Education for Health, Vol. 18, No. 1, March 2005, 45 – 61

Taylor & Francis
Taylor & Francis Group

PROBLEM-BASED LEARNING

On the Additional Value of Lectures in a Problem-Based Curriculum

HENK VAN BERKEL[1] & HENK SCHMIDT[2]

[1]Maastricht University, Maastricht, and [2]Erasmus University Rotterdam, The Netherlands

ABSTRACT **Objective:** *The purpose of this article is to report on a study conducted to investigate the additional value of lectures in problem-based learning. We hypothesized that lecture quality, as indicated by students, would have a positive influence on time students would spend on self-study, that they would increase intrinsic interest in subject matter and would have a positive influence on achievement. Lecture quality would, in this view, add to other factors that play a role in problem-based learning, such as prior knowledge of students, the quality of problems presented, the functioning of the tutor and small-group collaboration.*

Method: *The hypothesis on the influence of lecture quality on learning was assessed by means of testing a structural equating model, using data from 1500 students.*

Discussion and Conclusions: *We concluded that lecture quality does not affect time spent on study, does not add to achievement and does not influence intrinsic interest in subject matter. Post-hoc analysis, however, demonstrated the quality of lectures was related to other components of problem-based learning, not part of the presented model, such as students' perceptions of lectures as organizer of their study efforts or lectures as a means of putting the problems into a broader perspective.*

KEYWORDS *Lectures, lecture-based learning, problem-based learning, modelling, health sciences education.*

Introduction

The purpose of this article is to report on a study designed to investigate the additional value of lectures in problem-based curricula. Problem-based learning (PBL) is by now a well-established method of learning and instruction (Kaufman & Mann, 1999). Although originally developed in the context of

Author for correspondence: Henk van Berkel, Associate Professor, Maastricht University, Faculty of Health Sciences, PO Box 616, 6200 MD Maastricht, The Netherlands.
Tel: + 43 388 5750. Fax: + 43 388 5779. E-mail: h.vanberkel@educ.unimaas.nl

Education for Health ISSN 1357-6283 print/ISSN 1469-5804 online © 2005 Taylor & Francis Group Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13576280500042648

medical education, the method has been implemented in other domains as well (Gijselaers *et al.*, 1995). PBL emphasizes small-group work on problems, self-directed learning and de-emphasizes teacher-controlled activity such as lectures. Lecturing has repeatedly been shown to be a quite ineffective means of improving learning and achievement (Bligh, 1998). Lecturing does not particularly advance problem-solving skills, nor does it require creative or critical thinking or prepare students for the types of problems they will face as professionals (Johnson, 1999). These are the very the characteristics that PBL claims to impart on students. Before turning to the issue of lectures and PBL, we will review studies comparing effects of lecture-based learning (LBL) programs with those of PBL on student's learning outcomes.

Review of recent studies comparing PBL with LBL

In 1993 the British General Medical Council recommended major changes in the medical undergraduate course, e.g. reducing the duration of the curriculum. In response, the Nottingham School of Medical Education shortened the curriculum and replaced the lectures with problem-oriented small workshops. After this curriculum innovation, Singh *et al.* (1998) compared the attitudes of students towards psychiatry and mental illness. They concluded that, notwithstanding the substantial reduction, the new curriculum appeared to be as effective as a longer curriculum. Lancaster *et al.* (1997) and Kaufman & Mann (1997) also reported that the attitudes of PBL students were significantly more positive towards the content of the curriculum than those of LBL students.

The Medical Faculty of the University of Cologne in Germany also attempted to replace lectures with PBL. In their evaluation of this curriculum change, Antepohl and Herzig (1999) in particular investigated the changes in factual knowledge of the students. It seemed that PBL did not lead to less factual knowledge. Moreover, students considered PBL an effective learning method and favoured it over the lecture format. Students reported positive effects of PBL in terms of use of additional learning resources, interdisciplinarity, teamwork and learning fun. In general, results show a favourable impact of PBL on perceptions of the academic environment (Lieberman *et al.*, 1997; Lancaster *et al.*, 1997).

A study by Login *et al.* (1997) reported PBL students had better results on medical knowledge tests compared to LBL students. There is also evidence (Cariaga-Lo *et al.*, 1996) that students entering a PBL curriculum score higher on the Medical College Admission Test in the United States and have a higher undergraduate grade point average than students who entered the LBL curriculum. This suggests that sometimes the prior knowledge or even the learning skills of medical students, who are given the choice between a PBL and a LBL curriculum, may be different. On the other hand, Aaron *et al.* (1998) reported a controlled study which indicates that PBL students acquire better learning styles, leading to higher scores on exam questions which examine elaborated knowledge contrary to dispersed knowledge. To what extent this result can indeed be attributed to their learning styles is unclear. Menec *et al.* (1995) showed that although PBL students scored higher on achievement tests, the results are to a large degree explained by students' personal characteristics. Camp *et al.* (1994) also demonstrated that students' characteristics could cause the differences between the achievements scores.

