Adv Physiol Educ 29: 197–203, 2005;
doi:10.1152/advan.00066.2004.

Using student feedback to construct an assessment rubric for a concept map in physiology

Roger W. Moni,[1] Eileen Beswick,[1] and Karen B. Moni[2]

Schools of [1]Biomedical Sciences and [2]Education, The University of Queensland, Brisbane, Queensland, Australia

Submitted 30 November 2004; accepted in final form 25 May 2005

Moni, Roger W., Eileen Beswick, and Karen B. Moni. Using student feedback to construct an assessment rubric for a concept map in physiology. *Adv Physiol Educ* 29: 197–203, 2005; doi:10.1152/advan.00066.2004.—Constructing quality assessment rubrics can be challenging, especially when they are used for integrated, group-centered, applied learning. We describe a collaborative assessment task in which groups of second-year dentistry students developed a complex concept map. In groups of four, the students were given a written, simulated, medical history of a patient and required to construct a concept map illustrating relevant pathophysiological concepts and pharmacological interventions. This report describes a research project aimed at making educational goals of the task more explicit through investigating student and faculty member understandings of the criteria that might be used to assess the concept map. Information was gathered about the perceptions of students in relation to the learning goals associated with the task. These were compared with faculty member perceptions. The findings were used to develop an assessment rubric intended to be more accessible to learners. The new rubric used the language of both faculty members and students to more clearly represent expectations of each criterion and standard. This assessment rubric will be used in 2005 for the next phase of the project.

cardiovascular; criteria; dentistry; standards

A THOROUGH UNDERSTANDING OF PHYSIOLOGY underpins the clinical skills of dental practitioners. At The University of Queensland, the second-year cohort of dentistry students must complete the foundation course in which they learn physiological principles by integrating the discipline within applications they are likely to meet in their professional practices. Many of the learning and assessment tasks are achieved via small groups to foster collaborative working habits and deep learning that is meaningful and long lasting (4). Assessment of group activities has many advantages in terms of assessing how knowledge is constructed and is increasingly common in higher education institutions (2, 9). In these group activities, assessment grades can be negotiated among students and faculty members in various ways (11). For 2 years, we have used concept maps constructed by students working in groups to assess their knowledge of cardiovascular, respiratory, and renal physiology. In 2004, we made a study of this assessment practice.

Concept maps are advocated as graphical representations of the conceptual, relational, and hierarchical nature of knowledge (1). Explicit teaching of concept maps can facilitate meaningful learning and, when used for assessment, can reveal conceptual understanding (16). However, there remains considerable debate around their reliability as an assessment tool in higher education (13, 15). Concept maps are often marked using quantitative scoring schemes that may be highly variable or require sophisticated computational data mining programs, e.g., Pathfinder network analysis (18). The use of explicit criteria with descriptive standards provides an alternative approach to grading. Multiple criteria can be used with composite standards to create a grading rubric that enables competent assessors to make holistic, qualitative judgements (19). With the use of these assessment rubrics, conceptual knowledge can be measured in a meaningful, reproducible, and efficient manner (3). When complemented with exemplars, they are more meaningful to participants, simple to design, and draw upon the professional expertise of educators (14, 20). Criteria and standards are written using natural language descriptions. The choice of words in a rubric is important because it is through language that understandings of criteria are shared among students and faculty members (21).

The University of Queensland Assessment Policy and Practices has mandated the use of assessment rubrics with explicit criteria. Social context plays an important role in how teachers and students at universities negotiate and interpret assessment tasks (6, 17). Consequently, as part of this project, three faculty members collaborated to investigate the processes by which students come to understand criteria in a challenging concept mapping task. This report describes the first phase of an action learning project in which student and faculty member interpretations of assessment were gathered and used to develop verbal descriptions of criteria and standards for a revised assessment rubric to be used in the second phase of the project.

