

## Peer assessment: A valuable tool to differentiate between student contributions to group work?

CAROLINE STEENSELS<sup>1</sup>, LIES LEEMANS<sup>1</sup>, HERMAN BUELENS<sup>2</sup>, ELISABETH LAGA<sup>2</sup>, ANNIE LECOUTERE<sup>3</sup>, GERT LAEKEMAN<sup>1</sup>, & STEVEN SIMOENS<sup>1</sup>

<sup>1</sup>Faculty of Pharmaceutical Sciences, Centre for Drug and Patient Information, K.U.Leuven, Edward Van Evenstraat 4, 3000 Leuven, Belgium, <sup>2</sup>Information and Communication Technology in Education, University Education Support Office, K.U.Leuven, Naamsestraat 98, 3000 Leuven, Belgium, and <sup>3</sup>Faculty of Pharmaceutical Sciences, Centre for Teacher Training, K.U.Leuven, Edward Van Evenstraat 4, 3000 Leuven, Belgium

### Abstract

This paper aims to report experiences with peer assessment (PA) during the first year of operation in a pharmacy practice (PP) course. PA was carried out twice. The range and standard deviation of scores were larger during the second assessment, suggesting a more critical way of rating. This indicates that gaining experience with PA is a necessary condition for this tool to enable students to assess each other's contribution to group work. A significant correlation was observed between scores awarded by peers and external tutors, indicating that students assessed one another in the same way as external tutors. In conclusion, PA can be a valuable tool to differentiate between student contributions to group work if students are properly trained to conduct such an assessment. It is recommended that PA be integrated at an earlier stage of the curriculum to allow students to gain the required expertise.

**Keywords:** *Education, group project, peer assessment, pharmacy, tutor assessment*

### Introduction

In recent years, various forms of group work have been introduced in university courses. Group work is viewed as an instrument that enables students to develop a core set of skills that have been labelled in the literature as “transferable skills” (Assiter, 1995), “key competencies” (Mayer, 1992), “generic attributes” (Wright, 1995) or “capabilities” (Stephenson & Yorke, 1998). Students who have mastered such skills are able to work collaboratively with others, reflect critically on their own thinking, take responsibility for their own learning, communicate with others, develop a self-learning attitude and, ultimately, engage in a process of lifelong learning.

In 2004, a new pharmacy practice (PP) course was added to the curriculum of fourth-year pharmacy students at the K.U.Leuven, which covers five main topic areas: pharmaceutical care, communication,

pharmacotherapy on over the counter (OTC-) drugs, pharmaco-economics and drug policy. The PP course has moved away from the traditional practice of students learning in isolation and, instead, has embraced the concept of collaborative learning. This “social” approach to supporting students’ learning purports to facilitate the development of independent students/pharmacists who acquire a set of practical competencies in addition to theoretical knowledge.

To this effect, the course draws on a combination of different learning methods with particular emphasis on problem-based and collaborative learning. Group work is the mainstay of the new PP course. The theoretical aspects of pharmaceutical care, self-management and pharmaco-economics are put into practice during the group work. This group work is undertaken in collaboration with and supervised by external tutors from professional organizations or

---

Correspondence: C. Steensels, Faculty of Pharmaceutical Sciences, Centre for Drug and Patient Information, K.U.Leuven, Edward van Evenstraat 4, 3000 Leuven, Belgium. Tel: 32 16 323451. Fax: 32 16 323468. E-mail: caroline.steensels@pharm.kuleuven.be

industry. Additionally, an internal tutor from K.U.Leuven follows up group work. To facilitate discussion with peers, tutors and supervisors, each group of students can make use of a private IT-discussion board. This has been made possible by the launch of “TOLEDO” (“Toetsen en Leren Doel-treffend Ondersteunen”, i.e. “efficiently supporting learning and evaluation”), a virtual learning environment that not only facilitates group work, but also allows students to tackle several problem-based exercises on communication and self-management (Leemans, Verstraeten, Zwaenepoel, & Laekeman, 2003). Furthermore, during a role-play session, communication skills are taught and practiced in small interactive groups. Sufficient time has been set aside for students to spend on online exercises and group work as the number of classroom lectures has been reduced to less than 25% of total course time.