Within the context of continuing medical education, Doucet *et al.* (1998) conducted a study in which they compared PBL and LBL students in terms of knowledge and clinical reasoning with respect to the diagnosis of headache. The PBL group scored significantly higher on both dimensions. In addition, the satisfaction of PBL students with several program aspects was significantly higher that that of the lecture group. Schreiber (1997) and Richards *et al.* (1996) conducted a similar study in the domain of internal medicine. The results also favoured PBL.

There is also evidence that PBL students spent more time preparing for the Medical Licensing Examination (Richards & Cariaga-Lo, 1994). This result was seen as a positive effect of the PBL curriculum. In another study, however, the increase in study time was seen as a disadvantage (Rand & Bagliono, 1997). As will be discussed further on, time spent on individual study is a variable in educational research that is difficult to explain using elements of the learning situation, be it small-group learning or lecture-based learning. Van Berkel and Schmidt (2000), for instance, examined a path-analytic PBL model in which several elements of PBL were investigated, including self-study time. A multiple R-coefficient, indicating the extent to which self-study time was explained by other variables, only reached a value of 0.20.

Although some studies have revealed inconsistencies, the overall results indicate relatively strong evidence in favour of PBL, compared to LBL, on some important learning outcomes. However, do these results imply that lectures could be discarded from problem-based curricula? Lectures have some positive role to play in a PBL curriculum too. A lecture as a means of transmitting knowledge to students may be ineffective. However, lectures can clarify the underlying structure of the curriculum. They can guide students to undertake learning activities in a desirable direction, and they can arouse interest in the subject matter (Bligh, 1998). The question arises if it is possible to design a curriculum in which lectures are an integral part of the learning process, not disrupting but reinforcing it? Would these lectures, under such conditions, contribute to the learning outcomes?

Integrating lectures into a model of problem-based learning

The studies reviewed thus far concentrate on comparisons of outcomes of different types of curricula. These studies largely ignore those elements in the curricula that are responsible for the different outcomes. In particular, if one is interested in the relative contributions of the quality of lectures to students' learning processes, one needs a different approach, an approach more geared toward the processes of learning and instruction that lead to the various outcomes. One such approach is the model-of-school-learning approach advocated by Carroll (1963) and Bloom (1976). In their view, learning within the context of a school has three categories of variables. The first category contains input variables, such as the characteristics of students, the behaviour of teachers and the learning materials. The second category comprises the intervening process variables: learning activities carried out by students, time spent on study and features of the instructional process. The third category consists of cognitive output variables, e.g. achievement, and affective outcomes like interest in subject matter studied. In the view of the model-of-school-learning approach, relationships between these variables should be the focus of educational research. Gijselaers and Schmidt (1989) have proposed a theoretical model of problem-based learning that is based on this approach to learning and instruction (see Figure 1).

The input variables are the student's prior knowledge, the quality of the problems and the effectiveness of the tutor. The intervening process variables are the extent to which the tutorial group functions effectively and the amount of time students spend on self-directed activities. The output variables are identified as the score on the achievement test and the increased interest in the topic studied. The arrows in Figure 1 indicate the directions of the influences from one variable to another. According to the model, an increase in the magnitude of one of the variables causes an increase of the magnitudes of other variables.

In a series of studies applying structural equations modelling techniques (Bentler, 1989), Gijselaers and Schmidt (1989), Schmidt and Gijselaers (1990), and Schmidt (1999) found empirical support for this model. More recently, Van Berkel and Schmidt (2000) added another intervening variable to the model, namely the willingness, or commitment of students to actually engage in the problem-based learning process, operationalized in terms of tutorial group meeting attendance. This extension improved the model slightly but statistically significantly.

The present study extends this work. It directs attention to the additional value of lectures on the proposed problem-based learning model as represented in Figure 1. Lectures can be seen as an independent source of information on what to study and how to study (Bligh, 1998) and, therefore, are considered as an input variable, in addition to prior knowledge, the quality of a problem and the effectiveness of the tutor.

Methods

Subjects and procedure

Subjects were all undergraduate students of years 1–3 of a problem-based four-year health sciences curriculum. The number of students in each cohort was