METHODS

Research Design

The present study was the initial, exploratory phase of a larger project to investigate assessment in dentistry. The aims were to explore the processes by which faculty members and students come to understand criteria used for assessment and to develop a revised assessment rubric. A case study approach was selected because the processes and characteristics investigated in this specific university course may be used to illuminate aspects of a common challenge for all faculty members, i.e., to design appropriate assessment rubrics (10).

The study was undertaken in three stages. The first stage investigated how faculty members understood and used the rubric to grade group concept maps. In the second stage, students responded to a survey about their understandings of the criteria. In the third stage, a revised assessment rubric was developed.

Address for reprint requests and other correspondence: R. W. Moni, School of Biomedical Sciences, The Univ. of Queensland, St. Lucia, 4072 Brisbane, Australia (e-mail: r.moni@uq.edu.au).

Research Questions

Three research questions guided the data collection.

Question one. How do faculty members understand and use the assessment rubric of a concept mapping task?

Question two. How do students understand the assessment rubric of a concept mapping task?

Question three. How can the language of faculty members and students be used to develop a more appropriate assessment rubric?

The Course and Students

DENT2012 (Foundation Biological Sciences for Dentistry) is a 1-yr compulsory course to introduce students to fundamental biological concepts that underpin the practice of dentistry. The students were enrolled in the second year of an undergraduate Bachelor of Dentistry program. To progress to the second year, they had to achieve a grade point average of 6.2 from a maximum of 7.

The relevant learning goals were to *1*) become familiar with the biological processes that underlie normal human function and *2*) gain an appreciation of common pathological conditions relevant to dentistry.

The "Cardiovascular, Respiratory, and Renal" unit of work comprised 4 wk of the year-long course. Content was delivered in lecture format by four faculty members from the Physiology and Pharmacology departments. *Faculty members A* and *B* were tutors for this course. Key concepts included the control of blood pressure, hypertension, atherosclerosis, thrombosis, heart failure, and, as a minor component, the action of appropriate drugs.

Concept mapping constituted 6% of the total summative assessment. This complemented other assessment in the form of quizzes, practical reports, log books, problem-based exercises, multiple-choice questions, and a written final exam.

The Concept Map Assessment Task

This task was designed in 2003 to encourage a greater depth of learning and to assess basic concepts of physiology introduced in lectures. Students participated in a 50-min introductory workshop in which examples of concept maps on an unrelated topic were presented. They were then given guidelines to help them construct their own concept maps based on these examples.

Next, each group of students was given the same written, clinical scenario of a hypertensive patient exhibiting multiple risk factors and broad guidelines on inclusion of major physiological pathways and treatments for hypertension. Students were explicitly instructed to work collaboratively and were given 1 wk to complete the task, including a 2-h workshop in which all students discussed their work with faculty members. The task was completed by students in self-selected groups of four and submitted on large-format paper.

Marks out of 10 were allocated for comprehensive coverage (CC; 6 marks), logical sequence (LS; 2 marks), and overall presentation (OP; 2 marks). All members of a group received the same mark.

Student Surveys

A short survey was designed to explore what first-year dentistry students ($n = 62$) understood about the assessment rubric. There were three survey questions.

Question one. What things do you think are being assessed in this concept mapping task?

Question two. What things do you think should be assessed in this concept mapping task?

Question three. What do you think of this form of assessment task?

Before completing the survey, students were informed about the purposes of the study and advised that their participation was voluntary and anonymous; they could withdraw from the study at any time, and their decision would not affect their grades for the course.

Faculty Members

Three faculty members were involved with this project. Two were involved in teaching and assessing the task (*faculty members A* and *B*). The third, with more expertise in assessment research, acted as an educational advisor at each stage in the project.

In chronological sequence, *faculty members A* and *B* answered the survey questions for themselves, marked group concept maps separately using the original rubric (Fig. 1), moderated the final marks (moderation is the formal process by which faculty members negotiate consistency and agreement when grading using a rubric), and drafted the assessment rubric using the inputs of faculty members and students.

