PEER ASSESSMENT IN THE TEAM-BASED LEARNING CLASSROOM

by

Charlene A. K. Strumpel

BSN, Okanagan University College, 2001

A MAJOR PROJECT SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN NURSING

We accept this major project as conforming to the required standard

_____

Barb Pesut, PhD, RN, Supervisor

_____

Carole Robinson, PhD, RN, Committee Member

UNIVERSITY OF BRITISH COLUMBIA OKANAGAN

December, 2011

TABLE OF CONTENTS

Title of Major Project:

PEER ASSESSMENT IN THE TEAM-BASED LEARNING CLASSROOM

ABSTRACT

Although peer assessment (PA) is purported to be an essential element of Team Based Learning (TBL), there is little evidence to indicate the most effective method of PA in this context. This literature review was undertaken to examine the research relating to PA of individual students' contributions to group work. As such, this literature review might prove useful and informative for educators using PA in a variety of contexts, including problem-based learning, case-based learning, or group projects. A total of 45 research articles describing both qualitative and quantitative studies were examined. These articles were examined to investigate the potential benefits or drawbacks of PA; the validity and reliability of PA; what PA process and instrument constitute best practice; and the method that is most appropriate to use when calculating a student's grade from a PA score. Each of these issues was examined and, where possible, evidence and recommendations from the literature were presented. In addition to describing many of the factors that teachers should consider prior to implementing PA in a university course, this literature review also revealed many gaps in knowledge and potential avenues for research into PA.

INTRODUCTION/PROBLEM STATEMENT

In 2009, in response to increasing class size, and in an effort to increase student engagement in the classroom, several teachers in the second year of the Bachelor of Science in Nursing (BSN) program at the University of British Columbia Okanagan (UBCO) implemented Team-Based Learning (TBL) in a total of eight classroom courses. TBL is a very specific form of cooperative learning that will be described below. However, one key feature of TBL is that students are held accountable for their individual work as well as their group work. The method used to ensure accountability for group work is peer assessment (PA). In the context of TBL, students assess their peers' contributions to group work. At this point, it may be relevant to note that in the literature, the terms 'peer assessment' and 'peer evaluation' are often used interchangeably. For the sake of consistency, the term 'peer assessment' will be used in this paper.

Prior to the first term in which TBL was implemented, many of the teachers attended workshops, consulted with faculty who were experienced in TBL, and reviewed the literature on TBL. A key resource was Michaelsen, Parmelee, McMahon and Levine's (2008) book on *Team-Based Learning for Health Professions Education*. During this planning phase, teachers found it challenging to select a process for PA, since several methods were reported in the TBL literature, yet no single method was recommended above the rest. As a result, the teachers chose to use a PA process that seemed easy to understand, straightforward to explain to students, and easy to translate into a student grade. Although there were some variations in the way that PA was implemented in each course, the basic method was fairly consistent. An example of one of the first PA forms developed is shown in Appendix A. In this course, students were required to complete this PA form at midterm and at the end of the term, using an on-line program called

iPeer. Once all of the students had submitted their PAs, the results were released so that students could view their feedback. While students could see the numeric scores and comments from their peers, the names of the students who provided the feedback was kept confidential (although teachers could track the assessments of each student). The midterm PA provided the students with an opportunity to practice the process, as well as to receive formative feedback from their teammates. At the end of the term, each student received a score (out of 10) that counted for 10% of the student's final grade. This grade was simply calculated by averaging all of the PA scores awarded by each student's peers.

Over the next two years, even as these teachers became more skilled at facilitating the overall process of TBL in their classrooms, they began to express reservations about PA. Anecdotally, teachers questioned the reliability and validity of PA grades, and many students commented on their dissatisfaction with the PA process. A few of the teachers even stopped requiring students to engage in PA. These concerns provided an excellent opportunity to perform a literature search and examine the research on PA of students' contributions to group work.

Purpose of the Major Project

The purpose of this major project was to examine the current research on PA in post secondary education. It was anticipated that this evidence would support the revision of the current PA process to one that would be considered fair and equitable to both teachers and students. The broad question guiding the literature review was: What is the best evidence relating to PA of individual students' contributions to group work? The specific questions to be answered were:

1.   What are the potential benefits and drawbacks of PA?

2.   How valid and reliable is PA?

3.   What PA process and instrument constitute the best practice?

4.   What is the most appropriate method for incorporating PA in the calculation of a

student's final grade?

