-answer details.

		Experimental	Rated	Control	Rated
1	Questions solved: Question poser	42 (42/59=71%)	53	19 (19/42=45%)	29
2	Questions solved: Tutor 1	34	53	16	29
3	Questions solved: Tutor 2	38	53	17	29
4	Question* not closed but solved according to Tutor 1 and 2	2	4	4	6
5	Questions solved (integrated score) at least 2 agree	44 (44/59=75%)	57	22 ((22/42=52%)	35

Note: 6 of the 10 questions started, but not rated, did contain an answer, 4 were not really started.

Hypothesis B: The model treatment

For hypothesis B we looked if the experimental group outperformed the control group on responsiveness (the number of invitations required and the answer time) and quality (the number of solved questions, the level of the answer ratings) of the answers.

Before going into details, a first inspection suggests that the students in the experimental group participated more actively in question-answering. They posed more questions (though not significantly), solved more questions (see above), answered faster, and fewer invitations were required (Table 4, rows 1, 6 and 2 respectively). Their answers were also rated higher (Table 4, row 7). There was only one apparent anomaly, the control group (Table 4, rows 4-5) had a higher overall involvement. This, however, actually is to be expected as in the control group the algorithm attempts to achieve an optimal spread of the workload.

Table 4. Overview of the results.

		Experimental	Control	Total
1	The number of questions posed*	59	42	101
1a	- closed (so rated);	53	29	82
1b	- in discussion	4	6	10
1c	- failed i.e. the invited peer-tutors did not react to or refused the invitation	2	7	9
2	The number of questions with 2 invitations loop	12	21	33
3	The number of students that posed one or more questions	26	21	47
4	The number of students that assisted in answering one or more questions	29	36	65
5	The total number of students actively involved (posing or answering)	31	37	68
6	Average time to resolve a question (days)	5.6	9.6	7.0
7	Average answer rating of the rated questions (5-points scale, 1 is not answered - 5 is fully answered) by the question poser	4.0	3.4	3.8

* For 4 out of the 9 questions not started, invitations were sent out on the very last day that the course was available.

The following tables show the results of a detailed analysis on responsiveness, i.e. the number of invitations required (derived from Table 4, row 2) and the answer time (Table 4, row 6). For both the number of invitations required (Table 5) and the time to answer (Table 6) the experimental group scores significantly better.

Table 5. Number of invited students per question: 2 implies one invitation loop, 4 two loops
The chi-square = 9.81; df = 1; asymp sig = 0.002.

		Group		Total
		Experimental	Control	
Number of invited students	2	47 (79.7%)	21 (50.0%)	68 (67.3%)
	4	12 (20.3%)	21 (50.0%)	33 (32.7%)
Total		59 (100.0%)	42 (100.0%)	101 (100.0%)

Table 6. Answer time (in hours). Mann-Whitney U = 436.500; asymp. sig. = 0.001

	Group	N	Mean Rank	Sum of Ranks
Answer time in hours	Experimental	53	35.24	1867.50
	Control	29	52.95	1535.50
	Total	82		

The picture is confirmed when looking in detail at the quality, i.e. the number of question solved (Table 7). We already noted that the experimental group solved more questions, further analysis shows that difference to be significant. Moreover, when looking in detail at the rating of the questions, the two groups not just differ with respect to questions solved and not-solved (Table 7). The experimental group, also proves to be very clear about its judgement (Table 8). A large part of their ratings (60%) is in the segment of absolutely reject (rating 1) and absolutely accept (rating 5).

Table 7. Solved Questions. The chi-square = 6.90; df = 1; asymp sig = 0.009.