Coding Data

Student responses were coded using open, axial, and selective coding to identify themes (5). Each student survey response was typed verbatim into a Microsoft Excel file, with each separate concept written on a line (open coding). These "response elements" were considered to be the units of analysis. Using the "FIND" function, key words were counted and recorded to construct a frequency table. Words with similar stems were also searched, and, if the implied meaning was consistent, they were then tallied with the parent word, e.g., "logic," "logical," and "logically" were grouped together (axial coding). Themes were then ranked in order of frequency. Comments that were recorded only once or twice were then either collapsed into larger categories if they expressed similar ideas or formed new categories if they did not. This process (selective coding) was repeated until all the response elements were accounted for. The number of final categories was not prespecified.

Interrelator Reliability

Each faculty member was trained to independently code and categorize each response from the student survey. Interrelator reliability (IRR) was calculated as a correlation coefficient for each of four emergent categories. These values ranged from 0.79 to 1.00. The average correlation coefficient (as a correlation for agreement across all categories) was 0.88.

DENT2012 - Assessment for Cardiovascular Respiratory Renal Block, 11 June 2004.
This assessment task is worth 6% of your total marks.

Groups of 4 students should work together to produce a concept map of aspects of hypertension and its treatment.

A concept map is a visual representation of a body of knowledge organised so that a number of points of interest (concepts) are linked to show their relationship with one another. It is a useful learning tool to link your new knowledge with what you already know.

Consider a patient, Cheryl, aged 50, who has a resting blood pressure of 150/100. Her history shows that she is 30% above her ideal weight, smokes 30 cigarettes a day, has asthma, and does not exercise. You determine that the most likely cause of her hypertension is increased peripheral resistance associated with an increased blood volume.

Draw a simple concept map relating the following mechanisms. Fill in as many intermediate steps as you can, while ensuring that your map does not become unintelligible. You are not required to develop a model of how these risk factors contribute to hypertension, but may need to consider them when developing an overall treatment plan.

- hypertensive patient
- how increased TPR and increased blood volume could cause hypertension
- compensatory mechanisms in the CVS
- compensatory mechanisms in the kidney
- where drugs or other treatment might be used to reduce the hypertension
- where drugs might be contraindicated due to side effects on the respiratory system.

The concept map should include the major pathways that are involved in control of blood pressure. Superimposed on the normal pathways, you should indicate where changes have taken place, either to produce hypertension or as a compensatory mechanism. You could give some consideration to the time course of the changes, eg seconds, minutes, hours, days, years. The third layer should be drug and other treatments which can reduce blood pressure.

The number of steps you present in each pathway is a matter of judgement but it is important to keep in mind the overall presentation. Too many points will make the presentation untidy and difficult to comprehend. Too few will suggest that the student has not developed a deep understanding. Concepts should be linked where possible with words such as promotes, initiates, inhibits, contains, triggers, is necessary for, and transforms. The points should be presented in the correct sequence.

The overall presentation should be easy to understand. One or possibly two arrows crossing other arrows may be permissible but more than this is unacceptable.

Marks will be awarded as follows:

comprehensive coverage	6 marks
points in logical sequence	2 marks
overall presentation	2 marks

Fig. 1. Assessment task description and rubric in 2004.

The Revised Assessment Rubric

New criteria were developed using the major categories derived from reflections of faculty members and the student survey responses. The revised statements of standards that described grades for the group concept maps incorporated the language of faculty members and students. This assessment rubric will be used for the first time in 2005 with the next cohort of students.

RESULTS

Faculty Member Reflections on Research Questions

Faculty members A and *B* closely agreed in their answers to *questions one* and *two*. CC was considered to include factually correct, major physiological concepts; specific pharmacological targets; and lifestyle interventions. LS included adhering closely and consistently to the use of taught symbolic notation (nodes and arrows) and the factual correctness of connections. OP included clarity, ease of understanding, balance of detail across sections, and a limit to the number of cross-arrows.

Because *faculty member B* was new to this task, reflective notes were made to assist the moderation process to make explicit, and thus reduce, inconsistencies of grading between the two faculty members.

