Instructional Science **29:** 33–44, 2001.

Effects of problem-based discussion on studying a subsequent text: A randomized trial among first year medical students

WILLEM S. DE GRAVE[1], HENK G. SCHMIDT[2] & HENNY P.A. BOSHUIZEN[1]
[1]*Department of Educational Development and Research;* [2]*Department of Cognitive Psychology; PO BOX 616, 6200 MD Maastricht University, The Netherlands*
(Phone: +31-43-3881117; Fax: +31-43-3884140; e-mail: degrave@educ.unimaas.nl)

Received: 18 October 1999; in final form: 31 March 2000; accepted: 21 July 2000

Abstract. The Purpose of this study was to examine effects of group discussion of a medical problem on the comprehension of a subsequent problem-relevant text by first year medical students.

Forty-eight first-year medical students were randomly assigned to one of two conditions: The experimental group discussed a problem of blood pressure regulation, where the control group discussed a problem of vision. Subsequently, all students studied a text on the physiology of blood pressure regulation. Finally, a free recall test was administered. Numbers of propositions accurately recalled were analyzed using analysis of variance (ANOVA). Students who discussed the blood pressure regulation problem recalled 25% more from the text than those who discussed the control problem. This difference was statistically significant. The present study represented the first truly randomized trial in the ecologically valid context of a medical curriculum. It demonstrated the positive effects of problem-based tutorial group discussion on the comprehension of text. It confirmed earlier findings from laboratory studies that problem-based learning, in addition to positive emotional and motivational long-term effects well-documented in the literature, also has positive effects on learning.

Keywords: problem-based learning, randomized trial, knowledge construction, learning from texts, elaboration, cooperative learning

Introduction

What do students learn from a problem-based learning exercise? That means, what and how much knowledge do students acquire? People who ask this question have frequently been trained in discipline-based education. PBL, however is an approach that focuses on problem analysis, self-directed learning and knowledge application (Schmidt, 1993). This question is among the first asked by those who are confronted with the problem-based approach to medical education. Although the question is simple, it turned out that the answer is remarkably difficult to be given. A number of studies (e.g. Eisenstaedt et al., 1990; Albanese & Mitchell, 1993) have attempted to provide answers with mixed outcomes. Some studies suggest that problem-based learning leads to superior performance on tests of knowledge, whereas others demonstrate the opposite. Again other studies show no differences between problem-based and more conventional approaches to medical education. The reasons for these mixed results are not clear. A recent meta-analysis of studies presented at the 1998 meeting of the American Educational Research Association (Kalaian et al., 1998) suggests that zero or negative results may be due to the fact that many of these studies do not involve randomized groups of students.

The study presented in this contribution article aimed at testing effects of problem-based tutorial group discussion, prior to the individual study, on the learning retention of new information through using an experimental design in which students were randomly assigned to the conditions of the experiment, thereby avoiding biased results due to systematic differences between the groups studied. First however, we will briefly review the literature on effects of small-group instruction on cognition.

Problem-based discussion in a small tutorial group can be considered an educational procedure, aimed at increasing the interaction between knowledge already available in the participants and the new information to be learnt. It is supposed that this procedure promotes the integration of new information into the knowledge base already present (Schmidt, 1993). This integration is stimulated by elaboration, e.g. by constructing explanations for the problem at hand, especially when these explanations are self-generated and stimulated by the interaction in a tutorial group (Chi et al., 1989; Pressley et al., 1992). In the literature, several possible reasons can be found for the positive effects of elaboration (Hamilton, 1989; Mayer, 1996; Prawat, 1989). Hamilton (1989) summarizes these as follows: “Elaboration’s may increase the redundancy of stored information, impose an organizational scheme on stored information, increase the number of contextual elements that will overlap between the encoding context and the retrieval context, and/or increase the distinctiveness of stored information.”

