Medical Teacher, Vol. 24, No. 5, 2002, pp. 544–549

Taylor & Francis healthsciences

Student evaluation of expert and non-expert problem-based learning tutors

GRAHAM D. HENDRY, HUY PHAN, PATRICIA M. LYON & JILL GORDON
University of Sydney, NSW, 2006, Australia

Introduction

The role of the tutor is important in developing effective group process in educational programs built around small-group, problem-based learning (PBL). The tutor's role includes creating a supportive group climate, encouraging the involvement of group members and addressing group problems when they arise.

Good tutoring has the potential to enhance group process in both novice and experienced teams. It is worthwhile therefore to monitor the quality of tutorial teaching in PBL programs and provide valid and useful feedback to individual tutors.

In addition to 'expertise' in teaching, tutors may have content expertise in the area(s) covered by the patient problem(s). Tutors with content expertise may facilitate students' learning, for example, by intervening in students' discussion with timely statements or questions that evoke relevant ideas and/or clinical reasoning processes. However, studies of the effect of tutor content expertise on student learning have produced inconclusive results (Schmidt & Moust, 2000). Some research has found a positive effect on students' achievement in favour of content experts, while other studies have found mixed or no differences between 'expert' and 'non-expert' tutors (see Schmidt & Moust, 2000, for a comprehensive review). The inconclusive results from this research may be due to differences in (1) the definitions of content expertise and (2) the size of samples used, (3) participating students' level of experience of PBL, (4) students' levels of prior knowledge, and/or (5) the amount of structure provided by the medical school curriculum (Schmidt & Moust, 2000).

The curriculum in Years 1 and 2 of the University of Sydney medical program is structured around PBL tutorials, supported by lectures and other teaching sessions relevant to the problem. A total of 70 clinical problems are grouped into nine units or 'blocks' of study. The first block is introductory, seven blocks are based on body systems and the ninth block is concerned with oncology and palliative care. Prior to each block, all PBL tutors are provided with printed materials specially designed to support their teaching for each problem. These support materials, together with extensive web-based student resources linked to the weekly case (readers may explore the program web site at http://www.gmp.usyd.edu.au/visitors/), provide a highly structured teaching and learning environment.

In this paper, we report the results of an analysis of PBL tutor evaluation data from Years 1 and 2 of the medical program for the period 1998–2000. We conducted a confirmatory factor analysis on a 24-item tutor feedback form completed by students at the end of each block, and compared ratings of 'expert' and 'non-expert' tutors, where expert tutors were broadly defined as having medical training. We used this definition of 'expert' because tutors with medical training have different levels of clinical experience in addition to knowledge of both the basic and clinical sciences. Non-experts were further subdivided into staff with basic science and non-basic science backgrounds. Our analysis does not include an examination of the relationship between tutor expertise and student achievement in the USydMP, because students change their tutors several times during Years 1 and 2, and such an examination would involve a higher level of complexity beyond the scope of this article.

PBL in the University of Sydney Medical Program

The pattern of three weekly PBL tutorials, incorporating a well-defined clinical reasoning process, is illustrated in Figure 1. Students work together in groups of 8–9 with a tutor in three 90-minute tutorials. Problems are triggered by brief audiovisual summaries of patients' presenting conditions delivered to groups on a Faculty intranet prior to any formal teaching (readers may explore the program website at http://www.gmp.usyd.edu.au/visitors/).

In the first tutorial, groups think broadly about the problem and generate hypotheses and learning issues. Between the first and second sessions, students engage in individual reading and typically attend 2–3 lectures and at least one 'theme' session (a theme session is a class or tutorial that involves the practical application of students' knowledge in some way, e.g. laboratory classes in pathology or biochemistry). In the second tutorial, groups share what they have learned, plan their inquiry and obtain patient information before reaching a diagnostic decision. Students spend one day at their clinical school and attend remaining lectures and theme sessions before discussing patient management in the third tutorial.

Tutors

Tutors are recruited from departments and clinical schools within the Faculty of Medicine and vary widely in their specialty and/or discipline backgrounds (e.g. tutors may be qualified medical practitioners, basic science academics or academics with backgrounds in public health or education).