Comprehensive coverage. For CC, the six marks were allotted to three subcategories based on the written task given to students. The subcategories were as follows: blood pressure (mean arterial pressure, total peripheral resistance, cardiac output, and blood volume, with each contributing ½ mark); adaptive changes (baroceptor, autonomic, short term, and long

Table 1. *Summary of group scores before and after moderation by faculty members*

Statistics	*Faculty member A*	*Faculty member B*	Moderated
Average ± SD	7.97 ± 1.22	9.00 ± 0.71	8.25 ± 0.84
Median	8.25	9.25	8.50
Range	9.5–6 = 3.50	10–7.5 = 2.50	9.5–7.0 = 2.50

For statistics, $n = 16$ groups; scores were out of 10 marks. Range shows the upper score minus the lower score.

term, with each contributing ½ mark); and treatments (pharmacological, with 1 mark for accurate indications and ½ mark for the specific contraindications for asthma and life-style changes contributing 1 mark).

Logical sequence. For LS, three subcategories were generated. These were as follows: directionality (½ mark), "did the propositions make sense?" (½ mark), and the factual correctness of connectivity (1 mark).

Overall presentation. For OP, after a few maps were examined, four subcategories (½ mark to each) were generated. These were as follows: format (nodes within closed figures, line kinks, consistency of style, most content was confined to concept boxes); spelling (2 errors were permitted, conventional abbreviations were acceptable); legibility (readable script size, contrast of shades or colors); and clarity (graphic spacing, design, clutter, and neatness).

Selecting subcategories involved continuously making decisions by reflecting on aspects of the criteria that were implicit, examining student work and then judging whether students' implicit interpretations matched my expectations. As questions regarding the interpretation of criteria were proposed, new subcategories were defined and then contrasted with progressively formed definitions of concepts about criteria.

Student Group Marks Before and After Moderation

Group marks were moderated by both *faculty members A* and *B* (Table 1). A high academic standard of student work was evident. An example of the concept map that was scored as 9/10 marks is presented in Fig. 2.

Notes During Moderation by a Faculty member

Moderation (lasting 3.5 h) began with a comparion of group marks, followed by a discussion of grading rationales and then negotiation of the final group marks. The salient issues arising in discussions are listed below.

- There was close agreement in the final group marks, with overlapping mean scores given by *faculty members A* and *B*.
- For some maps, there were much larger differences in marks. These reflected differences in faculty member dispositions to grading severity and were an amalgam of threshold knowledge (*faculty member A* being an experienced physiologist and *faculty member B* being a pharmacologist), implicit knowledge of grading such tasks, and interpretations of grading schemes (refer to Table 1, in which marks allocated by *faculty member B* decreased as a result of the moderation process; in contrast, *faculty member A* adjusted the median score by only 0.25 marks).
- There was frequent use of mark deductions, e.g., one mark was removed for a spelling error in the title.