Acceptance of group work by students and the success of group work critically depends on a fair and credible assessment of the group process. Assessment by students themselves can be used as a tool to evaluate group work. The aim of this article is to describe our experience with developing and implementing a peer assessment (PA) tool for group work undertaken by students within the context of the PP course and to report findings of PA during the 2003–2004 academic year, the first year during which the course was taught. In particular, this analysis explores whether students are capable of assessing each other and, therefore, whether PA can be used as a tool to differentiate between student contributions to group work. Additionally, students’ opinion of PA is presented. This article may guide other course coordinators on how to develop a tool of student PA and on how to implement it.

#### *Peer assessment*

Student PA can be defined as a process in which each student evaluates the extent to which each of the other group members has exhibited certain traits, performed specific tasks or accomplished particular objectives (Kane & Lawler, 1978). PA may prove to be a valuable tool for evaluating group work for a number of reasons. As students are likely to have a better view on the contribution of each group member, PA may guarantee a more objective evaluation (Somervell, 1993; Dochy, Segers, & Buehl, 1999). Also, PA reinforces the message to students that not only individual achievement is evaluated, thereby inciting students to work together (Boud, Cohen, & Sampson, 1999). Finally, PA has been recognized as a skill necessary for professional practice (Des Marchais & Vu, 1996; Heylings & Stefani, 1997; Pond & ul-Haq, 1997; Thomas, 1997; Das, Mpofu, Dunn, & Lanphear, 1998; Sullivan, Hitchcock, & Dunnington, 1999; Sluijsmans, Moerkerke, Merrienboer, & Dochy, 2001).

On the other hand, academics need to be aware of various possible causes of ratter bias in PA. Several studies have reported poor correlations between marks awarded by students and by tutors (Kane & Lawler, 1978; Falchikov, 1986; Falchikov & Magin, 1997; Orsmond, Merry, & Reiling, 2000). These findings, relating to validity and reliability, were either linked to problems of fairness and the individual’s feelings toward group members (Kane & Lawler, 1978; Fox, 1989; Williams, 1992; Boud et al., 1999; Magin, 2001b; Pope, 2001), or to issues concerning the use of students as assessors, including their lack of ability to discriminate between levels of performance and their reluctance to judge their peers (Falchikov, 1995; Orsmond & Merry, 1996; Sullivan et al., 1999; Li, 2001; Sluijsmans et al., 2001).

PA can take the form of peer nomination, identification of best and worst student by group members; peer rating, assessment of each student by the rest of the group in terms of a set of performance characteristics; and peer ranking, a ranking of all individual students in the group from best to worst against a given set of characteristics. It is not clear which form of PA is to be recommended. Arguments that peer nomination could be subject to bias and of limited use in providing feedback have been made. Nonetheless, it is valid and reliable in identifying extremes of behaviour (Kane & Lawler, 1978; Boud et al., 1999). There is evidence indicating that peer ranking leads to (quite) similar observations (Mac-Alpine, 1999). Peer rating, on the other hand, has been found to be superior for use in formative assessment procedures, focusing upon feedback and fostering deep level learning (Kane & Lawler, 1978; Fry, 1990; Williams, 1992; Boud et al., 1999; ).

A crucial message emerging from this literature is that there are some pre-conditions to using PA as a tool for evaluating student contributions to group work. First, the group size cannot be too large and the group must interact frequently in order to have a salient view of each student’s work. Second, students have to be aware of the criteria used for assessment. Finally, students need to be skilled in carrying out PA. In other words, students have to learn to assess one another. Brown and Knight (1995) claim that students, who participate for the first time in PA, need the tutor to provide them with the performance criteria on which they have to assess their peers. When they have mastered this skill, students are capable of assessing their peers in an accurate way (Brown & Knight, 1995; Segers & Dochy, 2001; Dochy, Heylen, & van de Mosselaer, 2002).

