

Assessment in Problem-based Learning: The Role of the Tutor

A E Dodds,*^{MA (Melb)}, R H Osmond,**^{BA Hons (Melb)}**, S L Elliott,***^{MD, MBBS (Melb), FRACP}

Abstract

Introduction: Problem-based learning (PBL) in medicine emphasises the tutor's role in facilitating collaborative and integrated learning. While it is widely recognised that tutors come to know their students, they traditionally play little part in the formal assessment process. We introduced tutor assessment into a problem-based medical curriculum and examined the patterns of tutor marks in four subjects, in terms of within semester consistency and their relation to other forms of assessment in the course. **Materials and Method:** Seventy-four tutors assessed 187 students twice (formative assessment in mid-semester, summative assessment at the end of semester) in each semester, using an assessment tool that focussed on communication and learning skills. Tutor marks were examined using correlations, cluster analysis and chi-square coefficients. **Results:** There were consistent trends in the patterns of cluster membership for pairs of marks in high, average and low clusters from mid-semester formative to end-of-semester summative assessment. Cluster membership for pairs of marks was predominantly consistent. Most movement was from lower to higher clusters over the two assessments, reflecting student improvement after tutors' feedback and tutor responsiveness to feedback to them. There was little variance in the marks of different tutors. Tutor marks correlated moderately and significantly with other forms of assessment in each subject. **Conclusion:** Tutor marks contribute useful, distinctive dimensions to assessment in a problem-based medical curriculum.

Ann Acad Med Singapore 2001; 30:366-70

Key words: Collaborative learning, Marking consistency, Medical curriculum

Introduction

The tutor plays a key role in fostering the collaborative and self-directed learning that is central to problem-based learning (PBL).¹⁻⁵ It is reasonable, therefore, for the tutor to be involved in assessing the processes of collaborative learning. Tutors are in the best position to know how well their students work together and how well they demonstrate understanding and integration. Surprisingly, most PBL assessment procedures seem to rely on terminal examination marks, thus ignoring the processes that are endemic to PBL learning. The aim of our study was to investigate the value of PBL tutors' marks as part of the overall assessment strategy in a new medical curriculum in which formative assessment was used alongside summative assessment. What do analyses of tutor marks reveal about tutor monitoring of learning, and how do tutor marks relate to other forms of assessment?

The few reported studies of tutor assessment raise a number of issues. Kaufman and Hansell,⁶ for example, found only a poor relationship between tutor assessment

and student examination performance. Most (87%) of their 31 tutors could not predict students' examination performance, and tutor ratings of knowledge acquisition and integration were unrelated to students' examination marks. These findings seem at odds with the rhetoric of relying on tutors to facilitate and monitor student learning. Monitoring and assessing processes are conceptually related, and ought to be related in practice. Des Marchais and Vu⁷ go further, arguing that tutor ratings are psychometrically questionable, because of their inability to discriminate between higher performing students. Their own complex scheme, however, may have contributed to the lack of discrimination, since their tutors had to rate each student on 44 behaviours. We believed it was possible to give tutors a simpler, PBL appropriate set of marking criteria and tasks, and then to examine their contribution to overall assessment in conjunction with other marks.

A new PBL-based medical course presented an ideal opportunity to incorporate tutor marks into the assessment system. Teaching and learning is organised around the

* Senior Lecturer in Medical Education

** Graduate Research Assistant

*** Director

Faculty Education Unit

Faculty of Medicine, Dentistry and Health Sciences

The University of Melbourne, Australia

Address for Reprints: Agnes Dodds, Faculty Education Unit, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Victoria, 3010, Australia.

E-mail: agnesed@unimelb.edu.au.

body systems in a 14-week semesters, with PBL tutorials supported by lectures and practical classes. Ten students in each PBL class attend two-hour tutorials twice a week, giving tutors 56 contact hours with their students.

Our curriculum developers planned an assessment system congruent with PBL philosophy of integrated and collaborative learning. At the same time, they had to meet the university's requirement of individual graded assessment. In this context, it would have been inefficient as well as counter to PBL principles to ignore the tutor's privileged access to student performance over the semester. In order to assess a range of competencies, knowledge and group skills, our set of assessment instruments includes tutor marking of tutorial performance as well as written examinations (including multiple choice questions, short answer questions and case analyses).