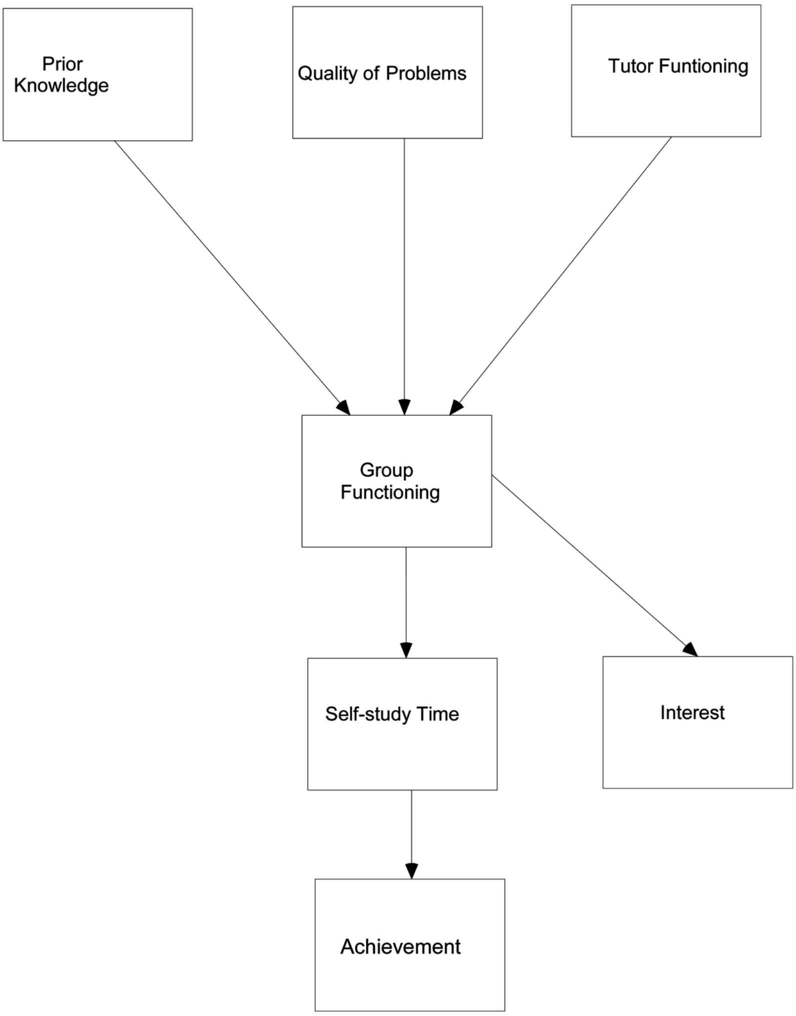

Figure 1. Theoretical Causal Model of Problem-Based Learning (Gijselaers & Schmidt, 1989).

around five hundred. The first three years of this curriculum consisted of series of consecutive units, each of them lasting six weeks. In the fourth year students write their master thesis. This final year was excluded from the present study. At the end of every unit, students filled out a questionnaire detailing aspects of the learning experiences in the particular unit. Because there are six units per year, the total number of questionnaires was 6 x 3 x 500. So the theoretical maximum response was 9000. As the overall response rate was 78%, (year 1 83%, year 2 80%, and year 3 74%) 7013 questionnaires form the basis of the investigation. However, data were aggregated, because some variables cannot be considered independent scores (Marsh, 1980). First, variables, such as lecture quality and quality of problems, were dependent on the specific unit of which they are part. Thus, scores on these variables were aggregated at the unit level. Next, all other variables were aggregated at the tutorial group level, because achievement and other variables were related to members of the same tutorial group and therefore cannot be considered independent scores. Finally, the level of analysis was the tutorial group. In total, 700 groups were involved in the study.

Instruments

The rating scale consisted of 23 Likert-type items, covering the various dimensions of PBL as outlined in Figure 1. The 5-point scale varied from ‘‘totally disagree’’ to ‘‘totally agree’’. The items are summarized in Table 1.

In previous factor-analytic reliability and validity studies, these dimensions have been shown to have good average intraclass-coefficients (as a measure of interrater agreement) and construct validity (Schmidt *et al.*, 1995).

An achievement test was administered to each student at the end of each unit. These so called unit tests usually consist of about 200 true-false items. The raw score was transformed to a 4-point scale (1 indicating ‘‘insufficient’’, 2 indicating ‘‘doubtful’’, 3 indicating ‘‘sufficient’’, and 4 indicating ‘‘excellent’’). The reliabilities of these tests (coefficient alpha) ranges from 0.81 to 0.93, which is relatively high.

Statistical analysis

The goal of this investigation was to study the supplemental merit of lecture quality to the PBL model. The two models were compared: one without the lecture quality (the general PBL model), and the new one with lecture quality as one of the components. To investigate and to compare two or more models, one has to apply model fitting techniques derived from structural equating theories (see Box 1). The statistical package AMOS (Arbuckle & Wothke, 1999) was used.

Results

Table 2 shows the intercorrelations, means and standard deviations of the variables.