#### **Materials and methods**

In line with the need for multi-uniformity in student assessment (Dochy et al., 2002), PA was incorporated as part of a triangulated approach to assessment. This

means that students were evaluated by multiple assessors: the internal university tutor at the university who evaluates the final report arising from the group work (accounting for 70% of marks on group work), the external tutor from a professional organization or industry (15% of marks) and the students as peers (15% of marks). Less weight was attached to PA as this was the first time this approach was used. Given that external tutors have little training or experience in evaluating students, their assessment accounted for only 15% of marks.

The PA tool used is based on an existing and validated instrument consisting of six performance characteristics (Dochy et al., 2002). Whereas some other tools only consist of two characteristics, this may deliver a too narrow evaluation of the student's commitment. The use of six criteria provides a more specific and comprehensive evaluation of the contribution of each student to the group work. The characteristics used are related to aspects of basic knowledge, problem solving, motivation, creativity, academic skills (e.g. independent data collection and critical reflection), and group functioning (working collaboratively with others). Students evaluated their peers on each performance characteristic using a four-point Likert scale (0, did not contribute; 1, below average contribution; 2, average contribution; 3, above average contribution). The tool was pilot tested among fifth-year pharmacy students. An open discussion with students revealed that two performance characteristics were vague and difficult to understand. This resulted in high grades according to social desirability. As a result, the instrument was revised before using it in the fourth year: two performance characteristics were clarified by adding a short description.

Table I illustrates how the PA tool was used in practice and shows a fictitious example of the points awarded to Student 6 by his/her five group members. The total score of Student 6 is calculated as follows. First, the score awarded to Student 6 by each of the five group members is calculated as the mean across the six performance characteristics. Second, the

average is taken of the score awarded to Student 6 by each of the other five group members.

A first assessment took place after three weeks of group work in March 2004. The main objective of this assessment was to get students acquainted with both the procedure and the process of rating each other. The outcome of the PA was followed by an open discussion about the contribution of each group member. If needed, the characteristics were explained to students in more detail. The discussion was also used to inform students of their contribution to the group. In particular, each student was informed of whether (s)he was working hard or needed to improve his/her contribution according to peers. A second assessment in May 2004, on the same group of students, was used for the purpose of student assessment. As other authors have stressed the need for confidentiality when students assess their peers (Lejk & Wyvill, 2001), students were able to complete the form on their own and send it back by e-mail.

A survey of the students was undertaken at the end of May 2004 to gain an understanding of their opinion of PA. This survey enquired about the following five features of the PA tool: accuracy of PA in the evaluation of group work; capability of students in evaluating each other fairly; suitability of students as compared to mentors/tutors to evaluate the group work; acquisition of skills to evaluate colleagues; and share of points dedicated towards PA. Figure 1 phrases the exact questions. Every student was invited to express his/her views on these statements with "I agree" or "I don't agree". Again, students could fill in the questionnaire on their own and send it back by e-mail. As this survey was carried out while students were taking the PP course, students may have felt pressured to give a positive assessment of PA. Therefore, the same survey was repeated 6 months later during the next academic year to explore whether their opinion of PA had changed after completion of the PP course.

In addition to PA, internal tutors evaluated the report prepared by group members, assigning a score

Table I. Fictitious example of PA score awarded to student 6 by five group members.

	Student 1	Student 2	Student 3	Student 4	Student 5	Score averaged across students
Basic knowledge	2	2	2	1	3	2
Problem solving	2	1	2	2	2	1.8
Motivation	3	0	2	3	2	2
Creativity	3	1	3	1	0	1.6
Academic skills (such as independent data collection, critical reflection)	2	1	3	1	2	1.8
Group functioning (working collaboratively with others)	2	0	2	2	1	1.4
Score averaged across performance characteristics	2.3	0.8	2.3	1.7	1.7	
Mean score of Student 6						1.7

Note: Characteristics were evaluated on a four-point Likert scale with: 0, did not contribute; 1, below average contribution; 2, average contribution; and 3, above average contribution.