Early experiments on effects of problem-based learning by Schmidt and his collaborators (Schmidt, 1982; Schmidt et al., 1989) provided support for some of these hypotheses. These studies however shared two shortcomings: (1) the use of problems with limited contextual validity for medical education, and (2) involvement of students who had no prior experience with problem-based learning. The experimental problem used in all these studies was the so-called ‘blood cell’ problem, a description of the behavior of a red blood cell under a microscope in water and in a salt solution. Although this problem refers to the biological processes of osmosis and diffusion, its complexity does not match the complexity of clinical problems routinely used in actual problem-based medical curricula. Since proponents of problem-based learning assume that its effectiveness is at least partially dependent on the use of contextually meaningful materials, the blood cell problem experiments cannot be considered entirely ecologically valid.

The second shortcoming of these experiments was that, for most of the students involved, problem-based learning was a new and unfamiliar educational method. None of these students had been trained in carrying out problem-based discussions. A survey of research of co-operative learning (Webb & Sullivan-Palincsar, 1996) shows that students who had been trained in co-operative learning show better interaction skills and greater learning progress. (These training activities involved practicing learning skills such as effectively explaining ideas to one another and asking questions.) A study by Bielaczyc, Pirolli and Brown (1995) illustrates the importance of training study skills such as individually generating explanations and self-regulation strategies for learning to solve problems. They investigated the influence of training in these areas on the way in which students give explanations and the effect of this on the problem-solving test. King (1992) trained students in asking questions to be used in a group setting. Her study also showed that providing training in this area was more effective than not doing so.

The aim of the present study was to find out whether effects of problem-based discussion on student learning observed in laboratory experiments can also be observed in a real problem-based medical curriculum, with problems actually used in that curriculum and with students who had prior experience with problem-based learning. Prior to studying a physiology text, 48 first-year medical students analyzed either a problem relevant to the text, or a control problem. Subsequently, they took a free recall test, measuring what had been learned from the text. The students' recall protocols were parsed into propositions. These propositions – or idea units – were either categorized as *explanatory* or *descriptive*. An explanatory proposition is defined as a statement that either characterizes a process or describes the conditions under which this process occurs. All other propositions were considered descriptive. An example of an explanatory proposition is: "Histamine causes vasodilatation." An example of a descriptive proposition is: "The vascular system is subdivided in three compartments." This distinction was made because problem-based discussion is supposed to have an influence on the integration of new information in the existing knowledge base. Because problem-based discussion leads to elaboration, this it is supposed to improve an individual's organization and awareness of knowledge (Prawat, 1989). An important characteristic of integrated knowledge is its organization. Such better-organized knowledge will be more easily accessible and more useful, and will hence lead to better recall of the text studied. The number of explanatory propositions remembered constitutes a measure for the degree of integration of the new information in the existing knowledge base. Mayer (1989) holds that the number of explanatory propositions in free recall constitutes an accurate measure of the level of integration of subject matter in available knowledge structures.

Mayer also assumes that instructional procedures such as problem-based discussion will decrease the recall of literal information from text and increase the occurrence of *inferences*. This is based on the assumption that problem-based discussion helps students reorganize the subject matter to be learned so as to match their conceptual model; when students actively reorganize the material, they tend to forget the literal content, which leads to an increase of the number of inferences during recall. Hence, another measure that can be used to determine the degree of integration of new information in existing knowledge is the number of inferences during recall. The occurrence of inferences in recall illustrates the constructive character of studying a text. Problem-based discussion within a group is supposed to promote inferences when a text is being studied (Schmidt, 1983). This leads to a situation in which it is no longer possible to clearly differentiate between previous knowledge and the contents of the text. As a result, there will be more inferences during recall.

Method

Participants

Participants were 48 first-year students of medicine (27 female, 21 male), who had been enrolled into Maastricht University, The Netherlands, 18 weeks before the experiment was conducted. They had entered university straight from secondary school and everyone had taken biology and physics at school. During the first unit of the study of medicine, which takes six weeks, they had received an elaborate introduction in the method of working in tutorial groups. Also during the subsequent units, explicit attention was given to the working method. This means that the students were taught from the beginning of their studies how to optimally carry out problem-based learning in tutorial groups. The students had acquired a basic knowledge of blood circulation in secondary education. During the first three six-week units of the problem-based medical curriculum, they had acquired relevant prerequisite knowledge, in particular knowledge regarding the nervous system. The participants were randomly allocated to the conditions of the experiment, which resulted in the creation of 9 groups: 5 experimental groups and 4 control groups.