The theoretical model displayed in Figure 1 was tested against these data. There were 700 groups involved in this study. This is a relatively large number, which implies that the fit indices commonly used, such as the Chi square and P, are not applicable (Arbuckle & Wothke, 1999). The values of other indices are: CMIN: 27.32; FMIN: 0.48; Fa: 0.46; PCLOSE: 0.00; RMSEA: 0.20; TLI: 0.38, and CFI: 0.64. These findings indicate that the theoretical model displayed in Figure 1 does not adequately describe the data. Other authors (Gijselaers & Schmidt, 1989; Schmidt, 1999; Van Berkel

Table 1. Variables and their operational formulations (abbreviation used in text)

Variables	Operational formulation
Amount of Prior Knowledge (Prior Knowledge)	– This unit fitted well with my prior knowledge
Quality of the Problems (Quality of Problems)	– The problems were clearly formulated – The problems were a good starting point for the group discussion – The problems were motivating to start the self-study
Tutor Performance (Tutor Functioning)	– The tutor functioned well – Rate the functioning of the tutor (1–5) – The tutor stimulated analysis of the problems – The tutor stimulated the group through her/his expertise – The tutor showed interest in my study activities
Lecture quality (Lecture Quality)	– The lectures matched the subjects I was studying at that moment – The subject-matter was clearly presented in relation to the goals
Tutorial-Group Functioning (Group Functioning)	– The meetings were productive – The meetings were pleasant
Time Spent on Individual Study (Self-study time)	– How many hours a week did you spend on self-study?(Indicate in whole hours)
Intrinsic Interest in Subject-Matter (Interest)	– The subject-matter was interesting

& Schmidt, 2000) found a similar model to be too restrictive as well. Therefore, we studied several feasible alternatives to the present model. The best of these is displayed in Figure 2.

The purpose of our research was to investigate the supplemental merit of lecture quality to the PBL model. Therefore, in addition to the modified PBL-model (model A) displayed in Figure 2, we added an alternative model (model B) which included the new variable ‘‘Lecture quality’’ (see Figure 3). We assumed, based on the theoretical considerations mentioned in the introduction part of this article, that lecture quality influences time spent on study, intrinsic interest in subject matter and achievement score.

The models from both Figure 2 and Figure 3 are compared in Table 3.

In general, there are no, or at least very small, differences between the coefficients derived from both models. This result implicates that adding lecture quality does not lead to a change of the squared multiple correlations and the beta weights. Table 4 shows the comparison of the fit indices between the two models.

Table 2. Intercorrelations, means and standard deviations of the variables ($N = 700$)

	1	2	3	4	5	6	7	8
1. Amount of prior knowledge	-							
2. Intrinsic interest in subject matter	0.48*	-						
3. Lecture quality	0.30*	0.35*	-					
4. Time spent on study	0.06	0.15*	– 0.02	-				
5. Achievement	0.06	0.07	0.09*	0.02	-			
6. Quality of the problems	0.42*	0.59*	0.58*	0.06	0.08*	-		
7. Tutor performance	0.18*	0.17*	0.16*	– 0.01	0.02	0.22*	-	
8. Small group functioning	0.27*	0.32*	0.24*	0.07	0.04	0.32*	0.58*	-
Means	3.5	3.8	7.2	16.5	3.1	10.9	19.3	7.6
Standard deviations	0.58	0.71	0.95	4.90	0.34	1.13	2.64	1.13

*Significant at the 0.05 level (2-tailed).