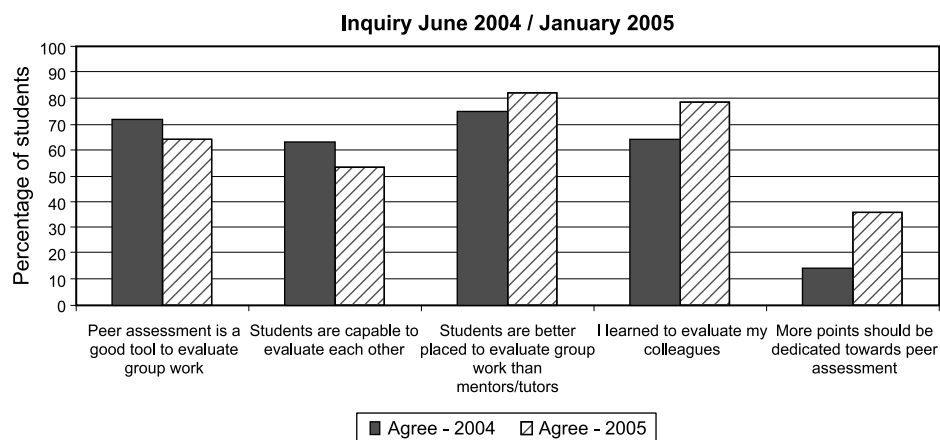


Figure 1. Survey results of students' opinion of PA.

on 20 points. Assessment by external tutors was as follows. As external tutors are not university staff and, thus, are unlikely to have experience with student assessment, tutors identified extremes of behaviour. They nominated the best and worst performing members of the group by allocating a "bonus" (" + ") and "malus" (" - "), respectively, on each performance characteristic. All group members started with a score of 7.5/15. Every "bonus" and "malus" accounted for an increase and decrease, respectively, of 1.25 points. A fictitious example of how an external tutor used this bonus/malus system to evaluate a group of students and how student scores were calculated is portrayed in Table II. Scores awarded by external tutors were then recoded to a scale from 0 to 3 to enable comparison with scores awarded by PA.

## Results

During the 2003–2004 academic year, a PA of group work that was undertaken by students within the context of the PP course was carried out. Twenty-eight students were allocated to five projects based on individual preferences. In practice, this implied that each student was assigned to the project of his/her first or second choice. The objective was that all groups would be more or less of the same size in order for each student to have a similar workload. This resulted in three six-member groups and two five-member groups.

The projects dealt with different subjects from the field of pharmaceutical care. Two of them were conducted in cooperation with pharmaceutical firms, Aventis and Janssen-Cilag. Two other projects were supervised by professional organizations and the last one was conducted at K.U.Leuven University. In the Aventis project, students carried out a pharmacoeconomic evaluation of two antibiotics in an intensive care setting. The Janssen-Cilag project made use of the "mystery-shopping" technique to get an idea of the most popular/sold cough syrups and the way they were

delivered by pharmacists. During one of the projects managed by professional organizations, students gained more in-depth knowledge about migraine and then designed a handout for patients about this subject. The other project group examined the effects of intracollegiate networking between pharmacists. In the last project group based at K.U.Leuven University, students examined gastro-intestinal and cardiovascular co-medication with COX-2 inhibitors or NSAIDs for the treatment of osteoarthritis in a sample of patients in ambulatory care.

The first PA undertaken in March 2004 indicated that students did not differentiate much when allocating grades to their peers. They tended to assign a very narrow range of scores (range = 0.5, standard deviation = 0.11), usually at the high end of the Likert scale (mean = 2.14). This narrow range suggested that students can not or do not wish to differentiate between their peers.

