Higher Education 14 (1985) 643–650
Elsevier Science Publishers B.V., Amsterdam – Printed in the Netherlands

PEER TEACHING: ACADEMIC ACHIEVEMENT OF TEACHER-LED VERSUS STUDENT-LED DISCUSSION GROUPS

MAURICE L. DE VOLDER, WILLEM S. DE GRAVE and WIM GIJSELAERS
Educational Development and Research, University of Limburg, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

ABSTRACT

Problem-based small-group discussions are the cornerstone of health sciences education at the University of Limburg (The Netherlands). In each of three courses, fifteen discussion groups of about eight students were randomly assigned a staff-tutor (control condition) or student-tutor (experimental condition).

In two of the courses no significant differences in cognitive test achievement between the two conditions were found. In one course students tutored by staff-tutors performed significantly better than students tutored by student-tutors. However, no significant differences in test performance were found between students tutored by high-achieving versus average-achieving student-tutors. These findings contradict the "congruence" hypothesis from information processing theory. Several other explanations are explored.

Health sciences education at the University of Limburg (Maastricht, The Netherlands) features problem-based small-group discussions alternated with periods of self-directed learning. This problem-based self-directed learning method is designed to teach problem-solving skills, self-learning skills and enhance motivation and knowledge retention (Schmidt and De Volder, 1984). One of the cornerstones of this method is the discussion group consisting of about eight students in which health science phenomena (so-called problems) are analyzed and learning goals are formulated (Schmidt, 1983). These groups are guided by tutors who are faculty members. Their role is to stimulate the discussion, not to give lectures. In 1981 it was decided to explore the possibility of assigning the role of tutor not only to faculty staff but also to (undergraduate) students. The main reason for this was what Goldschmid and Goldschmid (1976, p. 14) called "economic considerations": "Given today's high student–faculty ratios, particularly in the larger universities, an increase of the number of teachers appears highly desirable. At a time of financial constraints, indeed of budgetary *cuts*, however, such a proposition would be totally unrealistic. In this context, it should perhaps be underlined that peer teaching represents one of the few instructional innovations which does not call for an immediate additional investment".

It is important to note that in the educational programs described by Goldschmid and Goldschmid, peer teaching in small groups is only a "supplement" to the lectures given to large groups by faculty and is aimed mainly at non-cognitive goals such as motivation, socio-psychological needs, active involvement of the learner and skills in cooperation. As pointed out earlier, at the University of Limburg discussion groups are the main mode of instruction, and are used for cognitive as well as non-cognitive educational goals. Of course, the Faculty Board was especially interested whether student-led groups differed from teacher-led groups with respect to cognitive achievement. In two recent reviews of the literature (Goldschmid and Goldschmid, 1976; Collier, 1979) it was found that most studies: (a) deal with the comparison of, say, the lecture and discussion method, or independent study; (b) compared student-led versus teacher-led groups only with respect to non-cognitive outcomes; (c) are only of a descriptive nature (no outcome measurements at all). In our own review of the literature, we could only find two studies comparing cognitive achievement of student-led versus teacher-led discussion groups.

Rabe (1973) divided 70 students enrolled in two college-level introductory health courses into eight groups, represented as equally as possible concerning sex, age, year, and major in school. Student discussion leaders were chosen by democratic and volunteer methods. Students were pre- and post-tested with the Kilander Health Knowledge Test and the Meise Scale for Measurement of Attitudes towards Heathful Living. Based on results of the analysis and interpretation of the data, their conclusions were: (a) student-led discussion groups appeared to be at least as effective in transferring knowledge as teacher-led discussion groups; (b) student-led discussion groups seemed to be as effective in developing positive attitudes towards healthful living as teacher-led discussion groups; (c) variety of teaching methods should be used in teaching, since many students in both teacher-led and student-led discussion groups did not respond appropriately in knowledge gain or attitude change; and (d) direct contact with the teacher is not always mandatory for desired learning to take place; the student-led groups were without the teacher approximately one-third of the time but resulted in similar gains in knowledge and changes in attitudes as the teacher-led groups. However, it is unclear whether the teacher informally tested student progress during the course and perhaps corrected students from student-led groups more in the two-thirds of class time when "normal" instruction took place.