		Group		Total
		Experimental	Control	
Solved Questions	Not solved	17 (28.8%)	23 (54.8%)	40 (39.6%)
	Solved	42 (71.2%)	19 (45.2%)	61 (60.4%)
Total		59 (100%)	42 (100%)	101 (100%)

Table 8. Ratings of the answers

		Group		Total
		Experimental	Control	
Rating Question Poser	1	6 (11.3%)	6 (20.7%)	12 (14.6%)
	2	5 (9.4%)	4 (13.8%)	9 (11.0%)
	4	16 (30.2%)	10 (34.5%)	26 (31.7%)
	5	26 (49.1%)	9 (31.0%)	35 (42.7%)
Total		53 (100.0%)	29 (100.0%)	82 (100.0%)

So there is strong evidence that the experimental group outperforms the control group. However, it still needs to be shown that the selection algorithm chooses its peer tutors for the reasons we assume. A critical condition for this to be the case is that the question posed should be mapped to the relevant Activity Node (as the result of this mapping is used to determine the suitable peer-students). The staff-tutors were asked to indicate the main Activity Node for 71 questions (9 of the questions were not rated at all, because nobody had yet responded to them, 20 questions were rated as being out of scope with regard to the content of the Learning Network). The results for both tutors were almost identical. They differed in opinion only on 4 questions. In 62% of the cases the judgment of the tutors was identical to the one calculated by the system. This result imparts confidence on the algorithm, particularly if one takes into account that in practice it points to more than one Activity Node.

Hypothesis C: Supporting the student in answering

To establish to what extent the text fragments are of use in answering the questions, we collected the following results through the questionnaire. First of all, the majority of the respondents (34 of 50) valued the text fragments with a rating of 4 or above (Table 10). The resources used to answer are given in Table 9.

Table 9. The resources used to answer a question, the respondents could select one or more choice (n = 51; 1 skipped this question, 5 were not involved as peer-tutor).

		Experimental	Control	Total
Resources used	Text fragments	5	9	14
	Course content	14	11	25
	Prior-knowledge	17	12	29
	Others (unspecified)	11	9	20

Table 10. The appreciation of the text fragments (n = 50; 2 skipped this question, 5 were not involved as peer-tutor)

		Experimental	Control	Total
Text fragment rating (1=useless ; 5=good)	2	1	4	5
	3	10	1	11
	4	11	12	23
	5	3	8	11
Total		25	25	50

It is interesting to notice that the control group tends to appreciate the text fragments more; also they mention more often that they actually used the text fragments (both differences not significant). An explanation could be that the control group is more dependent on the text fragments since they have been selected at random. Also there is the in-built tension that a good selection strategy likely diminishes the need for text fragments.

General observations

Finally, through the questionnaire we received feedback on the usability and general acceptance of the model. The questionnaire was completed by 57 of the 110 students (52%) fairly evenly divided over the two groups, 29 (experimental) and 28 (control) respondents respectively. For both groups, almost all respondents agreed that answering a question is a good investment of time (25 experimental; 22 control). Positive responders could motivate their answer by indicating one or more reasons on a list. Two reasons were selected most often:

- "I am aware that other students also have questions" (24 students), and
- "It improved my knowledge and understanding" (29 students).

The overall usefulness (26 experimental; 17 control) and usability (22 experimental; 16 control) received a positive rating i.e. 4 or above on a 5-point scale. The figures, however, show that the respondents of the experimental group are more positive. This was confirmed when asked who would like to see this question-answering approach offered in other courses too. The students in the experimental group were significantly more interested, i.e. 25 students of the experimental group answered positive against 16 of the control group (chi-square = 5.177; df = 1; asymp sig = 0.023).

Discussion and conclusions

The tests of our three hypotheses lead to the conclusion that our system can successfully be used to help answering content questions students may have. For the experimental group the number of questions solved is clearly above 50% and there were no false-positives. The number of solved questions could have even been higher, had we not abruptly stopped the experiment after 8 weeks. We also unambiguously showed that the experimental group outperforms the control group both with regard to responsiveness and quality. Only the result for the last hypothesis, about the usefulness of the text fragments, is less clear. Though students seem to appreciate the text fragments, they make only limited use of them.