In the second PA, carried out in May 2004, a wider range of scores was observed (range = 1.8) and the standard deviation increased almost 4-fold (standard deviation = 0.42). The average score comes to 2.03. This wider range and standard deviation may reveal a more critical way of rating and show that students are able to assess each peer's contribution to group work as compared to that of other group members. No significant difference was found between average scores in March and May 2004. This suggests that, on average, students evaluated their peers on the same level, even though they were more likely to assign lower and higher scores to peers in May 2004.

*How do scores of (the second) PA compare to those awarded by external tutors?*

A statistically significant positive correlation was found between scores (Spearman correlation coefficient = 0.49,  $p = 0.009$ ), suggesting that students and external tutors assessed student contributions to group work in similar

Table II. Fictitious example of assessment by external tutor in five-member project group.

	Student 1	Student 2	Student 3	Student 4	Student 5
Basic knowledge	+				
Problem solving			+		
Motivation				+	
Creativity	+				-
Academic skills			+		
Group functioning		-			-
SCORE	10/15	6.25/15	10/15	8.75/15	5/15

ways. It is noteworthy that the average score granted by PA of 2.03 exceeded the average score given by external tutors of 1.59. In other words, students seem to award higher marks to their peers than external tutors.

The first survey (June 2004) questioned students about their opinion of PA and revealed a positive attitude towards PA (Figure 1). The majority of the students (71%) viewed PA as a suitable instrument to evaluate group work. Sixty-four percent of students stated that they had learned to evaluate colleagues and that they are now capable of evaluating each other. Most students felt better placed than internal/external tutors to evaluate their group work (75%). This corroborates the value of using a two-step approach in which students first carry out an assessment to gain expertise with PA and then conduct a second assessment for the purpose of evaluation. Even though students generally had a positive attitude towards PA, 86% of students felt that PA should not gain more weight in the overall assessment of group work (PA accounted for 15% of marks on group work).

The same survey, conducted during the next academic year 6 months later (January 2005), produced slightly different results. Fewer students (64% vs. 71%,  $p = 0.775$ ) thought of PA as a valuable tool to evaluate group work. Even though a growing majority of students felt that they had learned to evaluate each other (79% vs. 64%,  $p = 0.375$ ), fewer students were convinced of their capability of evaluating colleagues (54% vs. 63%,  $p = 0.665$ ). Nevertheless, eight out of ten (82% vs. 75%,  $p = 0.745$ ) felt better placed than internal/

external tutors to evaluate their group work. When asked in June, 14% of students wished to see more points dedicated to PA. Six months later, this was 36% ( $p = 0.123$ ). In other words, there was no significant difference between the two surveys.

Each project group consisted of 5-6 students and each student was assessed by the other members of the group. To investigate whether the scoring happened in a fair and objective way, the standard deviation of the scores that a student received from his/her peers in the project group was calculated. A low standard deviation would indicate that peers have given similar scores to a student and may be taken as an indicator that PA was objective. Figure 2 shows the standard deviation in every project group. In four of the five project groups the standard deviation fluctuates around 0.2. In one group, this standard deviation is remarkably higher. The scores of two students even have a standard deviation around 1. According to the external tutor, two students in this group isolated themselves and almost did not participate at all. These two students gave each other high marks, but obtained lower marks from their other colleagues, giving rise to the high standard deviations. This suggests that PA can be a valuable tool to evaluate student contributions insofar that students assess each other objectively.

## Discussion

This study has described an approach to developing a PA tool to evaluate the contribution of each student to