Clement (1971) found no differences between student-led and instructor-led discussion groups on an immediate post-test of cognitive achievement, but found on an achievement test administered six weeks later that the student-led discussion groups outperformed the instructor-led discussion groups. Clement noted that both groups benefited from discussion but student-led groups showed better retention of this effect. According to Clement, this is due to the increase in number of responses by each student and the greater likelihood of encoding in the student-led groups. Remarkable was the fact that three student-led groups each consisted of four students, while the instructor-led group comprised eleven students. This somewhat compromises the methodological attraction observed by Cornwall (1979) who pointed out that comparing test results of teacher-led versus student-led groups is methodologically attractive because the experimental variable – the kind of teacher – can be changed without causing any other incidental but significant change in conditions, as is the case when one compares, for instance the discussion group and lecture methods. Cornwall also reports that there exists a theoretical line of reasoning supporting the claim that peer teaching stimulates and promotes the relatively small-scale restructuring and tuning of the learner's semantic network at least as effectively as the regular teacher. When learning is viewed as information processing (Lindsay and Norman, 1977), the likely degree of "congruence" between the student-teacher and the student's semantic network representation of the particular subject matter and between their more general global or background cognitive network, would result in more effective tutoring if provided by a peer than by the teacher-tutor. Role theory, also, suggests that students will learn more effectively from people of approximately their own age (role models) than from teachers who are older and different in general outlook and culture (Sarbin, 1976).

In the following section we will describe an experiment to test the difference in cognitive achievement between student-led and teacher-led discussion groups in the setting of a problem-based curriculum.

Method

The University of Limburg offers a four year Health Sciences Program (De Volder and Thung, 1983). The first year is divided into six periods of six weeks each, the so-called block periods or blocks. Each block is devoted to a certain theme. In the academic year 1982–1983 the themes of the six blocks were, in chronological order: general introduction to the program, introduction to research methodology, introduction to social health, introduction to health education, introduction to nursing science, and introduction to health care administration. Study groups of about eight students meet twice a week for two hours to discuss theme-relevant topics. Groups are monitored by a faculty member, the so-called tutor. The role of the tutor consists mainly of stimulating the discussion, not giving lectures. In the academic year 1982–1983 an experiment was carried out. Students of the fourth year (who are still undergraduates) were allowed to function as tutors in the first year. In 1982–1983 there were 148 first year students divided into fifteen groups. Composition of the groups was changed at random after each block. Students who (in their third year) volunteered to tutor were permitted to do so in the following year. No selection was made and students were accepted until the necessary quotum was reached. Student-tutors followed the same three-day tutor training course as staff-tutors were required to follow. This training took place in the beginning of the academic year. This means that student-tutors could only be used in the last four blocks of the year. Tutors are free to indicate their preference for a certain block and, if possible, their choices are honored. Within a block, however, tutors are allocated to groups at random, and only one group per tutor. It appeared that, in general, student-tutors preferred the earlier blocks and staff-tutors the later blocks. As it turned out, in the third block nine out of fifteen tutors were students, in the fourth block five out of fifteen, and in the sixth block three out of fifteen. As to the fifth block, data collection was aborted when we discovered that the planning committee responsible for the fifth block decided not to construct an achievement test to be taken by students at the end of the block. So, we ended up with three blocks in which we could compare performance of students tutored by staff with performance of students tutored by senior students on an achievement test consisting of true–false items and assessing mainly factual knowledge. The block three test contained 98 items, the block four test 65 items and the block six test 75 items. Reliabilities (Cronbach's alpha) were respectively 0.78, 0.64 and 0.70. For each block, a *t*-test for independent samples was performed with student-tutor versus staff-tutor as independent variable and student performance on the true–false test as dependent variable. For the third block only, a *t*-test for independent samples was performed with high-achieving versus average-achieving student-tutors as independent variable and student performance on the true–false test as dependent variable. Because there were only a few student-tutors in the fourth and sixth blocks, we decided not to perform the last analysis in blocks four and six. Student-tutors were labeled high-achieving when in the previous academic year they scored more than one standard deviation above the mean of their year group on the comprehensive examination (three true–false tests each containing 100 items) representing all disciplines in the programme. Student-tutors who scored between minus one and plus one standard deviation from the group mean were labeled average achievers. Thus, in the third block five student-tutors were labeled high-achievers and the other four average--achievers. Probably due to a process of self-selection, no student-tutor fell into the category of low-achievers (one standard deviation below the group mean).