Materials

The materials used consisted of two problem descriptions (‘Stinging incident’ and ‘Looking too far’), a text and a free recall test. The problem descriptions for the experimental group consisted of a problem (‘Stinging incident’) that is actually used in the curriculum, i.e. the second problem of the fifth unit in first year. The title of the latter unit is ‘Regulation disorders.’ The problem is part of the first sub-theme of this unit dealing with blood circulation. The problem description for the control group (‘Looking too far’) consisted of a task was taken from the first unit of the second year, which also aimed at acquiring basic knowledge. The title of the latter unit is ‘Consciousness, senses and emotions.’ Criteria to chose this problem was that it was not related to the topic to which the experimental group’s problem referred, i.e. the regulation of blood pressure, but appealed in a similar way to the previous knowledge of the students.

The explanatory expository text provided for individual study described the process of blood pressure regulation. The text had the following basic structure: (1) Introduction, including the main factors affecting blood pressure regulation; (2) pressure and flow in the vascular system; hemodynamics, (3) pressure drop in the vascular system; three compartments, (4) regulation of heart performance, (5) short-term regulation of blood pressure, and (6) shock. The text was largely a selection of sections from a well-known physiology textbook (Bernards & Bouman, 1988). This textbook is also referenced to in the unit from which the problem for the experimental group originated from, and is also often used by students. The introduction made use of a section from a quality English textbook on the physiology of the heart and the vascular system (Berne & Levy, 1992). The part on shock was taken from a section from the textbook of Physiology (Guyton, 1991). The structure of the explanatory text was similar to those used in other textbooks. In total, the text concerned was 10 pages long. It consisted of a total of 347 propositions and 2556 words.

The free recall test consisted of three blank pages and a front page with the following instruction: “Write down anything you can remember of the text on blood pressure regulation.”

Procedure

The participants were allocated randomly to one of the two groups of the experiment, i.e. to the experimental group in which a relevant problem was discussed ($N = 27$), or to the control group ($N = 21$). Equivalency of group members was assessed by reference to prior national test scores. The average final score on a national test for biology in the experimental group was 7.11, with a standard deviation of 0.75. The control group showed an average final score of 7.00, with a standard deviation of 0.77. The average final score for physics in the experimental group was 6.70, with a standard deviation of 0.91, while the control group had an average final score of 6.74 and a standard deviation of 0.93. These differences were not significant.

The experimental group of 27 participants was randomly divided into five tutorial groups consisting of either five or six students. The control group of 21 persons was randomly divided into subgroups of five or six students. Each group had a tutor that was also randomly assigned. These tutors were fourth-year medical students, who had worked on the same problems in the past. They were at the same time experiment-leaders. These students had more than three and a half years' experience with problem-based learning. They were all blind to the purpose of the experiment.

The experiment took place in rooms that are normally also used for study group meetings. Overall, use was made of nine rooms. These rooms were equipped with a blackboard and a flip-chart. The experiment was held on a morning when no other study activities were planned. In the experimental condition, the 'Stinging incident' problem was analyzed. The control groups analyzed the problem of 'Looking too far.' Otherwise, the participants under both conditions carried out the same activities.

The tutor explained to the participants what was expected of them. The problem was presented with a brief explanation of its context. In the case of the 'Stinging incident' problem, it was said that the topic concerned circulation. In the case of 'Looking too far,' it was said that this was part of a unit on consciousness, senses and emotions. Having read the problem, students first briefly defined the problem (5 minutes), which was followed by a 30-minute brainstorming session in which they generated possible explanations for the problem. The tutor acted as chairperson and regularly presented a summary of the different ideas.[1] He/she also stimulated the production of statements and making notes of the statements on the flip-chart and/or blackboard. The tutor had previously been given detailed instructions about how to act in this respect. He or she had been told not to give any clues on the basis of which the group could draw any conclusions as to whether they were right or wrong, or whether the statements given were relevant. Using a tape-recording of the discussions, it could be verified that the tutor had not given any information that could be used by the participants to draw conclusions regarding the underlying physiological mechanisms of the two problems.