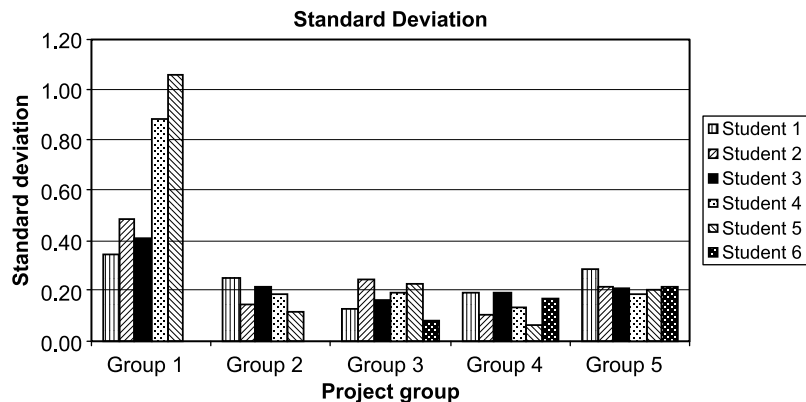


Figure 2. Standard deviation on PA scores.

group work as part of a PP course taught to fourth-year pharmacy students. Findings of the first year during which the PA tool was implemented were presented.

This experience with PA shows that students tended to assign a narrow range of scores, usually at the high end of the rating scale, which corresponds with the findings of PA undertaken in other courses by different types of students (Conway, Kember, Sivan, & Wu, 1993; Falchikov, 1995; Pond & ul-Haq, 1997; Das et al., 1998; Cheng & Warren, 1999; MacAlpine, 1999; MacPherson, 1999; Sullivan et al., 1999; Purchase, 2000; Miller, 2003). However, a comparison of the two assessments indicated that the standard deviation and range of scores increased. This shows that PA can be a valuable tool to differentiate between student contributions to the group, if students acquire the necessary skills to carry out a PA. This condition supports our rationale for adopting a two-step approach to PA. The aim of the first assessment was to get students acquainted with both the procedure and the process of rating each other. This assessment was followed by a discussion about the contribution of each group member. In this way, students learned to assess each other before the second assessment that was used for the purpose of evaluation.

Such an approach is in line with that of other researchers who claim that student PA of group work serves the purposes of both forming students and evaluating them (Orsmond & Merry, 1996; Das et al., 1998; Topping, 1998; Purchase, 2000; Li, 2001). From our perspective, PA was primarily used as a tool to evaluate student contributions to group work. However, the formative aspect of PA needs to be emphasized as well. PA also aims to stimulate students to become reflective practitioners who are able to be critical of both their own learning experiences and those of their peers/colleagues. Topping (1998) concluded that PA can bring students up to a high level of responsibility over their own learning. This formative aspect of PA is widely supported in the literature (Boud, 1989; Fox, 1989; Fry, 1990; Williams, 1992).

Various studies have reported poor correlations between marks awarded by students and by tutors (Kane & Lawler, 1978; Falchikov, 1986; Falchikov & Magin, 1997; Orsmond et al., 2000). These findings, relating to validity and reliability, were either linked to problems of fairness and the individual's feelings toward group members (Kane & Lawler, 1978; Fox, 1989; Williams, 1992; Boud et al., 1999; Magin, 2001b; Pope, 2001), or to issues concerning the use of students as assessors, including their lack of ability to discriminate between levels of performance and their reluctance to judge their peers (Falchikov, 1995; Orsmond & Merry, 1996; Sullivan et al., 1999; Li, 2001; Sluijsmans et al., 2001). This literature indicates that the validity and reliability of PA is

inhibited by students' lack of capability to score their peers' performance. Findings in this study reveal that this does not have to be the case if students are first taught how to evaluate their peers. An improvement in the ability of students to assess their peers was noted as evidenced by the larger standard deviation and range of scores during the second PA.

A significant correlation was observed between scores awarded by peers and by external tutors. This shows that students appear to have assessed one another in the same way as external tutors and, thus, may be taken as an indicator of an objective assessment. Apparently, students do not let personal relations interfere with their scoring. Furthermore, it is important to stress that the tool of student PA was not set up as a means of validating scores granted by external tutors. Diverging opinions of the purposes of multiple assessments have been formulated. Multiple assessment can be considered either as a tool to achieve convergence, i.e. agreement among the assessors, or as a way to ensure completeness, i.e. the uncovering of multiple perspectives on student behaviour (Breitmeyer, Ayres, & Knalf, 1993; Sim & Sharp, 1998). Previous studies dealing with PA have commonly concentrated on agreement between peers themselves and between peers and faculty assessors (Falchikov & Goldfinch, 2000; Magin, 2001a). On the other hand, from our perspective, PA was viewed as a valid evaluation tool in its own right that made up one element of a multiform approach to assessment. This is based on our opinion that PA reveals another aspect of group work that cannot be evaluated by the internal or external tutor. The motivation in this research to introduce PA was to reveal the exact contribution of each member of the group, from three different perspectives: the point of view of peers, external and internal tutors.