Results and Discussion

The results are shown in Table I. In the third block, students in student-led discussion groups scored significantly lower on the achievement test than students in teacher-led discussion groups. In the fourth and sixth block, there were no significant differences in test achievement between students from student-led versus teacher-led discussion groups. In the third block, there were no significant differences between test scores of students tutored by high-achieving student-tutors and students tutored by average-achieving student-tutors (see Table II).

TABLE I

Test Achievement of Students from Student-led versus Teacher-led Discussion Groups. (The maximum score on each test is 12 credit points.)

	Block period 3		Block period 4		Block period 6	
	Student-led	Teacher-led	Student-led	Teacher-led	Student-led	Teacher-led
N	84	56	47	91	23	110
Mean	7.15	7.97	9.59	9.90	8.36	7.34
S.D.	2.27	2.12	2.26	2.55	3.04	2.97
t-value	−2.16		−0.69		1.50	
d.f.	138		136		131	
prob. (2-tail)	0.03		0.49		0.14	

TABLE II

Test Achievement of Students from Discussion Groups led by Average-achieving (AA) versus High-achieving (HA) Student-tutors. (The maximum score on each test is 12 credit points.)

	AA	HA
N	26	36
Mean	7.61	7.02
S.D.	2.37	2.00
t-value	1.07	
d.f.	60	
prob. (2-tail)	0.29	

Without question, the most interesting finding of our research was that, in the third block, students in discussion groups tutored by other, more advanced, students performed less well on an achievement test than did students in discussion groups tutored by staff-tutors. Although we were not able to design a double-blind experiment, it is hard to theorize why and how this effect could be explained away by spurious factors such as the Hawthorne-effect, or other biases. In fact, the Hawthorne-effect would predict higher achievement scores for the experimental group, i.e., the student-tutor group. Other biases, such as a bias against student-tutors perceived as non-experts, could influence the appraisal of tutor functioning but it is hard to conceive that students would go as far as to underachieve on an objective achievement test used for summative purposes.

In the third block (we shall discuss results from the two other blocks later) our findings are opposite to the hypotheses derived from information-processing theory (Clement, 1971; Cornwall, 1979) and role theory (Sarbin, 1976). It is not so easy to explain why. One often-heard and somewhat maliciously phrased explanation is called the "pooling-of-ignorance" effect which shows in student-led groups. Staff-tutors possess more knowledge than student-tutors: therefore, they should be better able to monitor the fine details of the discussion. In the eyes of many teachers this is most probably the case. We tried to verify this assumption by comparing achievement of students tutored by high-achieving versus average-achieving student-tutors. Since these tutors also differ in knowledge, one would expect that students tutored by high-achieving student-tutors would perform better on the end-of-block test than students tutored by average-achieving student-tutors.

However, we found no significant differences in test results of students tutored by these two categories of student-tutors. On the contrary, the trend favored the average-achieving student-tutors. Although this does not completely disqualify the "pooling-of-ignorance" hypothesis, it does mean a serious setback for it. So we kept searching for other possible explanations for the apparent superiority of teacher-led groups. Is it because staff-tutors know better what is expected of students on the achievement test? This is not very likely. Although the content of the test is not known beforehand by the tutors, staff-tutors could be better informed if they were the ones who made up the tests. However, this is generally not the case. The content of an end-of-block test is decided upon by the planning committee responsible for that block. Of course, a tutor may be a member of that committee. As it happened, this was not the case in block three. Perhaps student-tutors are even more acquainted with the test than staff-tutors since they took a similar test in the past while the staff-tutors did not. Another possible explanation is that staff-tutors – more than their student colleagues – were tempted to teach in a classical way, that is to give lectures. Moreover, if this were so, they would have an advantage in terms of knowledge fund. Unfortunately, we are unable to verify this intriguing assumption since we do not possess data relevant to this question.