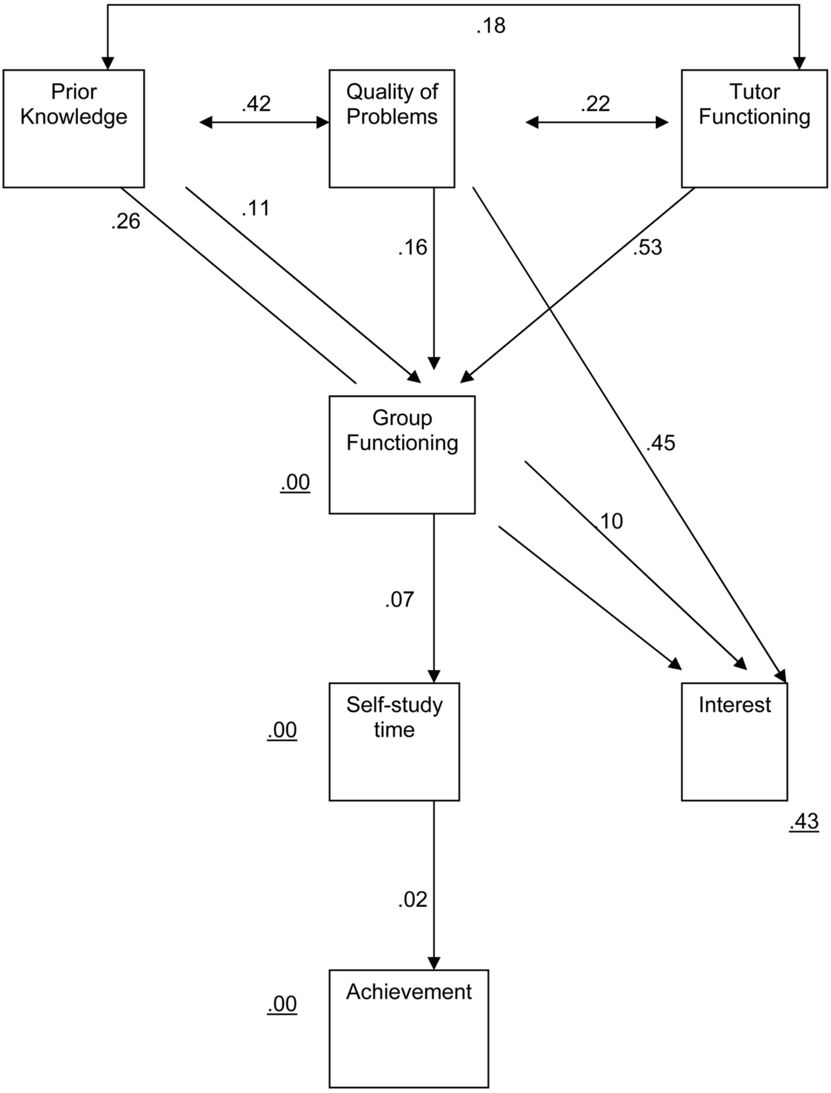

Figure 2. Alternative Causal Model of Problem-Based Learning.

As can be deduced from RMSEA and the corresponding *p*-value PCLOSE, both models yield a good fit with the data. Based on the fit indices, neither model is preferred.

A third way to compare the two models is to investigate the differences of the chi-square statistic. The null-hypothesis is that both models fit the PBL-model equally well. A test of the stronger model (Model B: more variables) against the weaker one (Model A: less variables) can be obtained by subtracting the smaller chi-square statistic from the larger one. The new statistic is 0.225 (i.e.

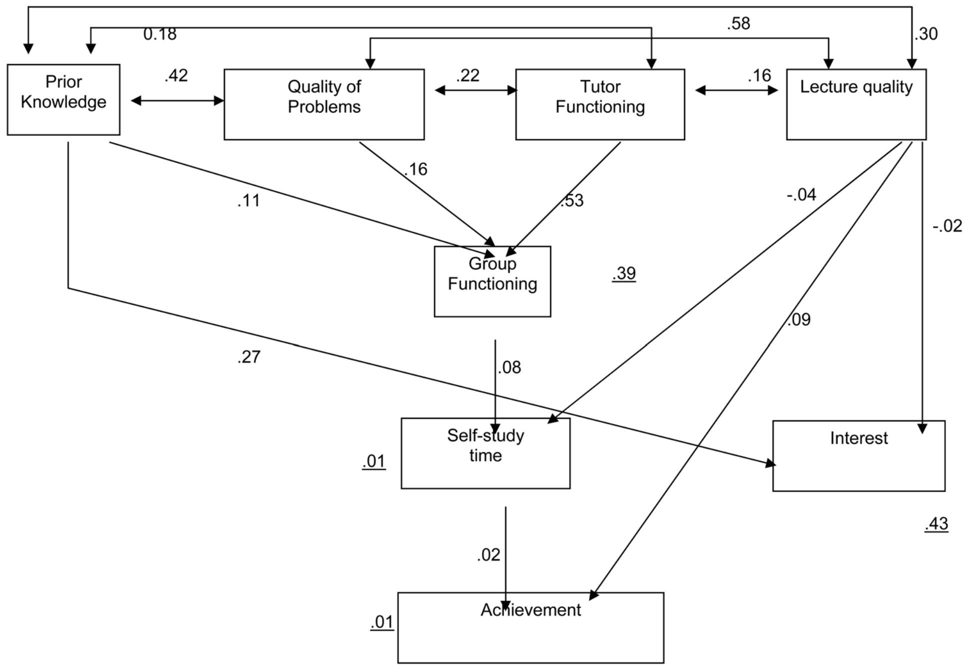

Figure 3. Alternative Causal Model of Problem-Based Learning including Lectures.

22.865–22.640). If model B is correctly specified, this statistic will have an approximate chi-square distribution with degrees of freedom equal to the difference between the degrees of freedom on the competing models. The difference in degree of freedom is 1 (i.e. 11–10). That is, Model B imposes all of the parameter constraints of Model A, plus an additional 1. With 1 degree of freedom, chi-square values greater than 3.841 are significant at the 0.05 level. Based on this test, the null-hypothesis is not rejected. Adding a new variable to the PBL model, lecture quality, does not lead to a better explanation of the PBL model.