After the problem-based discussion, the participants in both environments received the expository text on blood pressure regulation. In all groups, the flip-charts with the minutes of the problem-based discussion were first removed. The time allowed to study this text was 30 minutes. Students were permitted to make notes in the text. After 30 minutes, the expository text was collected. Subsequently, the free recall test on blood pressure regulation was handed out to both the experimental and the control groups. No time limit was set on the completion of the free recall test.

Analysis

First, the free recall protocols produced by the participants were divided by two independent assessors into subject-predicate units (or propositions), each represented representing a single idea. The inter-assessor agreement for this task was 96%. The boundaries of two adjacent propositions can be found by identifying appropriate linguistic markers in the text: conjunctions, adverbs, relative pronouns, semicolons, and full stops. The 47 protocols were divided between two assessors, who then evaluated the propositions produced for their accuracy. Subsequently, two different assessors reassessed the accuracy of the propositions of all 47 protocols. The inter-assessor agreement between the first and the second assessment was 87%. Then the propositions were classified by two assessors as being explanatory or descriptive. An explanatory proposition was defined as a statement that either characterizes a process or describes the conditions under which this process occurs. All other propositions were considered descriptive. An independent assessor evaluated a sample of 15 protocols as to the distinction between explanatory and descriptive propositions. The agreement between the first and the second assessment on this classification of propositions was 86%. Differences of opinion between assessors were solved in a consensus procedure. The propositions were also matched with the original text. It was determined whether the proposition was a literal reproduction or a paraphrase of the text, or concerned an inference. This means that information in the protocols extends beyond the information in the expository text. On the basis of matching of protocols and expository text, the number of produced inferences was determined. An independent assessor evaluated a sample of 15 protocols as to this match between protocols and the expository text. The agreement between the first and the second assessment of propositions was 83%. Differences between assessors were solved by means of discussion.

Table 1. Averages and standard deviations for the number of accurate explanatory and descriptive propositions, produced in recall

	M	SD	N
Stinging incident problem (total)	50.1	9.7	27
Explanatory propositions	25.0	6.1	
Descriptive propositions	25.1	6.9	
Control problem (total)	40.5	11.8	21
Explanatory propositions	20.1	7.3	
Descriptive propositions	20.3	5.6	

Results

Free recall

The analysis of variance of the free recall produced the following results: a significant difference in the number of accurately remembered propositions was found between groups analyzing the ‘Stinging incident’ problem and the control condition: $F(1,46) = 9.68$, $MS_e = 56.53$, $p < 0.01$. Hence, the group that analyzed the ‘Stinging incident’ problem produced significantly more accurate propositions than the control group. This confirms the hypothesis that problem-based discussion prior to the expository text to be learnt results in more information being remembered from text. There was no significant difference in the number of inaccurate propositions recalled ($F(1,46) = 0.01$, $MS_e = 12.90$, $p < 0.92$).

The analyses also show that the number of explanatory and descriptive propositions in the recall did not differ significantly, $F(1.46) = 0.03$, $MS_e = 29.28$, $p < 0.86$. Neither is there any significant interaction between treatment and type of proposition, $F(1.46) = 0.01$, $MS_e = 29.28$, $p < 0.98$. Table 1 provides a summary of the results.

The hypothesis was that the group dealing with the ‘Stinging incident’ problem would show a relatively greater number of explanatory than descriptive propositions during recall in comparison with the control group. This was assumed on the basis of the supposed degree of integration of previous knowledge and newly learnt learned information. This hypothesis is not confirmed by the data. The average number of explanatory propositions accurately produced by the experimental group, however, differs significantly from the average number of accurate declarative propositions produced by the control group $F(1,46) = 6.24$, $MS_e = 44.44$, $p < 0.05$. The assumption that the number of descriptive propositions during recall of the ‘Stinging incident’ problem group would be less in comparison with the control group, must be rejected. The number of descriptive propositions during recall of the ‘Stinging incident’ problem group also turns out to be significantly higher than of the control group $F(1,46) = 6.54$, $MS_e = 41.53$, $p < 0.05$. Apparently, both kinds of propositions are remembered better due to problem analysis.