As to the other two blocks, we did not find significant differences in test achievement between student-led and teacher-led discussion groups. There are a number of possible interpretations for this lack of significance. The differences in reliabilities between the three achievement tests could be responsible, but these differences are altogether not so impressive (from 0.64 to 0.78). Then it is not impossible that the effect is related to a certain subject matter content, although it is hard to understand why. Perhaps it is even too farfetched to look for interactions with subject matter, because it is a fact that on the whole the conditions for finding something in block three were simply more favorable than in the other two blocks: test reliability was after all slightly higher and the ratio of student-tutors to staff-tutors was closer to one.

We did not combine chance probabilities of the three statistical tests because of the interrelatedness of the samples: not only were the same students (subjects) involved – although in differently constituted groups – but also to a certain degree the same tutors, staff-tutors as well as student-tutors. Especially the latter observation could be important for interpreting our findings. Exactly half of the student-tutors of blocks four and six had already tutored in block three. This means that by gaining more practical experience, student-tutors could eventually catch up with staff-tutors. Post hoc analysis of the data, however, did not support this hypothesis since we found no differences between first-time and second-time student-tutors with respect to test achievement of the students they were tutoring. This is consistent with the findings of De Volder (1982) who found no significant correlation between years of practical tutor experience of staff-tutors and tutor functioning as evaluated by students.

Since the Faculty Board decided to continue the experiment with student-tutors on a small-scale basis, we will be able to investigate whether our findings can be replicated. In addition, more attention will be given to "process variables," by which we mean the actual tutoring behavior of student-tutors and staff-tutors. When these have been studied, we will be able to look for possible explanations in terms of tutoring behavior for the differences – if any – in achievement of students from student-led versus teacher-led discussion groups.

References

Clement, D. E. (1971). "Learning and retention in student-led discussion groups," *Journal of Social Psychology*, 84: 279–286.

Collier, K. G. (1980). "Peer-group learning in higher education: the development of higher order skills," *Studies in Higher Education* 5: 55–62.

Cornwall, M. G. (1979). *Students as Teachers: Peer teaching in Higher Education*. University of Amsterdam: C.O.W.O.

De Volder, M. L. (1982). "Discussion groups and their tutors: relationships between tutor characteristics and tutor functioning," *Higher Education* 11: 269–271.

De Volder, M. L. and Thung, P. J. (1983). "Implementing a problem-based curriculum: a new social health programme at the University of Limburg, the Netherlands," *International Journal of Institutional Management in Higher Education* 7: 225–233.
Goldschmid, B. and Goldschmid, M. L. (1976). "Peer teaching in higher education: a review," *Higher Education* 5: 9–33.
Lindsay, P. H. and Norman, D. A. (1977). *Human Information Processing.* (2nd ed.) New York: Academic Press.
Rabe, A. N. (1973). "A comparison of student-led discussion groups to teacher-led discussion groups of teaching college introductory health courses." Paper presented at the Scientific Forum of the American School Health Association, Chicago, Illinois (E.R.I.C. nr. ED103410-SP009017).
Sarbin, T. R. (1976). "Cross-age tutoring and social identity," in: V. L. Allen (Ed.), *Children as Teachers.* New York: Academic Press.
Schmidt, H. G. (1983). "Problem-based learning: rationale and description," *Medical Education* 17: 11–16.
Schmidt, H. G. and De Volder, M. L. (Eds.) (1984). *Tutorials in Problem-based Learning: A New Direction in Health Professions Education.* Assen, The Netherlands: Van Gorcum.