Discussion

The purpose of the study was to investigate the auxiliary role of lectures in PBL. We hypothesized the quality of lectures to be an independent input variable, which would entertain causal relationships with time spent on study, intrinsic interest in subject matter and achievement. The fits of two models were computed: one with lecture quality included as a determinant of learning, the other without lecture quality. If the hypothesis that lecture quality plays a positive role in problem-based learning was valid, the model, which included

Table 3. Squared Multiple Correlations, and beta weights of two models

	Model A (excluding lecture quality)	Model B (including lecture quality)
Squared multiple correlations		
Tutorial-group functioning:	0.39	0.39
Time spent on individual study	0.00	0.01
Intrinsic interest in subject-matter	0.43	0.43
Achievement	0.00	0.02
Beta weights		
Quality of problems - > Tutorial-group functioning	0.16	0.16
Tutor performance - > Tutorial-group functioning	0.52	0.52
Amount of prior knowledge - > Tutorial-group functioning	0.11	0.11
Tutorial-group functioning - > Time spent on study	0.07	0.08
Tutorial-group functioning - > Intrinsic interest in subject matter	0.10	0.10
Time spent on study - > Achievement	0.02	0.02
Quality of problems - > Intrinsic interest in subject matter	0.45	0.46
Amount of Prior Knowledge - > Intrinsic interest in subject matter	0.26	0.27
Lecture quality - > Time spent on study		– 0.04
Lecture quality - > Achievement		0.09
Lecture quality - > Intrinsic interest in subject-matter		– 0.02

Table 4. Fit indices of two models (see the box for the meaning of the abbreviations)

	Model A (excluding lecture quality)	Model B (including lecture quality)
Chi-square	22.64	22.86
Df	10	11
p	0.01	0.02
CMIN/DF	2.26	2.08
RMSEA	0.04	0.04
PCLOSE	0.65	0.74
FMIN	0.03	0.03
FO	0.02	0.01
TLI	0.97	0.95
CFI	0.98	0.99

lecture quality, should account for more of the variance of the components, and the general fit should be better. This hypothesis could not be confirmed. Lecture quality appears neither to increase time spent on study nor to increase the test score or intrinsic interest in subject matter.

This conclusion, based on our data, seems to be at odds with other observations made in problem-based programs. For instance, generally lectures are well attended by students. In addition, students find lectures moderately useful as demonstrated by data obtained from program evaluations (Schmidt *et al*., 1995). However, useful for what? As demonstrated in this study, the quality of lectures does not contribute to higher achievement nor does it increase interest in subject matter. There are, however, some interesting correlations between lecture quality on one hand and some model variables on the other (some correlations are reported in Figure 3) which can, at least partly, explain the positive attitudes of students to lectures. There is a relatively high correlation between lecture quality and the quality of the problems in the unit guide (0.58). The higher the evaluations of the lectures, the higher students perceive the quality of the problems (or vice-versa). This result suggests that lectures clarify the problems, which is maybe why students actually attend the lectures. In addition, there is also a significant correlation (0.33) between the evaluation of the lectures and the perceptions of students of the quality of the organization of the educational unit. This result is not studied in depth in this investigation, but it is possible that lectures help students to organize their study efforts. A third correlation is also of interest: the correlation (0.32) between lecture quality and the perceived relevance of the educational unit as a whole. The higher students evaluate the lectures, the more the educational unit is relevant for them, or vice versa. It seems that lectures function as a tool that put the content of the unit into a broader perspective. By attending the lectures, students get an overall view of the unit and of the place of the unit in the whole curriculum. Although this will not lead to a higher achievement score, it can add to the effectiveness of the unit as a means of preparing students for professional practice.

In conclusion, our study does not support the idea that the quality of lectures adds something to the problem-based learning model in terms of its outcomes. In fact, the model as tested in previous studies (e.g., Schmidt, 1999; Van Berkel & Schmidt, 2000), was confirmed in the present study. However, this does not imply that we suggest that lectures should be deleted from these curricula. We believe that lectures may be an important means to help students put their knowledge into a broader, often professionally relevant, perspective. This is an important aspect of higher education that often can only be attained through the direct confrontation between students and these professionals, their teachers. However, if this is true, then it is important for teachers to be aware that lectures mainly fulfil this role, and that using lectures as a means of knowledge transmission is a waste of time. As unearthed in the present study and others, students in PBL seem to have more efficient means to acquire knowledge, as demonstrated by the relations between achievement and other elements of PBL.

References

AARON, S., CROCKET, J., MORRISH, D., BASUALDO, C., KOVITHAVONGS, T., MIELKE, B. & COOK, D. (1998). Assessment of exam performance after change to problem-based learning: Differential effects by question type. *Teaching and Learning in Medicine*, *19*, 86–91.

ANTEPOHL, W. & HERZIG, S. (1999). Problem-based learning versus lecture-based learning in a course of basic pharmacology: A controlled randomized study. *Medical Education*, *33*, 106–113.

ARBUCKLE, J. L. & WOTHKE, W. (1999). *Amos users' guide; version 4.0*. Chicago, IL: SmallWaters Corporation.