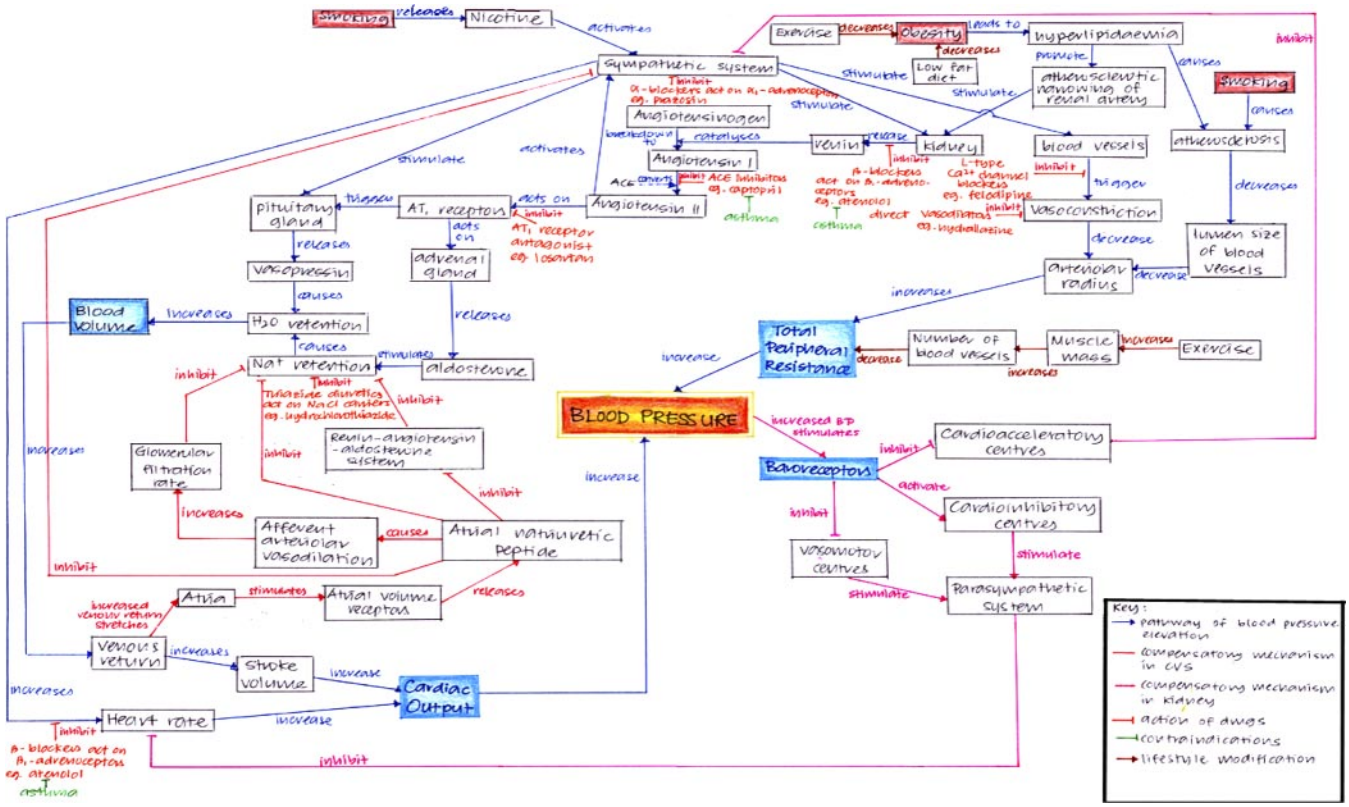

Fig. 2. Example of a concept map constructed by one group of students. TPR, total peripheral resistance; CVS, cardiovascular system.

• Consensus on what constituted "top" (10/10) and "lowest" (6/10) maps was quickly reached. These acted as reference marks and defined a grading range of 6–10 marks.

• There were occasional inconsistencies of marks allocated to different criteria.

• As the criteria used in the original rubric were not consistently clear, "trade offs" had to be negotiated, e.g."insightful" additions should be rewarded and could partly compensate for some omissions on other areas of the map. This was recognized as problematic by the authors and needed to be addressed in the revised rubric.

• There was a need to have more information about what was presented to the students in lectures. It should be noted that in this unit of the course, the authors were not involved in presenting course content. The information made available to the authors about lecture content was general and not specific and detailed.

• While the task was considered to be useful, in the future it needed to be more specific to the needs of dentistry.

• It became clear that faculty member understandings about standards of criteria were greatly enhanced by discussion around exemplars.

Student Survey Results

In total, forty-five (73%) survey forms were returned (Table 2). However, four surveys were returned as groups and not individually as requested. Given that most groups had four students, it can be assumed that all or nearly all students responded to the survey. Typically, each returned survey had two to three comments per question; therefore, the number of response elements was greater than the number of survey forms returned for each question.

For *question one*, 1 of 45 surveys had answers that were either irrelevant or incomprehensible, whereas 1 of 45 answered "I don't know." For *question two*, no answers were submitted from 6 of 45 surveys. For *question three*, no answers were submitted from 2 of 45 surveys.