Correspondence: Dr Graham D. Hendry, Department of Medical Education, A27, University of Sydney, Sydney NSW 2006, Australia. Tel: +161-2 9351-5681: fax: +161-2 9351-6646; email: grahamh@gmp.usyd.edu.au

ISSN 0142–159X (print)/ISSN 1466–187X online/02/050544–06 © 2002 Taylor & Francis Ltd
DOI: 10.1080/0142159021000012603

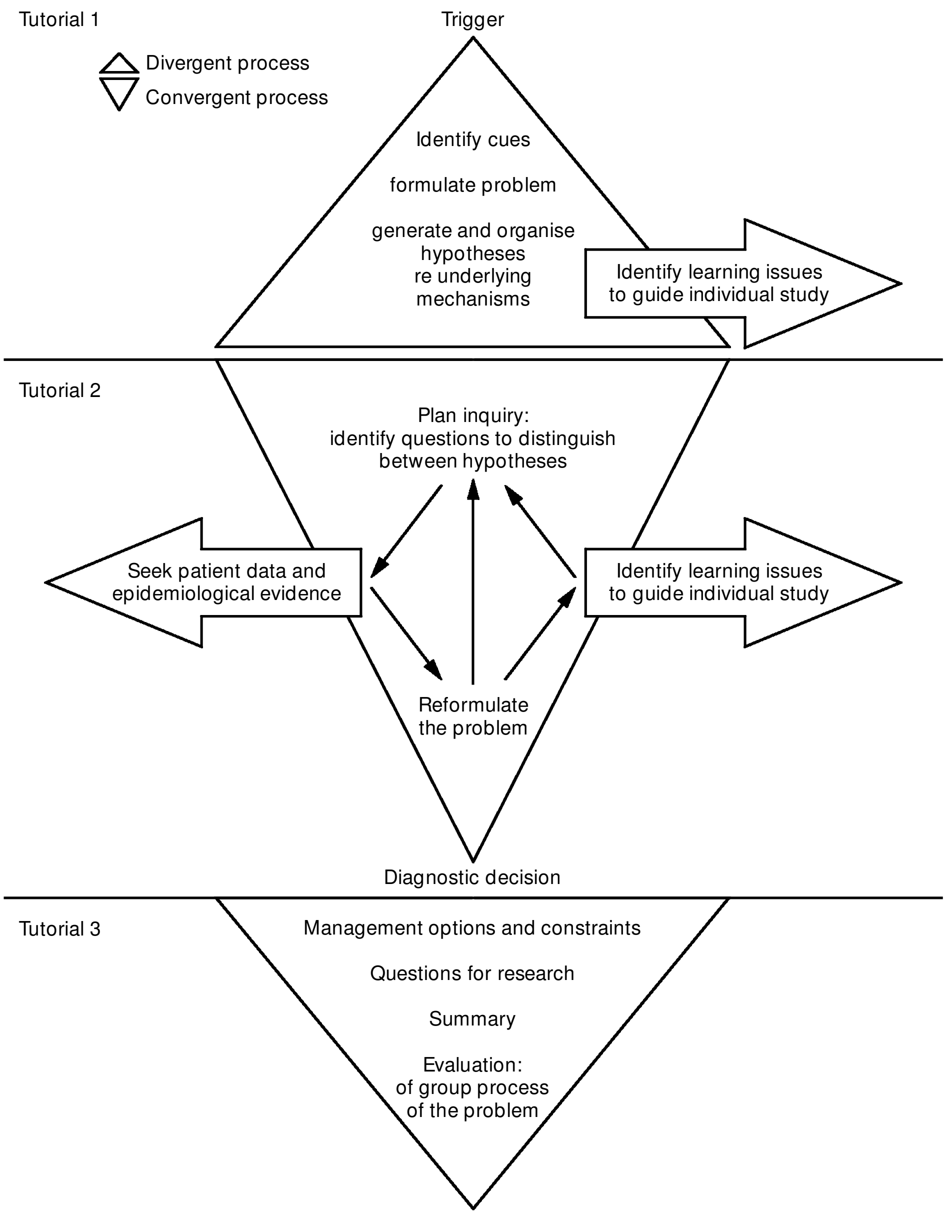

Figure 1. The weekly problem-based learning and clinical reasoning process (adapted from Neame, 1989).

Students in the combined degree program who have deferred to undertake Masters or PhD studies are also invited to be tutors. PBL groups change tutors each block, giving them experience of nine different tutors in the first two years.