The inclusion of tutor assessment in the overall grading scheme, nevertheless, was a contentious issue for some curriculum planners. Some claimed that giving tutors the responsibility of assessing performance would not only lead to conflict between their assessor and facilitator roles, but also would promote inequity through the variability of marks across tutors. While this tension between teaching and assessing exists for any good teacher, we were interested in systematically examining any variability in tutor assessment that would reflect conflicting interests. Consequently, our emphasis on the dual tutor functions of facilitation and assessment made us careful in selecting and training the 21 to 24 tutors recruited each semester. They mostly were academics from basic science departments of the faculty, with fewer than 15% clinically qualified.

PBL tutoring is taken seriously. Although more than half the tutors were experienced teachers, all were required to participate in a pre-semester two-day workshop including training in assessment philosophy and procedures. During the semester, weekly tutor meetings also involved briefing on assessment techniques.

The PBL Tutor Assessment Scheme

Typically, across the curriculum, tutors assess student performance over the whole semester using a written assessment tool that is shown as Appendix I. This instrument covers two areas of: (1) communication skills, and (2) learning skills and knowledge building. Tutors give each student a 0 to 10 mark for each area, using a set of indicators that assist them to construct two composite scores. They add written comments on strengths and areas for the student's attention.

The instrument is used twice each semester. In a mid-semester interview with individual students, the tutor allocates marks and records comments on the form that is handed back to the student. These marks are formative

indicators of student progress and do not count towards final assessment. Subsequently, these tutors are given feedback on the full range of tutor marks and PBL group means in a meeting focused on maintaining marker consistency. At the end of semester, the instrument is used again, but without the interview. The second set of marks becomes 10% of the student's overall mark for the course. We can ask how variable are the products of this tutor assessment scheme, and how well the tutor's contribution relates to other, more traditional, forms of assessment.

Materials and Method

Participants

Seventy-four tutors made two assessments of the performance of 187 students of 195 (107 females, 88 males) in four subjects of the Bachelor Medicine/Bachelor Surgery course. Students were randomly allocated to tutorial groups, stratified for gender. Numbers varied slightly between analyses as some students withdrew or returned from leave of absence.

Method

For our analyses of relationships between formative and summative tutor marks for four subjects, consistency across tutors, and the relation of tutor marks to other pieces of summative assessment, we report descriptive statistics on the original variables. Calculations were made on transformations of several negatively skewed variables. We used Ward's method of hierarchical cluster analysis with a squared Euclidean distance measure⁸ to categorise tutors' mid-semester formative marks and end-of-semester

APPENDIX I: PBL TUTORIAL ASSESSMENT FORM (EXTRACT)

Communication Skills

1. Interacts appropriately with other group members
2. Demonstrates presentation skills
3. Contributes knowledge and resources to group discussion
4. Participates regularly in a variety of roles
5. Provides constructive feedback to others
6. Facilitates the group process by follow-up, extension, reframing

Marks/10

Learning Skills and Knowledge Building

1. Generates relevant hypotheses
2. Identifies key learning issues
3. Evaluates hypotheses with relevant evidence
4. Demonstrates understanding of concepts and mechanisms
5. Demonstrates ability to locate and appraise information
6. Demonstrates ability to apply relevant information to the case

Marks/10

Comments

(Including: (1) strengths and (2) areas needing attention)

TABLE I: SPEARMAN'S CORRELATIONS BETWEEN FORMATIVE AND SUMMATIVE TUTOR MARKS, WITH DESCRIPTIVE STATISTICS

	PBS form.	PBS sum.	NDM form.	NDM sum.	CRL form.	CRL sum.	CGD form.	CGD sum.
PBS form.		.71	.38	.52	.41	.40	.37	.38
PBS sum.			.41	.50	.44	.48	.42	.42
NDM form.				.68	.37	.47	.42	.41
NDM sum.					.53	.57	.52	.54
CRL form.						.76	.51	.53
CRL sum.							.54	.60
CGD form.								.79
Mean	12.60	14.46	13.95	15.10	14.29	15.03	14.53	14.88
SD	3.20	2.96	3.23	2.69	2.67	2.67	2.40	2.71
Minimum	4.00	4.00	3.00	4.00	4.00	6.00	6.00	3.00
Maximum	20.00	20.00	20.00	20.00	19.00	20.00	19.50	19.50

PBS: Principles of Biomedical Science; NDM: Nutrition, Digestion and Metabolism, CRL: Cardio-Respiratory and Locomotor;

CGD: Control Systems Growth and Development

All Spearman's correlations significant at 0.01 level (two-tailed).