Nevertheless, it is likely that there is some overlap in the evaluation by the external tutor and the peers as they both evaluate group work. This does not apply to the internal tutor who evaluates the result of group work—the report—rather than group work itself. On the other hand, external tutors assess group work from another point of view than that of peers. Also, differences in commitment of external tutors to supervise group work, with some tutors being more active than others, is likely to influence their ability of assessing each student's contribution to group work. Moreover, external tutors had not acquired the same level of expertise in evaluating students. These arguments prompted us to not only rely on an assessment by tutors, but also to initiate PA, with a view to evaluating students in a comprehensive and fair manner.

Finally, the confidentiality incorporated in the approach used in this study may be considered to be a weakness. Students completed both PAs on-line. Confidentiality was guaranteed aiming at honest and

differentiating assessment scoring. However, confidentiality was not explicitly checked. Students could have agreed upon how to score each other before filling in their forms. Such behaviour cannot be excluded and any pre-arranged agreement may have had a distorting effect on the results.

In conclusion, our experiences have highlighted the importance of the formative aspect of PA. PA can serve as an instrument to enable students to become reflective practitioners who can assess their own learning experiences and those of their peers. For the purpose of evaluation, PA appears to be a valuable tool to differentiate between student contributions to group work if students have acquired the necessary skills to carry out a meaningful and reliable evaluation. Therefore, in the opinion of the researchers, PA needs to be integrated in the curriculum at an earlier stage. This would allow students to develop expertise throughout their study years at university and become reliable assessors of their own learning experiences and that of others by their fourth or fifth year. In that way, PA could be used as a credible assessment tool in the master program. By then, students would be skilled “reflective practitioners” with a co-operative learning style, who have developed confidence in their self-learning with a sound judgement of their value and a sense of responsibility for their own life-long learning.

### Acknowledgements

The authors are indebted to fourth-year pharmacy students who carried out the PA and gave their opinion of it. No sources of funding were used to assist in the preparation of this manuscript. The authors have no conflicts of interest that are directly relevant to the content of this manuscript.