Background

Peer assessment is reported to be an integral part of TBL (Michaelsen & Sweet, 2008), yet

many educators report that implementing such PA is fraught with challenges (Levine, 2008).

However, to gain an understanding of the context in which PA is applied, it is first necessary to

provide a brief overview of TBL. After this summary of TBL, a description of the PA process –

as it was implemented in the 2nd year of the BSN program at UBCO – is provided.

Team-Based Learning

Parmelee (2008) explains that TBL was originally developed by Larry Michaelsen, a

faculty member in the business school at the University of Oklahoma. When confronted with

increasing class sizes, Michaelsen elected to incorporate group activities in his business course

rather than to utilize a lecture format. Over the years, Michaelsen's method has evolved into a

highly structured format, now known as TBL. TBL is now used in a variety of educational

settings, including (but not limited to) economics (Espey, 2008); music (Parker, 2007); law

(Dana, 2007); dentistry (Pileggi & O'Neill, 2008); management (Fairfield & London, 2003); and

nursing (Clark, Nguyen, Bray, & Levine, 2008). TBL has also been extensively used and

researched in medical education (Levine, Kelly, Karakoc, & Haidet, 2007). Fink and Parmelee

(2008) estimated that TBL is used in approximately 88 medical schools around the world, and

suggest that TBL allows students to learn complex information, gives them the opportunity to

apply that information to real-world problems, and helps them to develop communication and

teamwork skills.

Michaelsen and Sweet (2008) explain that there are four essential principles in TBL: (1) forming student teams appropriately and managing them effectively, (2) ensuring that students are accountable for their individual and team work, (3) ensuring that students receive frequent and timely feedback, and (4) promoting both learning and teamwork skills through team assignments. At the start of the term, they recommend that students should be assigned to teams of approximately 5-7 members. Teachers may use a variety of methods to assign students to teams, but regardless of the method, the process should be transparent to the students, and the method should distribute expertise and resources equally among the teams. It is also recommended that each course be divided into approximately 5-7 instructional units or modules. Prior to the start of each module, the teacher assigns readings or other preparatory work that will allow students to learn the basic concepts. On the first day of a new module, students start the class by completing the Readiness Assurance Process (RAP). In the RAP, students first complete a short multiple choice quiz. Next, the students work in their teams to complete the *same* multiple choice quiz. The teams may use Immediate Feedback Assessment Technique (IF-AT) cards to score their answers (Epstein Educational Enterprises, n.d.). Similar to a lottery ticket, the possible answers to each multiple choice question are covered by an opaque film. Students scratch off the film, looking for a star under the correct answer to discover if they have answered the questions correctly. If the students select an incorrect answer, they continue looking for the correct answer until they find it. Students get more marks for finding the star on the first try than if they need to scratch off more than one choice to find the answer. After completing the individual and team tests, any team may appeal a question if they disagree with the teacher's answer. After class, the teacher considers each appeal and may choose to allow the appeal if the students have provided a good rationale for their argument.

Based on the results of the individual and team tests, the teacher should be able to judge whether the students have a good understanding of the basic concepts. If students seem to be confused about any portions of the content, the teacher may elect to provide a short, focused lecture. If the students have a good grasp of the content, then the teacher may ask the students to complete team or individual activities in which the students practice applying the content to real-world problems.

TBL is structured in such a way as to ensure that students are accountable for learning the course content. First, the individual readiness assessment tests (and any other individual assignments or tests) ensure that students are held individually accountable for their work. Second, each student is accountable to the other members of his or her team. For a team to be successful, each student must come to class prepared for the team activities. Success of the teams also depends on students' communication and teamwork skills, ensuring effective collaboration on problems. To ensure that students are held accountable to their team members, the students in each TBL team assess the contributions of each of their peers. This PA assures students that each team member will be held accountable for his/her work, discouraging students from the "social loafing" (Levine, 2008) that occurs in other group settings. A more complete description of TBL is provided in the book, *Team-Based Learning for Health Professions Education* (Michaelsen et al., 2008).

Peer Assessment in the BSN Program at UBCO

After the decision was made to implement TBL in the second year of the BSN program, all the teachers with courses starting in September met prior to the fall term to plan how the various aspects of TBL (including PA) would be conducted. Unless otherwise specified, the descriptions of the PA processes (below) are those used by me, in the classes that I taught. The way in which

I conducted PA was quite similar to – but not exactly the same as – that used by the other

teachers.