Despite these promising results, a number of limitations to the experiment have to be considered. The experiment ran with a fixed group and for a fixed period. This situation is different from our target situation, with an 'unending' course and lifelong learners starting and finishing at any time. Another concern is the limited complexity of the contents studied, which were typically at the beginner level. This was confirmed indirectly as a substantial part of the responders indicated that they used prior knowledge to answer the questions. Results may be different with more demanding topics, the questions may then be too difficult to answer. Finally, students were aware that they participated in an experiment. This likely will have boosted their responsiveness. In the reality of a long-lasting Learning Network with many activity nodes, policies to ensure sufficient participation may be required (Berlanga *et al*., in press).

Also the particular implementation of the model deserves further attention. The wiki proved to be a tool that was unknown to the students. They used it much in the same way as a forum, contributing in turn and not editing the texts of others, while we had hoped them to become involved in a collaborative writing process. The students also 'complained' that they were not properly informed once a question was resolved and, even more important, that it was not possible to continue a contact through the system. Both we consider very serious issues because an important additional objective of this system is to assist lifelong learners in becoming (self-)organised into communities (Kester *et al*., 2007). Therefore it is important that contacts are well-established and can be followed up if desired. Finally, students mentioned they would have liked to be able to study all question-answer pairs, not only the ones they contributed to. Whether or not this is a good idea is not obvious. It may improve the question-answering efficiency but it will likely lessen the need for and the benefits of a discussion between students.

Further research is needed to address the issues raised and also to get a better insight into the effects of different values for the parameter settings of the model. So far, however, we believe that we showed that the system developed offers a promising line for efficient and effective support for elearning in general and lifelong learning in particular.

Acknowledgements

References

Berlanga, A., Sloep, P.B., Kester, L. Brouns, F., Van Rosmalen, P., Koper, R. (in press). *Ad Hoc Transient Communities: Towards Fostering Knowledge Sharing in Learning Networks*. International Journal Learning Technologies.

De Jong, A., Brouwers, M., Brouns, F., Van Rosmalen, P., Kester, L.,Sloep, P.B. (2007). *ATL: A Tutor Locator*. URL: http://hdl.handle.net/1820/960.

De Vries, F., Sloep, P., Kester, L., Van Rosmalen, P., Brouns, F., de Croock, M., Pannekeet, C. & Koper, R. (2005). *Identification of critical time-consuming student support activities that can be alleviated by technologies*. Research in Learning Technology (ALT-J), 13 (3), 219-229.

Howell, K. (2003). *Question Generation and Answering Systems R&D for Technology-Enabled Learning Systems Research Roadmap*, retrieved November 2004 from http://www.thelearningfederation.org/qa.pdf.

Janssen, J., Tattersall, C., Waterink, W.,Van den Berg, B., Van Es, R., Bolman, C., & Koper, R. (2007). *Self-organising navigational support in lifelong learning: how predecessors can lead the way. Computers & Education,* 49(3), 781-793.

Giles, J. T., Wo, L., & Berry, M. W. (2001). *GTP (General Text Parser) Software for Text Mining*. Retrieved online January, 2005 at: http://www.cs.utk.edu/~berry/papers02/GTPchap.pdf

Kester, L., Sloep, P. B., Van Rosmalen, P., Brouns, F., Koné, M., Koper, R. (2007). *Facilitating Community Building in Learning Networks Through Peer-Tutoring in Ad Hoc Transient Communities*. International Journal on Web Based Communities. Vol. 3 (2), 198-205.

Van Rosmalen, P., Sloep, P., Brouns, F., Kester, L., Koné, M. & Koper, R. (2006). *Knowledge matchmaking in Learning Networks: Alleviating the tutor load by mutually connecting learning network users*. British Journal of Educational Technology Vol. 37 (6), 881-895.

Van Rosmalen, P., Sloep, P., Kester, L., Brouns, F., De Croock, M., Pannekeet, K. & Koper, R. (in press). *A learner support model based on peer tutor selection*. Journal of Computer Assisted Learning. Available online at: doi: 10.1111/j.1365-2729.2007.00245.x.