It was also expected that the number of inferences during recall would increase under the influence of elaboration on the analysis of a relevant problem because this would lead to the integration of previous knowledge and newly learnt learned knowledge. Analysis of variance, however, yielded no significant differences between the two groups with respect to the number of inferences during recall $F(1,46) = 0.914$, $MS_e = 34.56$, $p < 0.34$. The averages and standard deviations for the number of inferences produced in free recall for the two groups are: Stinging incident problem (mean 9.56; standard deviation 6.19) and Control problem (mean 11.19 and standard deviation 5.45). Apparently, problem-based discussion does not result in a greater number of inferences during recall immediately after studying the information in the expository text.

Discussion

The results from this experiment clearly demonstrate that problem-based discussion in a small group, prior to studying a relevant expository text, facilitates the learning of such a text. They also demonstrate that finding effects of problem-based learning are not confined to laboratory experiments using relatively artificial materials, but can be found within working curricula as well, using actual curricular materials. The results are generally in line with findings of Schmidt and his co-workers (1992, 1989). They are consistent with an elaboration hypothesis of cognitive effects of problem-based learning as far as the analysis phase concerned. This hypothesis suggests that problem-based discussion encourages elaboration based on prior knowledge, thereby causing greater integration of new information into existing knowledge, and, hence, greater accessibility of such knowledge. The recall patterns found, however, are different from what was expected. No interaction occurred between treatment and type of propositions. This finding corresponds with observations made by Schmidt (1982), that problem-based discussion has a general facilitating effect, rather than a facilitative effect on explanatory knowledge acquisition alone. Also, the absence of differences between experimental and control group with regard to the number of inferences produced is intriguing. On the basis of studying the origin of inferences, Van den Broek et al. (1993) presented a number of explanations for the occurrence or absence of inferences during recall. They concluded that length of the text, time available for studying the text, characteristics of the criterion test, and time of testing affect the extent of inferential processes. Inferential reconstruction becomes stronger the longer the interval between learning and memory assignment. In the present experiment, participants were asked to remember the information from the text immediately after learning. Inferential reconstruction, therefore, may have occurred only to a limited extent; preventing potential differences between both conditions of the experiment to show.

In experiments such as the one reported there are always a number of competing hypotheses that have to be considered. The most important one is that students who discussed the 'Stinging incident' problem may have become more interested in acquiring further information about the physiology of blood pressure regulation, and, hence, may have studied harder. This motivation hypothesis has been tested in studies by De Volder and his associates (1989). These studies show that indeed discussing a relevant problem has an effect on intrinsic interest of students. However, this effect is counteracted by the fact that students, who have previously elaborated upon a certain topic, tend to spend less time on reading a subsequent text. A second alternative hypothesis is that students who discussed the 'Stinging incident' problem may have selectively attended to information that was directly relevant to the problem at hand and, therefore, may have processed that information more extensively. However, selective attention for a part of a text goes hand in hand with spending less time and energy on other parts of the text (Reynolds, 1992). A selective attention mechanism active under the experimental condition, therefore, cannot explain why the subjects in that condition recall more information in general.

In summary, our findings suggest that it is the opportunity for problem-based elaboration on existing knowledge, prior to studying a relevant text, that is responsible for the strong effect demonstrated in the present study. It seems that problem-based discussion encourages better integration of new information in the existing knowledge base, making the resulting knowledge more accessible and remembered better.

This study has several characteristics that restrict the generalizability of the results. This study used first year medical students in a European model straight out of high school. These students differ from students from other medical schools, e.g. like those in North America, where students are older and more educated when they enroll in medical school. Furthermore this study has been done in only one medical school. Other PBL schools have different curricula and different implementations of PBL, which might lead to different study approaches by the students. Finally, the results can be an underestimation of the true learning effects of the problem-based learning exercise since we measure student's new knowledge directly after individual study. Students did not get the opportunity to report and discuss and hence to validate and further elaborate the newly acquired knowledge in the tutorial group. Since effects of problem-based learning seem to be robust at least at the level of individual problems, it is important to demonstrate effects at the course level. Although random assignment of subjects to conditions – in our view a prerequisite for a careful study of effects of the innovation – may be impossible at that level, the present study and those of others (Kalaian et al., 1998) call for careful control of extraneous variables that may moderate or even diffuse effects of problem-based learning. There are some indications that in particular the study of long-term effects may be interesting to pursue (Eisenstaedt et al., 1990; Martensen et al., 1985; Tans et al.,1986).