Tutors are provided with a tutor guide and patient data sheet for each problem. The tutor guide is a printed handout that includes the trigger text and the main issues which the problem is expected to generate for students. It provides an overall 'map' of the problem, divided into sections for each tutorial session. Sections contain headings, optional questions for discussion, and lists of suggested hypotheses, or patient history questions or treatment options, depending on the session. It also includes the titles of lectures and theme sessions that have been scheduled for the week. To enable tutors to 'play the patient', i.e. provide appropriate data in response to groups' inquiry plans in the second tutorial, staff receive a patient data sheet that contains information on the patient's background, history of presenting problem, results of a physical examination and any investigations. The patient data sheet also contains information about patient management and outcome.

Tutors meet once a week, after the second tutorial, with the Block Chair to review groups' progress with the current problem, and to preview the next case with a 'case coordinator' or problem writer, who highlights the key issues and emphasis of the case.

Tutor feedback form

In the final week of each block, students are sent a tutor feedback form via email. The form consists of 24 items with a five-point scale and three open response questions. After the block is completed, each tutor receives the collated results and students' comments from his or her PBL group. Tutors can compare their performance with summary statistics for the total group of tutors.

The 24 rating items are grouped into five sections and an overall satisfaction item. The sections or 'scales', entitled 'Tutor approach and style', 'Clinical reasoning process', 'Independent study', 'Group function' and 'Feedback', were originally based on coherence of the face validity of items. It was also thought that by grouping items, the form would be easier for students to complete. Items under 'Independent study' and 'Clinical reasoning process' were derived from steps in PBL and the clinical reasoning process, illustrated in Figure 1. Other items were derived from PBL tutor evaluation forms then in use at the University of Newcastle (New South Wales) and Flinders University of South Australia.

To test the construct validity of the feedback form we conducted a confirmatory factor analysis. We also sought to analyse the relationship between derived factors and tutors' experience and broad knowledge background. We speculated that support provided to tutors in the form of tutor guides, patient data sheets and weekly meetings may help to reduce any possible variance in tutor effectiveness based on broad background.

Method

Subjects were 223 tutors who had taught in blocks 1–9 in the University of Sydney medical program during the period 1998–2000. They included members of Faculty, doctoral students enrolled in the Faculty and one medical program student.

We used the Statistical Package for the Social Sciences (SPSS) for Windows, Release 10.0.5, to calculate reliability estimates for each scale on the feedback form ('Tutor approach and style', 'Clinical reasoning process', 'Independent study', 'Group function', 'Feedback').

Confirmatory factor analysis (CFA) was carried out using the LISREL mainframe computer program (LISREL V8.30, Joreskog & Sorbom, 2000). We performed CFAs on the data for each block of the medical program across the three years 1998–2000. The overall satisfaction item was excluded from analyses, because a single item cannot be used to define a construct (e.g. Mueller, 1996). For the purposes of analysis, we grouped the 23 individual items in pairs or 'trios' i.e. *composites*. This strategy offers three advantages: (1) it creates a high ratio between the number of subjects and the number of variables, (2) it produces more valid and reliable indicators as well as making the multivariate normality assumption more likely to hold, and (3) it reduces the effects of idiosyncrasies in the wordings of individual items (Vispoel, 1995; Marsh *et al.*, 1998).

The overall fit of a factor model with the data was assessed with the ratio of chi-square to degrees of freedom test (χ^2/df), the non-normed fit index (NNFI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). The principal indices (e.g. NNFI, CFI) have values ranging from 0, reflecting a poor fit, to 1, indicating a perfect fit (Bentler, 1990). The RMSEA is often used as it is relatively insensitive to sample size, with values below 0.10 considered as 'good' and below 0.05 as 'very good'. For the χ^2/df ratio, a value less than 2.0 is usually indicative of a well-fitted model (e.g. Bollen, 1989; Mueller, 1996; Loehlin,

Table 1. Indices of goodness-of-fit, first-order factor models (1, 2, and 3) for each block.