Questions one and two. The final categories are listed in decreasing order of frequency using student language (Fig. 3). For *question one*, a fifth category of "Working in Groups to Research and Communicate" (5% of responses) was considered to be assessable only as a process but not as a product. Therefore, it was included in "Presentation." For *question two*, "Group Communication" (1.8% of responses) was included in "Presentation." A significantly large category, "Assessing the Block with Criteria" (13% of responses), was considered as addressing the purpose of the survey and not referring to assessment. Therefore, it was omitted when the criteria were developed.

Table 2. *Number of response elements for student survey questions*

Research Question Number	Number of Response Elements	Number of submissions (%Response)
One	147	43/45(96)
Two	123	39/45(87)
Three	93	43/45(96)

Question one was "What things do you think are being assessed in this task?" *Question two* was "What things do you think should be assessed in this task?" *Question three* was "What do you think about this form of assessment?" Survey response elements (separate ideas contributed by students) were the units of analysis.

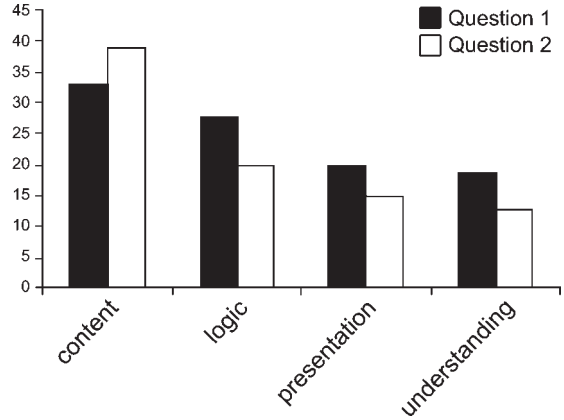

Fig. 3. Student responses to survey *questions one* and *two* on the assessment of concept maps. *Question one* was "What things do you think are being assessed in this task?" *Question two* was "What things do you think should be assessed in this task?"

The IRR was 0.87 and 0.96 for *questions one* and *two*, respectively. These were derived from IRR for specific categories ranging from 0.79 to 1.00. The negotiation of categories and frequencies of responses was minimal.

Question three. Most of the students responded to the concept map in favorable terms (Table 3); 2 of 45 submissions recorded no answer, whereas 1 of 45 submissions was undecided.

The following are typical of the negative comments: it was too time consuming, the process of "design" was inappropriate, students wanted clearer guidelines, they had difficulty in determining the detail and depth required, and they felt that the task did not test depth of understanding.

Summary of Student Feedback

Across *questions one* and *two*, the two most important themes were Content and Logic. Presentation was also important to the students but clearly less so. An unexpected finding was the relatively low priority placed on Understanding for both questions. Student responses to *question three* indicated that, overall, they approved of the task. However, there were some challenges evident. These included the time constraints (especially with the task close to semester block exams). In addition, students perceived that they needed more structured

Table 3. *Summary of student responses to question three*

Category		Examples of Student Answers
Approval	42%	• *Concept maps help put everything together linking multiple concepts* • *It's fun and challenging*
Qualified approval	40%	• *Challenging . . . but a good alternative learning method* • *A bit time consuming . . . but it's worth the effort*
Did not like the exercise	11%	• *Not very effective . . . 90% of the remaining time tweaking and perfecting [the concept map] for assessment was a complete stupid waste of time*

Question three was "What do you think about this form of assessment?"

guidelines to assist them in completing the task more efficiently.

The Revised Assessment Rubric

Faculty member reflections and student feedback were integrated to produce the assessment rubric (Fig. 4). The order of criteria was altered from Table 3 to reflect the opinions of faculty members that Presentation was the least important criterion. In addition, faculty members reasoned that the criterion of Understanding could only be represented through the logic implied by the particular use of arrows and active linking verbs. Understanding was therefore included within the criterion of Logic. The criteria were equally weighted with five provisional grades designated as *grades A–E*.