TABLE II: PERCENTAGES OF FORMATIVE AND SUMMATIVE TUTOR MARKS IN THREE CLUSTER GROUPINGS FOR FOUR SUBJECTS

Formative tutor mark (Mid-semester)	Cluster	Summative tutor mark (End of semester)		
		low	average	high
(a) Principles of Biomedical Science (PBS, n = 185)				
	low	4.3	11.9	1.1
	average	1.1	27.6	25.9
	high	0	2.2	25.9 ^a
(b) Nutrition, Digestion and Metabolism (NDM, n = 178)				
	low	3.9	3.9	0.6
	average	4.5	38.8	16.3
	high	0	3.9	28.1 ^b
(c) Cardio-Respiratory and Locomotor (CRL, n = 176)				
	low	9.1	0.6	0
	average	5.7	31.3	5.1
	high	0	23.9	24.4 ^c
(d) Control Systems, Growth and Development (CGD, n = 187)				
	low	10.2	6.4	0.5
	average	1.6	35.8	12.8
	high	0	4.8	27.8 ^d

^a X^2 (4, n = 185) = 78.63, P = 0.000;

^b X^2 (4, n = 178) = 88.49, P = 0.000;

^c X^2 (4, n = 176) = 123.20, P = 0.000;

^d X^2 (4, n = 187) = 146.18, P = 0.000.

summative marks into three high, average, and low clusters. Then we examined consistencies and movements between the three cluster groups for pairs of marks for each student, using chi-square analyses. This procedure allowed us to examine sensitively consistency or inconsistency in tutor marks across three levels identified from the distributions for each subject.

Results

Relationship Between PBL Formative and Summative Tutor Marks

There were strong correlations between formative and summative marks within each subject, and these were larger than either form of mark across different subjects, as shown in Table I's correlations between formative and summative tutor marks for four subjects: Principles of Biomedical Science (PBS); Nutrition, Digestion and Metabolism (NDM), Cardio-Respiratory and Locomotor (CRL); Control Systems, Growth and Development (CGD). All Spearman's correlations were significant at the 0.01 level, and were not inflated, since one does not constitute a percentage of a correlated score.

The correlations in Table I suggest consistency, but not sameness. A series of repeated measures t -tests ($P \leq 0.005$) revealed that, for all four subjects, the students' assessment was consistently higher in the end-of-semester summative assessment than in the mid-semester formative assessment. Tutors had refined their approaches by the second half of semester.

Table II shows the percentages of tutor marks that consistently fell in high, average and low clusters between mid-semester formative to end-of-semester summative assessment in the four subjects, and the percentages that changed across clusters.

There were three consistent trends in the patterns of cluster membership for pairs of individual students' marks from the middle to the end of semester. Firstly, overall patterns of cluster membership were different between the two times and forms of assessment, as revealed by the significant chi-square coefficients in Table II. Secondly, in each subject, the cluster membership of pairs of marks was predominantly consistent. Thirdly, changes in cluster

membership mostly involved movements from lower to higher clusters over the two assessments. For example, 107 (58%) pairs of formative and summative marks were in the same cluster (high, average, low) for Principles of Biomedical Science (PBS). This was the first subject in which the PBL tutor scheme was used. For the other three subjects, percentages of marks in the same cluster were higher: (b) NDM, 70.8%; (c) CRL, 77.0%, and (d) CGD, 73.8%. Cluster movement in PBS involved 72 (38.9%) moves to a higher cluster (48 from average to high), but only six (3.2%) to a lower cluster. Similar changes upward were: (b) NDM, 21.0%; (c) CRL, 5.7%; and (d) CGD, 19.8%. Overall, these patterns indicate a reasonable level of consistency in tutors' pairs of marks for four subjects, with a general trend towards higher marking in the final (summative) assessment.

Tutor Consistency

Tutors were generally consistent, with little variance across their marks, as calculated using intra-class correlations. In the formative marks, the variance across tutors was small, with most variability in PBS (30.0%), followed by NDM (22.5%), CRL (12.2%) and CGD (3.3%). In the summative marks, there was far less variability: 12% in PBS, with no variability at all in NDM, CRL, or CGD. It seems that tutors effectively used the feedback on their

formative marks, so that their summative marks come closer together.

Relations between Tutor Marks and other Assessment Procedures

For the summative assessment, correlations revealed that the tutors were contributing in distinctive ways to students' overall assessment, without their marks being at variance with other marks. Table III shows the correlations with three other marks and with the total mark for the four subjects. All Spearman's correlations were significant at the 0.01 level. Here, correlations between summative marks and their associated subject totals are inflated, because the former is a percentage of the latter. Hence, the tutor mark that represents only 10% of the total has lower correlations than, for example, the PBL exam that represents 25%.