	Model	χ^2:df	NNFI	CFI	RMSEA
Block 1 (n = 94)[a]	1	461.08: 54	0.81	0.84	0.16
	2	299.51: 53	0.88	0.90	0.12
	3	196.71: 51	0.93	0.94	0.09
Block 2 (n = 168)	1	325.68: 54	0.77	0.81	0.19
	2	231.46: 53	0.85	0.88	0.16
	3	169.87: 51	0.90	0.92	0.13
Block 3 (n = 123)	1	268.14: 54	0.75	0.80	0.20
	2	224.19: 53	0.80	0.84	0.17
	3	168.63: 51	0.85	0.89	0.12
Block 4 (n = 348)	1	441.18: 54	0.87	0.89	0.17
	2	237.03: 53	0.94	0.95	0.11
	3	145.75: 51	0.97	0.97	0.07
Block 5 (n = 220)	1	334.03: 54	0.82	0.86	0.16
	2	238.13: 53	0.88	0.91	0.13
	3	144.73: 51	0.94	0.95	0.09
Block 6 (n = 332)	1	417.48: 54	0.85	0.87	0.16
	2	221.61: 53	0.93	0.94	0.10
	3	144.63: 51	0.96	0.97	0.07
Block 7 (n = 130)	1	223.83: 54	0.87	0.89	0.17
	2	167.06: 53	0.91	0.93	0.13
	3	125.52: 51	0.94	0.95	0.11
Block 8 (n = 303)	1	494.23: 54	0.80	0.84	0.19
	2	300.05: 53	0.89	0.91	0.14
	3	218.65: 51	0.92	0.94	0.11
Block 9 (n = 79)	1	186.10: 54	0.66	0.73	0.20
	2	108.92: 53	0.85	0.88	0.11
	3	94.56: 51	0.88	0.91	0.09

Note: [a]n refers to number of student responses.

1998). We tested differences in fit between factor models using direct tests of chi-square differences (Hoyle & Panter, 1995).

The number of times each tutor had taught in the same and/or different blocks across the years 1998–2000 was used as the measure of tutor experience. Tutor background was classified using the following three categories: (1) medical training, (i.e. 'expert') and (2) basic science and (3) non-basic science (taken together, 'non-expert'). Tutors were assigned to each category on the basis of their academic qualifications. Non-basic science tutors included academics with backgrounds in nursing, public health, education and psychology.

We used General Linear Modeling, which consists of univariate analysis to examine the relationships between tutor experience, broad background and gender, the overall satisfaction score and any derived factors. For the purposes of analysis, we used group average scores on the 24 rating items for individual tutors in each block.

Results

All five scales on the feedback form and the total score displayed acceptable levels of reliability across the nine blocks of the medical program. Coefficient alpha estimates of reliability ranged from 0.70 to 0.96 for the scales, and 0.92 to 0.97 for the total score.

Results for CFAs conducted for each block are summarized in Table 1. CFAs performed for a single first-order factor model, where all measured items were specified on a single factor (Factor 1), indicated a weak model fit (e.g. χ^2/df = 8.54, NNFI = 0.81, CFI = 0.84, for block 1, 2, etc.). A two first-order factor model, where measured items from the 'Clinical reasoning process' and 'Independent study' scales loaded on Factor 1, and items from the 'Tutor approach and style', 'Group function' and 'Feedback' scales loaded on Factor 2, showed an increase in model fit (e.g. χ^2/df = 5.65, NNFI = 0.88, CFI = 0.90, for block 1, 2, etc.), but still failed to fit the data adequately. With a three first-order factor model, the various goodness-of-fit indices indicated a well-fitted model (e.g. χ^2/df = 3.86, NNFI = 0.93, CFI = 0.94, for block 1, 2, etc.).

In the three-factor model, items from the 'Tutor approach and style', 'Group function' and 'Feedback' scales loaded on Factor 1, items from the 'Clinical reasoning process' scale loaded on Factor 2, and items from the 'Independent study' scale loaded on Factor 3. The fit of model 3 differed significantly from the fit of models 1 and 2 ($\chi^2 = 264.37, p < 0.05$ between models 3 and 1, and $\chi^2 = 102.80, p < 0.05$ between models 3 and 2).

Tutor experience and background

The maximum number of times a person had tutored in the same and/or different blocks for the period 1998–2000 was 6. General Linear Modeling revealed that tutor experience was significantly related to all factors, i.e. staff who had tutored more often were rated better on all factors.

Gender was significantly related to Factor 1, $F(1, 365) = 6.73, p < 0.05$, with female tutors scoring lower (i.e. better) on Factor 1 than male tutors. Tutor background was significantly related to Factor 1 and Factor 2 (Clinical reasoning process), with non-basic science tutors scoring lower on Factor 1 than basic science tutors, $F(3, 365) = 2.50, p < 0.05$, and medically trained tutors scoring lower on Clinical reasoning process than basic science tutors, $F(3, 365) = 3.01$, $p < 0.05$. There was no difference between medically trained and non-basic science tutors on Clinical reasoning process.