DISCUSSION

This research project was initiated to improve the quality of assessment of physiological concepts in a second-year dentistry course. The project has reinforced the commitment of faculty members to using group-constructed concept maps with criteria-based, assessment rubrics. Together, these are perceived as effective learning and assessment options to integrate complex information in physiology.

The project clearly demonstrates the importance of faculty members working together in all assessment stages: the design, construction of tasks and grading rubric, grading, and evaluation (17). Multiple opportunities for reflection and discussion of the issues around assessment of the concept maps played a critical role in collaborative learning among faculty members. For example, initial grading indicated much agreement between *faculty members A* and *B*, which was further improved by the moderation process (Table 1). The same process quickly isolated threshold knowledge as an issue that could then be specifically addressed. Tacit understandings of disciplinary knowledge and understandings of assessment accruing from personal experiences were made explicit. Knowledge representations differed and had to be explicitly defended and resolved during discussions that formed the moderation process (12). This "externalization of interpretative frameworks" (8) led, in our case, to designing explicit criteria and standards likely to increase reliability of future assessment. It was particularly useful for faculty members to deepen their understanding of the concept mapping task, e.g., reaching a consensus that the provisional criterion of Understanding could only be demonstrated and therefore reliably assessed if it were included within the criterion of Logic.

Student perceptions were comprehensively examined through the analysis of the survey questions. Content, Logic, and Presentation were considered by the students to be important criteria. However, the criterion of Understanding was considered less important for both survey *questions one* and *two*. This did not match with the opinions of faculty members: that Understanding was critical and only demonstrable using the logical structure of propositions. Students were engaged in offering feedback (Tables 2 and 3) and, we infer, are very interested in improving assessment of (and learning through) this task. Their responses echoed the concern of faculty members for needing more structured guidelines to assist them in completing the task more efficiently.

Negotiated assessment between faculty members and students requires capturing and equitably representing the perceptions of both faculty members and students (7). In redrafting the assessment rubric for grading concept maps, we gave considerable effort to language selection and use (21). The language of both faculty members and students are highlighted and differentiated for the purpose of this paper (Fig. 3).

This project reinforced the importance of using language that reflects the understandings of both faculty members and students. The findings of this project were clearly useful for faculty members in the short term. We expect the grading rubric will be useful to students in future years. In the 2005 phase of research, we will explore the processes by which dentistry students use the assessment rubric we developed. We

	A	B	C	D	E
CONTENT	▪ All relevant *concepts* of P & P and *mechanisms* are included and correct ▪ All *drug* names, *sites* of action and *uses* are correct ▪ Non-drug interventions are correct and **comprehensive**	▪ All relevant *concepts* of P & P and *mechanisms* are included and correct ▪ Most *drug* names, *sites* of action and *uses* are correct ▪ Non-drug interventions are correct but not **comprehensive**	▪ Most relevant *concepts* of P & P and *mechanisms* are included and correct ▪ Most *drug* names, *sites* of action and *uses* are correct ▪ Most non-drug interventions are correct, but not **comprehensive**	▪ Few relevant *concepts* of P & P and *mechanisms* are included or correct ▪ Some *drug* names, *sites* of action and *uses* are correct ▪ Some non-drug interventions are correct	No submission
LOGIC and UNDERSTANDING	▪ *Understanding* of *facts* and *concepts* of P & P is clearly demonstrated by correct *links* and active verbs	▪ *Understanding* of *facts* and *concepts* of P & P is demonstrated by correct *links* but some verbs are not appropriate	▪ *Understanding* of *facts* and *concepts* of P & P is demonstrated but with some errors (e.g. incorrect *links* and verbs)	▪ Poor *understanding* of *facts* and *concepts* of P & P with significant errors	No submission
PRESENTATION	▪ Concept map is *neat*, **clear**, **legible** and has *easy-to-follow* links ▪ has no **spelling errors**	▪ Concept map is *neat*, **clear**, **legible** and has *easy-to-follow* links ▪ has some **spelling errors**	▪ Concept map is *neat*, **legible** but with some links difficult to follow ▪ has some **spelling errors**	▪ Concept map is untidy with links difficult to follow ▪ has some **spelling errors**	No submission

Fig. 4. Revised assessment rubric. P & P, physiology and pharmacology. Words in italics reflect students' language. Boldface words reflect additional faculty member language not used frequently by students

anticipate a 2006 phase of research in which we explore peer and self-assessment around this task.