Most summative tutor marks were significantly related to other pieces of assessment in the same subject, with correlations in the 0.30 to 0.50 range, and only one lower than 0.30 (NDM, tutor mark, exam, $r = 0.14$, $P > 0.05$). Correlations of the tutor mark with the total mark were higher, especially for NDM (0.75) and CRL (0.61), as would be expected for a component mark. These component-total correlations are lower than those of the examination components, but in the same order. The tutor component is a consistent, modest contributor that does not distort overall results.

Discussion

Tutors of PBL groups are well placed to judge the development of collaborative and integrated learning in their students. They spend more time with students than any other teachers and observe first hand the group process, adaptation and integration that are corner stones of the PBL system. We found that tutors are able to incorporate student assessment into their role as facilitators and monitors of student learning. When provided with clear criteria and feedback on their initial marks, tutors are able to achieve considerable consistency. Most of their marks are consistent, or increase slightly over the second half of a semester. Now that general consistency has been demonstrated, we are in a position to investigate these clusters with more fine-grained analysis.

In the infrequent cases of large discrepancies (i.e., movement across more than one level in pairs of marks), tutors were asked if there were any exceptional circumstances or if the movement involved error. In most cases, shifts could be explained by changes in a student's health or personal circumstances. Even in the very first semester of the new curriculum (PBS), tutor variability was not contributing to differences in patterns of student marks. The variability in the formative marks was largely

TABLE III: SPEARMAN'S CORRELATIONS OF SUMMATIVE TUTOR MARKS WITH ASSESSMENT PROCEDURES AND TOTAL MARKS FOR FOUR SUBJECTS

Assessment procedure	Tutor mark	Total mark
(a) Principles of Biomedical Science (PBS)		
Exam Paper 1: Multiple Choice Questions	.42	.69
Exam Paper 1: Short Answer	.30	.73
Exam Paper 2: PBL Exam	.30	.65
Total Mark	.57	
(b) Nutrition, Digestion & Metabolism (NDM)		
Exam Paper 1: Multiple Choice Questions	.42	.70
Exam Paper 1: Short Answer	.44	.77
Exam Paper 2: PBL Exam	.14*	.44
Total Mark	.75	
(c) Cardio-Respiratory and Locomotor (CRL)		
Exam Paper 1: Multiple Choice Questions	.41	.85
Exam Paper 1: Short Answer	.48	.87
Exam Paper 2: PBL Exam	.40	.76
Total Mark	.61	
(d) Control Systems, Growth and Development (CGD)		
Exam Paper 1: Multiple Choice Questions	.51	.83
Exam Paper 1: Short Answer	.49	.83
Exam Paper 2: PBL Exam	.50	.73
Total Mark	.70	

*not significant, all other Spearman's correlations significant at 0.01 level.

eradicated after tutor's feedback and briefing on marking criteria. This mid-semester formative mark served two useful purposes. It gave students information on their progress, and gave tutors information on their own use of the marking criteria.

The tutor marks fit well into the assessment profile, and do not disadvantage students. The moderate and significant correlations with other assessment procedures suggest that the tutor mark is contributing a specific dimension to assessment, without skewing the overall focus. The tutor's influence on assessment was neither inflated nor lost.

REFERENCES

1. Barrows H S, Tamblyn R M. Problem-based learning, approaches to medical education. New York: Springer-Verlag, 1980:38-47.
2. Norman G R, Schmidt H D. The psychological basis of problem-based learning: a review of the evidence. *Acad Med* 1992; 67:557-65.
3. Schmidt H D. Foundations of problem-based learning: some explanatory notes. *Med Educ* 1993; 27:422-32.
4. Thomas R E. Problem-based learning: measurable outcomes. *Med Educ* 1997; 31:320-9.
5. Davis M H, Hardien R M. AMEE Medical Education Guide No. 15: Problem-based learning: a practical guide. *Med Teacher* 1991; 21: 130-40.
6. Kaufman D M, Hansell M M. Can non-expert PBL tutors predict their students' achievement? An exploratory study. *Acad Med* 1997; 72: 516-18.
7. Des Marchais J E, Vu N V. Developing and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. *Acad Med* 1996; 72:274-83.
8. Aldenderfer M S, Blashfield R K. Cluster analysis. Beverley Hills: Sage, 1984:16-22.