On the overall satisfaction item, both medically trained and non-basic science tutors were rated lower (i.e. better) than basic science tutors, $F(3, 365) = 3.22, p < 0.05$. Means for gender and background for each factor are shown in Table 2, where a mean score of 1 represents best tutorial teaching.

Discussion

Our results show that a three first-order factor model offers the best fit for student evaluation data on PBL tutorial teaching in Years 1 and 2 of the University of Sydney medical program. The three factors load reliably and consistently from block to block regardless of block length or content. Although the original scales, 'Clinical reasoning process' and 'Independent study', on the tutor feedback form were confirmed as Factors 2 and 3 respectively, the third factor identified included items from three original scales. Specifically, most of these items addressed group qualities, including a group's climate or atmosphere, group members' level of

Table 2. Means and standard deviations for tutor background and gender for each derived factor and overall satisfaction score.

		Factor 1		Factor 2 (Clinical reasoning)		Factor 3 (Independent study)		Overall satisfaction	
Background	n^a	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Expert									
Medically trained	89	1.92	0.47	1.69^b	0.48	1.82	0.49	1.63^b	0.61
Non-expert									
Basic science	191	1.95	0.53	1.83	0.55	1.92	0.52	1.76	0.73
Non-basic science	21	1.67^b	0.40	1.72	0.34	1.82	0.31	1.44^b	0.43
Gender									
Female	139	1.79^b	0.44	1.78	0.51	1.79	0.45	1.61	0.62
Male	162	2.04	0.54	1.78	0.53	1.96	0.53	1.78	0.72

Notes: [a]Tutors may be counted more than once in *n* according to the number of times they have tutored; $^b p < 0.005$. A mean score of 1 is best, 5 is worst.

involvement, focus and level of teamwork, and a tutors' commitment to group and individual development. We interpreted Factor 1 to be a tutor's overall approach to facilitating the development of effective group process.

We examined the loadings for the composite 'Teacher approach and style 1' (TAS1) and on the basis that it consistently showed loadings between 0.40 and 0.70 (loadings less than 0.60 are considered poor), we deleted the two TAS1 items from Factor 1 and created the scale 'Group process' on the revised tutor feedback form, which is shown in Appendix A. This revised form can be used as a valid and reliable measure of the quality of PBL tutorial teaching in problem-based programs similar in curricular structure to the University of Sydney medical program.

Our results also show that in Years 1 and 2 of the medical program, there are significant effects related to gender and PBL tutors' broad background on facilitation of group process and clinical reasoning process. Students rate female tutors as better at facilitation of group process, while both medically trained tutors (experts) and non-basic science tutors (non-experts) are rated as better than basic science tutors (non-experts) at facilitation of clinical reasoning process and group process. Overall, tutors who have a non-basic science background are rated just as highly on their facilitation of students' learning as tutors who have medical training. Basic science tutors may tend to rely more on their own 'discipline expertise' instead of following the tutor guide, and/or may tend to be less student-centred than non-basic science tutors. The extent to which tutor guides, patient data sheets and weekly preview meetings help non-basic science tutors to achieve an effective standard of facilitation, regardless of the content of a block, requires further examination.

In their review of research on factors that effect students' learning in PBL tutorials, Schmidt & Moust (2000) conclude that effective tutors possess three key qualities: appropriate knowledge about the topic or case in question, an empathic attitude toward students' small-group learning, and an ability to express themselves at students' levels of knowledge. It may be that in our study some basic science tutors who are more comfortable with traditional, didactic approaches to teaching are less empathic in their attitudes and/or less able to express themselves at students' levels of knowledge in comparison with other tutors, particularly those staff who have had greater experience with small-group teaching.

In conclusion, it should be noted that although we found differences in the quality of tutoring between medically trained, non-basic science and basic science tutors, overall, the standard of PBL tutor facilitation of group process and clinical reasoning in Years 1 and 2 is rated very highly by students (see Table 2). A limitation of this study is that tutor effectiveness is determined by students' ratings, whereas it would be interesting to also examine tutor effectiveness using students' performance at assessment. However this would be difficult, because students may have up to nine different tutors over the course of Years 1 and 2.

The more often a tutor has taught, the more highly he or she is rated by students on all aspects of tutor effectiveness. We cannot be certain whether this is due solely to a 'practice effect', i.e. frequent experience, or is a result of other intrinsic tutor factors, such as a high level of commitment to facilitating group process, high self efficacy and/or a student-centred theory of teaching. Further research is needed to determine the relationships between potential intrinsic factors, and frequency and quality of tutoring.