Note

1. Normally, in the Maastricht curriculum, one of the students themselves would act as chair. However, to minimize random noise, it was decided that the student-tutor would perform this role.

References

Albanese, M.A. & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine* 68(1): 52–81.

Bernards, J.A. & Bouman, L.N. (1988). *Fysiologie van de mens* (5 edn.). Houten: Bohn Stafleu Van Loghum.

Berne, R.M. & Levy, M.N. (1992). *Cardiovascular Physiology* (6 edn.). St Louis: Mosby.

Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction* 13(2): 221–252.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science* 13: 145–182.

De Volder, M.L., Schmidt, H,G., De Grave, W.S. & Moust, J.H.C. (1989). Motivation and achievement in cooperative learning: the role of prior knowledge. In J.H.C. van der Berchen & F. Halisch, eds, *Achievement and Task motivation*. Berwyn: Swets North America.

Eisenstaedt, R.S., Barry, W.E. & Glanz, K. (1990). Problem based learning: Cognitive retention and cohort traits of randomly selected participants and decliners. *Academic Medicine* 65 (Supp): S11.

Guyton, A.C. (1991). *Medical Physiology* (8 edn.). Philadelphia: Saunders.

Hamilton, R. (1989). The effects of learner-generated elaborations on concept learning from prose. *Journal of Experimental Education* 205–216.

Kalaian, H.A., Mullan, P.B. & Kasim, R.A. (1998). Hierarchical linear modeling of the effects of problem-based learning on medical students' performance on NBME I and NBME

II: A meta-analytic perspective. *Paper presented at the Annual Meeting of the American Educational Research Association*. San Diego, CA, April 13–17.

King A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist* 27(1): 111–126.

Martensen, D., Eriksson, H. & Ingelman-Sundberg, M. (1985). Medical chemistry: Evaluation of active and problem-oriented teaching methods. *Medical Education* 19: 34–42.

Mayer, R.E. (1989). Models for understanding. *Review of Educational Research* 59(1): 43–64.

Mayer, R.E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review* 8(4): 357–371.

Prawat, R. (1989). Promoting access to knowledge, strategy and disposition in students: A research synthesis. *Review of Educational Research* 59(1): 1–41.

Pressley, M., Wood, E., Woloshyn, V.E., Martion, V., King, A. & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist* 27(1): 91–109.

Reynolds, R.E. (1992). Selective attention and prose learning: Theoretical and empirical research. *Educational Psychology Review* 4(4): 345–391.

Schmidt, H.G. (1982). Activatie van voorkennis, intrinsieke motivatie en de verwerking van Tekst. (Activation of prior knowledge, intrinsic motivation and comprehension of text). PhD-thesis, Apeldoorn: Van Walraven.

Schmidt, H,G. (1993). Foundations of problem based learning: Some explanatory notes. *Medical Education* 27: 422–432.

Schmidt, H.G., De Volder, M.L., De Grave, W.S., Moust, J.H.C. & Patel, V.L. (1989). Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology* 81: 610–619.

Tans, R.W., Schmidt, H.G., Schade-Hoogeveen, B.E.J. & Gijselaers, W.H. (1986). Sturing van het onderwijsleerproces door middel van problemen: een veld experiment (Directing the learning process by means of problems: A field experiment). *Tijdschrift voor Onderwijsresearch* 11: 35–46.

Van den Broek, P., Fletcher, C.R. & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes* 16: 169–180.

Webb, N.M. & Sullivan-Palincsar, A. (1996). Group processes in the classroom. In D.C. Berliner & R.C. Calfee, eds, *Handbook of Educational Psychology*, pp. 841–873. New York: Simon & Schuster Macmillan.