BENTLER, P.M. & BONNETT, D.G. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

BENTLER, P.M. (1989). *EQS: Structural Equations Program Manual*. Los Angeles, CA: BMDP Statistical Software Inc.

BENTLER, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

BLIGH, D. (1998). *What's the use of lecture?* (5th ed). Exeter: Intellect.

BLOOM, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

BOLLEN, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

BROWNE, M.W. & CUDECK, R. (1993). Alternative ways of assessing model fit. In, K.A. Bollen & J.S. Long (Eds), *Testing structural equation models* (pp. 36–162). Newbury Park, CA: Sage.

BROWNE, M.W. & MELS, G. (1992). *RAMONA User's guide*. Columbus, OH: The Ohio State University.

CAMP, D.L., HOLLINGSWORTH, M.A., ZACCARO, D.J., CARIAGA-LO, L.D., & RICHARDS, B.F. (1994). Does problem-based learning curriculum affect depression in medical students? *Academic Medicine*, *69* (Suppl.), 25–27.

CARIAGA-LO, L.D., RICHARDS, B.F., HOLLINGSWORTH, M.A., & CAMP, D.L. (1996). Non-cognitive characteristics of medical students entry to problem-based and lecture-based curricula. *Medical Education*, *30*, 179–186.

CARROLL, J.B. (1963). A model of school learning. *Teachers College Record*, *64*, 723–733.

DOUCET, M.D., PURDY, R.A., KAUFMAN, D.M. & LANGILLE, D.B. (1998). Comparison of problem-based learning and lecture format in continuing medical education on headache diagnosis and management. *Medical Education*, *32*, 590–596.

GIJSELAERS, W.H. & SCHMIDT, H.G. (1989). Towards a causal model of student learning within the context of a problem-based curriculum. In, Z. Nooman, H.G. Schmidt & E. Ezzat (Eds), *Innovation in medical education: An evaluation of its present status* (pp. 95–113). New York: Springer Publishing.

GIJSELAERS, W.H., TEMPELAAR, D.T., KEIZER, P.K., BLOMMAERT, J.M., BERNARD, E.M. & KASPER, H. (Eds) (1995). *Educational innovation in economics and business administration: The case of problem-based learning*. Dordrecht, London, Boston: Kluwer Academic Publishers.

JÖRESKOG, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202.

JOHNSON, P.A. (1999). Problem-based, cooperative learning in the engineering classroom. *Journal of Professional Issues in Engineering Education and Practice*, *125*, 8–11.

KAUFMAN, D.M. & MANN, K.V. (1997). Basic sciences in problem-based learning and conventional curricula: Students' attitudes. *Medical Education*, *31*, 177–180.

KAUFMAN, D.M. & MANN, K.V. (1999) Achievement of students in a conventional and problem-based learning (PBL) curriculum. *Advances in Health Sciences Education*, *4*, 245–260.

LANCASTER, C., BRADLEY, E., SMITH, I.K., CHESSMAN, A., STROUP-BENHAM, C.A. & CAMP, M.G. (1997). The effect of PBL on students' perceptions of learning environment. . *Academic Medicine*, *72* (Suppl.), 10–12.

LIEBERMAN, S.A., STROUP-BENHAM, C.A., PEEL, J.L. & CAMP, M.G. (1997). Medical student perception of the academic environment: A prospective comparison of traditional and problem-based curricula. *Academic Medicine*, *72* (Suppl.), 13–15.

LOGIN, G.R., RANSIL, B.J., MEYER, M., TRUONG, N.T., DONOFF, R.B. & MCARDLE, P.J. (1997). Assessment in preclinical problem-based learning versus lecture-based learning. *Journal of Dental Education*, *61*, 473–479.

MARSH, H.W. & YEUNG, A.S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of Math and English constructs. *American Educational Research Journal*, *35*, 705–738.

MARSH, H.W. (1980). The influence of student, course, and instructor characteristics in evaluation of university teaching. *American Educational Research Journal*, *17*, 217–237.

MENEC, V.H., PERRY, R.P. & STRUTHERS, C.W. (1995). The effect of adverse learning conditions on action-oriented college students. *Journal of Experimental Education*, *63*, 281–299.

RAND, J.S. & BAGLIONI, A.J. (1997). Subject-based problem-based learning in the veterinary science course at the University of Queensland. *Australian Veterinary Journal*, *75*, 120–125.

RICHARDS, B.F. & CARIAGA-LO, L.D. (1994). Curriculum type and sophomore students' preparation time for the USMLE step 1 examination. *Evaluation and the Health Professions*, *17*, 329–343.

RICHARDS, B.F., OBER, K.P., CARIAGA-LO, L.D, CAMP, M.G., PHILP, J., MCFARLANE, M., RUPP, R. & ZACCARO, D.J. (1996). Ratings of students' performances in a third-year internal medicine clerkship: A comparison between problem-based and lecture-based curricula. *Academic Medicine*, *71*, 187–189.