Notes on contributors

GRAHAM D. HENDRY is a Lecturer in Medical Education at the University of Sydney. In the last several years he has conducted research in constructivist learning theory, and coordinated the design and implementation of a multi-level evaluation system for the University of Sydney Medical Program.

HUY PHAN is a PhD student in the School of Learning and Development at the University of Sydney. He is currently in the final stage of his doctoral research, titled 'Self-efficacy in academic settings: a developmental approach'.

PATRICIA M. LYON is a Lecturer in the Department of Surgery at the University of Sydney. Research interests include teaching and evaluation in surgical education, with particular reference to the operating theatre, staff development and the use of qualitative methods in medical education.

References

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

BOLLEN, A.K. (1989) *Structural Equations with Latent Variables* (New York, Wiley).

HOYLE, R.H. & PANTER, A.T. (1995) Writing about structural equation models, in: R.H. Hoyle (Ed.) *Structural Equation Modeling: Concepts, Issues and Applications*, pp. 158–176 (Thousand Oaks, CA, Sage).

JORESKOG, K.C. & SORBOM, D. (2000) *LISREL 8.30*, Student Edition, February 2000 (Mooresville, IN, Scientific Software).

LOEHLIN, J. C. (1998) *Latent Variable Models: An Introduction to Factor, Path and Structural Analysis*, 3rd edn (Mahwah, NJ, Lawrence Erlbaum).

MARSH, H.W., CRAVEN, R. & DEBUS, R. (1998). Structure, stability, and development of young children's self-concepts: a multicohort–multioccasion study, *Child Development*, 69(4), pp. 1030–1053.

MUELLER, R.O. (1996) *Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS* (New York, Springer).

NEAME, R.L.B. (1989). Problem based medical education: the Newcastle approach, in: H.G. Schmidt, M. Lipkin, M.W. De Vries, & J.M. Greep (Eds) *New Directions in Medical Education*, pp. 112–146 (New York, Springer-Verlag).

SCHMIDT, H.G. & MOUST, J.H.C. (2000) Factors effecting small-group tutorial learning: a review of research, in: D.H. Evensen & C.E. Hmelo (Eds) *Problem-Based Learning: A Research Perspective on Learning Interactions*, pp. 19–51 (Mahwah, NJ, Lawrence Erlbaum).

VISPOEL, W.P. (1994) Self-concept in artistic domains: an extension of Shavelson, Hubner, and Stanton (1976) model, *Journal of Educational Psychology*, 87, pp. 134–153.

Appendix A: Revised University of Sydney Medical Program PBL tutor feedback form.

PBL Tutor Feedback Form

Tutor name:	Please choose a rating for each question.	
PBL Group ID:	General Ratings	Overall Rating (Last Item)
	SA – Strongly Agree	A – Excellent
Block Number:	Agree	
	A – Agree	B
	U – Undecided	C – Satisfactory
	D – Disagree	D
	SD – Strongly Disagree	E – Poor
	Disagree	

Group process
Our tutor . . .

1. appeared to be enthusiastic about tutoring
2. did not dominate group discussion
3. created a supportive group climate
4. showed concern with progress of individuals
5. invited constructive feedback on his/her performance
6. encouraged involvement of group members
7. kept the group focused on the task
8. encouraged us to reflect on and evaluate how well the group worked together as a team
9. addressed group problems when they arose
10. gave the group feedback on performance
11. gave me feedback on my performance when I asked

Clinical reasoning process
Our tutor encouraged us to . . .

12. identify the relevant clues in the problem presentation
13. hypothesize logically and broadly
14. ask for patient information required to test hypotheses
15. say how patient information distinguished between hypotheses
16. summarize and restate the problem as we proceed through the problem
17. make a diagnostic decision based on probabilities
18. think logically and broadly about planning patient management

Independent study
Our tutor encouraged us to . . .

19. identify what we needed to find out more about in relation to the problem
20. seek out appropriate learning resources
21. communicate effectively and efficiently to the group what we learned on our own

Overall

22. All things considered, how would you rate your tutor?

What aspect of your tutor's teaching has been *most useful* for your learning? Please give your reasons:

What aspect of your tutor's teaching has been *least useful* for your learning? Please give your reasons:

What suggestions do you have that would assist your tutor in his/her tutorial teaching?