### References

- Assiter, A. (Ed.) (1995). *Transferable skills in higher education*. London: Kogan Page.
- Boud, D. J. (1989). The role of self-assessment in student grading. *Assessment & Evaluation in Higher Education*, 14, 20–30.
- Boud, D. J., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24, 413–426.
- Breitmeyer, B. J., Ayres, L., & Knafl, K. A. (1993). Triangulation in qualitative research: Evaluation of completeness and confirmation purposes. *IMAGE Journal of Nursing Scholarship*, 25, 237–243.
- Brown, S., & Knight, P. (1995). Assessing learners in higher education. In *Teaching and learning in higher education*. London: Kogan Page.
- Cheng, W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education*, 18, 34–56.
- Conway, R., Kember, D., Sivan, A., & Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment & Evaluation in Higher Education*, 18, 45–56.
- Das, M., Mpofu, D., Dunn, E., & Lanphear, J. H. (1998). Self and tutor evaluations in problem-based learning tutorials: Is there a relationship? *Medical Education*, 32, 411–418.
- Des Marchais, J. E., & Vu, N. V. (1996). Developing and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. *Academic Medicine*, 71, 274–283.
- Dochy, F., Segers, M., & Buehl, M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research*, 69, 147–188.
- Dochy, F., Heylen, L., & van de Mosselaer, H. (2002). *Assessment in onderwijs: Nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs*. Utrecht: Lemma.
- Falchikov, N. (1986). Product comparison and process benefits of collaborative peer group assessment. *Assessment & Evaluation in Higher Education*, 11, 146–166.
- Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Educations and Training International*, 32(2), 175–187.
- Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group progress work. *Assessment & Evaluation in Higher Education*, 22, 385–396.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.
- Fox, D. (1989). Peer assessment of an essay assignment. *HERDSA News*, 11, 6–7.
- Fry, S. A. (1990). Implementation and evaluation of peer marking in higher education. *Assessment & Evaluation in Higher Education*, 15, 177–189.
- Heylings, D. J., & Stefani, L. A. (1997). Peer assessment feedback marking in a large medical anatomy class. *Medical Education*, 31, 281–286.
- Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin*, 85, 555–586.
- Leemans, L., Verstraeten, A., Zwaenepoel, L., & Laekeman, G. (2003). The use of a virtual learning environment during the internship of final year pharmacy students. *Pharmacy Education*, 3, 73–76.
- Leijk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment & Evaluation in Higher Education*, 26, 61–73.
- Li, L. K. (2001). Some refinements on peer assessment of group projects. *Assessment & Evaluation in Higher Education*, 26, 6–18.
- MacAlpine, J. M. K. (1999). Improving and encouraging peer assessment of student presentations. *Assessment & Evaluation in Higher Education*, 24(1), 15–25.
- MacPherson, K. (1999). The development of critical thinking skills in undergraduate supervisory management units: Efficacy of student peer assessment. *Assessment & Evaluation in Higher Education*, 24, 273–284.
- Magin, D. J. (2001a). A novel technique for comparing the reliability of multiple peer assessments with that of single teacher assessments of group process work. *Assessment & Evaluation in Higher Education*, 26, 139–152.
- Magin, D. J. (2001b). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education*, 26, 53–63.
- Mayer, E. Chair of Committee. (1992). *Putting general education to work: The key competency report*. Canberra: Mayer committee, Australian Government Publishing Service.
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28, 383–394.
- Orsmond, P., & Merry, S. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, 21, 239–251.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education*, 25, 23–38.

- Pond, K., & ul-Haq, R. (1997). Learning to assess students using peer review. *Studies in Educational Evaluation*, 23(4), 331–348.
- Pope, N. (2001). An examination of the use of Peer rating for formative assessment in the context of the theory of consumption values. *Assessment & Evaluation in Higher Education*, 26, 235–246.
- Purchase, H. C. (2000). Learning about interface design through peer assessment. *Assessment & Evaluation in Higher Education*, 25, 341–352.
- Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value-added of the students' perspective. *Studies in Higher Education*, 26, 327–343.
- Sim, J., & Sharp, K. (1998). A critical appraisal of the role of triangulation in nursing research. *International Journal of Nursing Studies*, 35, 23–31.
- Sluijsmans, D. M., Moerkerke, G., Merrienboer, J. J., & Dochy, F. J. (2001). Peer assessment in problem-based tutorials. *Studies in Educational Evaluation*, 27, 153–173.
- Somervell, H. (1993). Issues in assessment, enterprise and higher education. *Assessment and Evaluation in Higher Education*, 18, 221–233.
- Stephenson, J., & Yorke, M. (Eds.) (1998). *Capability and quality in higher education*. London: Kogan Page.
- Sullivan, M. E., Hitchcock, M. A., & Dunnington, G. L. (1999). Peer and self-assessment during problem-based tutorials. *American Journal of Surgery*, 177, 266–269.
- Thomas, R. E. (1997). Problem-based learning: Measurable outcomes. *Medical Education*, 31, 320–329.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68, 249–276.
- Williams, E. (1992). Student attitudes towards approaches to learning and assessment. *Assessment & Evaluation in Higher Education*, 17, 45–58.
- Wright, P. (1995). *What are graduates? Clarifying the attributes of "Graduateness"*. London: The Higher Education Quality Council (HEQC).