SCHMIDT, H.G. (1999, April). *Testing a causal model of problem-based learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, CA .

Schmidt, H.G. & Gijselaers, W.H. (1990, April). *Causal modeling of problem-based learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Schmidt, H.G., Dolmans, D.H.J.M, Gijselaers, W.H. & Des Marchais, J.E. (1995). Theory-guided design of a rating scale for course evaluation in problem-based curricula. *Teaching and Learning in Medicine*, *7*, 82–91.

Schreiber, M. (1997). Comparing how well two undergraduate curricula teach knowledge of internal medicine. *Academic Medicine*, *72* (Suppl.), 127–129.

Singh, S.P., Baxter, H., Standen, P. & Duggan, C. (1998). Changing the attitudes of 'tomorrow's doctors' towards mental illness and psychiatry: A comparison of two teaching methods. *Medical Education*, *32*, 115–120.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–189.

Van Berkel, H.J.M. & Schmidt, H.G. (2000). Motivation to commit oneself as a determinant of achievement in problem-based learning. *Higher Education*, *40*, 231–242.

Model fitting: A brief introduction

The model fit procedure (Bollen, 1989) yields several statistics that allow the investigator to assess the extent to which the empirical data fit the theoretical proposed. Unfortunately, there is no single best statistic that gives insight into the fit of the model. In addition, there are no criteria to evaluate the statistic, e.g. by a test of significance (Arbuckle & Wothke, 1999), although some authors give rules of thumb. The best a researcher can do is to compute several statistics which reflect the fit of the model. The coherence of these measures indicates the model-fit. Browne and Mels (1992) recommend restricting the statistics to the following indices: CMIN (based on minimum value of the discrepancy), P (Probability level), FMIN (also based on the minimum value of the discrepancy), FO (Estimated population discrepancy), PCLOSE (probability of close fit), and RMSEA (Root mean square of approximation). However, while following Browne and Mels (1992), we will add indices recommended by TLI (Tucker-Lewis index) and CFI (Comparative fit index).

These indices can be characterized as follows. Indices, such as CMIN, P, and FMIN, reflect the discrepancy between the covariances in the sample and those in the population. They differ from each other in the way they handle the constraints. Arbuckle and Wothke (1999) argue that the CMIN is the best index of fit. This index is computed by dividing the minimum discrepancy C by its degrees of freedom. The ratio must be less than 5 and preferably less than 3 and close to 1 for correct models. FMIN is another index based on the discrepancy between the sample and the population. Values below 0.05 indicate good fit. Based on the well-known chi-square and the degrees of freedom, a level of significance (p) is computed. In order not to reject the model, it is preferred (in the context of structural equating modeling) that the p-value is higher than .05. In investigations with a large sample (as is the case in this study), p is, however, not a proper index to reflect the fit of the model. This has to do with the fact that statistical hypothesis testing can be a poor tool for choosing a model (Joreskog, 1967). Models are never perfectly correct, which means that there is always a gap between the theoretical basis and practice. Thus models can always be rejected on statistical grounds. Large samples often lead to significant results and to rejecting the null hypothesis. In the case of model-fitting, the null hypothesis is: The covariances in the sample are the same as in the population. The larger the sample, the more likely the null-hypotheses is to be rejected (Arbuckle & Wothke, 1999; Mulaik, personal communication).

Steiger (1990) recommend the use of the discrepancy function (FO) obtained by fitting a model to the population moments rather than to the sample moments. From FO there are other indices deduced depending on the estimation of FO and related quantities. FO incorporates no penalty for model complexity and tends to favor models with many parameters. A general rule is: ‘‘The more parameters there are, the better the fit’’. Steiger (1990) suggests compensating for the effect of model complexity by dividing FO by the number of degrees of freedom. Taking the square root of the resulting ratio gives the population root mean square error of approximation, called RMSEA (Browne & Cudeck, 1993). They also give a rule of thumb for the fit. A value of 0.05 or less would indicate a close fit of the model in relation to the degree of freedom. Of course, a value of zero indicates exact fit. PCLOSE is a ‘‘*p*-value’’ for testing the null hypothesis that the population RMSEA is no greater than 0.05. (By contrast, the ‘‘*p*’’ mentioned above represents the *p*-value testing the hypothesis that the population RMSEA is zero.) Besides indices that compare the hypothesized model with the sample or population moments, there are other interesting indices, which compare the discrepancy of the model with the independence model. In an independence model, it is hypothesized that the observed variables are not correlated. The model therefore fits very badly. By comparing the observed fit with the worst-case scenario, the outcomes of the fit procedure obtain some perspective. Two of these indices are the Tucker-Lewis index, TLI (Bentler & Bonnett, 1980), and the comparative fit index, CFI (Bentler, 1990). They both yield values between zero and one (although theoretically TLI can be higher). Values close to one indicate a very good fit.