REFERENCES

1. **Ausubel DP, Novak JD, and Hanesian H.** *Educational Psychology: a Cognitive View*. New York: Holt, Rhinehart and Winston, 1978.
2. **Ballantyne R, Hughes K, and Mylonas A.** Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment Evaluation Higher Educ* 27: 427–441, 2002.
3. **Besterfield-Sacre M, Gerchak J, Lyons M, Shuman LJ, and Wolfe H.** Scoring concept maps: an integrated rubric for assessing engineering education. *J Eng Educ* 93: 105–116, 2004.
4. **Biggs J.** *Teaching for Quality Learning at University* (2nd ed.). Queensland, Australia: Society for Research into Higher Education and Open Univ. Press, 2003, p.16–17.
5. **Bogdan RC and Biklen SK.** *Qualitative Research for Education: an Introduction to Theory and Methods* (2nd ed.). Boston, MA: Allyn and Bacon, 1992.
6. **Goos M and Moni K.** Modelling professional practice: a collaborative approach to developing criteria and standards based assessment in preservice teacher education courses. *Assessment Evaluation Higher Educ* 25: 73–78, 2001.
7. **Gosling D.** Using Habermas to evaluate two approaches to negotiated assessment. *Assessment Evaluation Higher Educ* 25: 293–304, 2000.
8. **Howe KR.** The interpretive turn and the new debate in education. *Educational Researcher* 27: 13–20, 1998.
9. **Johnston L and Miles L.** Assessing contributions to group assignments. *Assessment Evaluation Higher Educ* 29: 751–768, 2004.
10. **Krathwohl DR.** *Methods of Educational and Social Science Research: an Integrated Approach*. New York: Longman, 1998.
11. **Lejk M, Wyvill M, and Farrow S.** A survey of methods of deriving individual grades from group assessments. *Assessment Evaluation Higher Educ* 21: 267–280, 1996.
12. **McCaghie WC, McCrimmon DR, Mitchell G, Thompson JA, and Ravitch MM.** Quantitative concept mapping in pulmonary physiology: comparison of student and faculty knowledge structures. *Adv Physiol Educ* 23: 72–81, 2000.
13. **McClure JR, Sonak B, and Suen HK.** Concept map assessment of classroom learning: reliability, validity, and logistical practicality. *J Res Sci Teaching* 36: 475–492, 1999.
14. **Orsmond P, Merry S, and Reiling K.** The use of exemplars and formative feedback when using student derived grading criteria in peer and self-assessment. *Assessment Evaluation Higher Educ* 27: 309–323, 2002.
15. **Ruiz-Primo MA, Schultz SE, Li M, and Shavelson RJ.** Comparison of the reliability and validity of scores from two concept-mapping techniques. *J Res Sci Teaching* 38: 260–278, 2001.
16. **Ruiz-Primo MA, Shavelson RJ, Li M, and Schultz SE.** On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment* 7: 99–141, 2001.
17. **Rust C, Price M, and O'Donovan B.** Improving students' learning by developing their understanding of assessment criteria. *Assessment Evaluation Higher Educ* 28: 147–164, 2003.
18. **Rye JA and Rubba PA.** Scoring concept maps: an expert map-based scheme weighted for relationships. *School Sci Math* 102: 33–45, 2002.
19. **Sadler DR.** Evaluation and the improvement of academic learning. *J Higher Educ* 13: 191–209, 1983.
20. **Sadler DR.** Specifying and promulgating achievement standards. *Oxford Rev Educ* 54: 60–79, 1987.
21. **Woolf H.** Assessment criteria: reflections on current practices. *Assessment Evaluation Higher Educ* 29: 479–493, 2004.