Levine (2008) describes and provides examples of several different PA tools used by other

TBL educators. However, she emphasizes that "many health science educators have encountered

difficulties when attempting to incorporate a peer evaluation program into their TBL curriculum"

(p. 103). At UBCO, all of the teachers initially decided to use a PA tool similar to those

described as the Koles and Texas Tech methods (Levine, 2008, pp. 107-108, 114-116) as these

methods seemed to be the simplest to understand and use. Using the chosen method, students

scored their peers on a number of teamwork-related criteria. To ensure the criteria were

meaningful, the criteria were developed in collaboration with the students at the start of the term.

Each student's PA scores were averaged to give a mark which was worth a portion of the

student's final grade. As described in the Koles method, I initially decided to make PA worth

10% of each student's final grade (see Appendix A). I also followed Michaelsen's (as cited in

Levine, 2008) suggestion to require students to discriminate in their PA scores. In my version of

PA, students awarded each other a score out of 10. The instructions on my PA form (see

Appendix A) stated that "You are required to show some discrimination in your scoring. (That

means that you may not assign 10/10 to all of your teammates)." Since Levine suggests that

students need to learn how to give feedback, I facilitated a session at the start of the term in

which the students decided on the criteria that would be used in the PA. Students were also

required to complete a midterm PA (not for marks) in order to practice this strategy. Students

were required to provide comments to each peer, describing something that the peer was doing

well, in addition to suggesting something that the peer could do to improve his/her teamwork

skills. Lastly, I planned to have students complete the PAs confidentially, using an online

program called iPeer.

I found both benefits and challenges in the first term that I employed PA. I believed the

process in which the students set the criteria for the PA to be a useful exercise, and most of the

comments provided by students to their peers were thoughtful and constructive. However, I

experienced great difficulty with the online iPeer program (due to the absence of technical

support for this program in the 2009/2010 academic year). As a result, I had to modify the PA

process and subsequently required the students to complete the final PA on paper. In the final

PA, I found that students discriminated very little between the contributions of their peers – most

of the students awarded their peers scores of 9 or above (out of 10). In addition, several students

volunteered comments that it was unfair to prevent them from awarding 10 points to everyone in

their team. As a result, a few students ignored this rule entirely. At the end of the term, I was left

feeling somewhat dissatisfied. Although I believed that the process of learning to give peer

feedback was useful to the students, the PA mark inflated the grade of almost every student –

which I did not believe to be appropriate.

The following term, I continued to implement PA in my class, but decreased the value of

the PA mark to 5% of each student's final grade. The basic format of the PA remained the same

– however, there were no restrictions put on the scores that students could award to their peers.

At the end of this term, the results of the PA were very similar to the previous term. Students

generally provided constructive feedback to their peers, and most students received PA marks

between 4.5 and 5 (out of 5). A few of the students commented that they had difficulty

remembering which students were in their TBL teams – making me wonder how valid their PA

marks actually were. In hindsight, I recognized that I had not allowed sufficient time for students

to work together in their teams (giving them time to develop a clear picture of the contributions

of each their peers), as well, I had not asked students to complete their PAs immediately after a

teamwork assignment (so that they would not forget each peer's contributions).

At the start of the next academic year, I made a few modifications to the PA process. While

the PA was still worth 5% of the final grade, students received the full PA grade for simply

completing both the midterm and final PA forms. Also, at the start of the term, when the PA

form was developed in collaboration with the students, the students decided to make the

'comments' section optional (rather than requiring students to include a positive comment and a

suggestion for improvement). At the end of the term, every student received the full 5% PA

mark, but many of the comments were not as constructive as had previously been the case. While

previous year students provided quite specific feedback to their peers (e.g. "wish you would

speak up more"), this year, students tended to write very general comments (e.g. "great working

with you!"). In hindsight, I recognized that requiring students to provide specific comments (as

had been the case the previous year) led to higher quality student feedback.

By this time, some of the teachers were beginning to doubt the value of PA. Anecdotally,

teachers reported finding the paper version of PA to be quite a bit of work to collect, collate,

calculate, and hand back. Teachers also had questions regarding the validity and reliability of

PA. As a result, the team of teachers I was working with the following term decided not to

implement PA.

Significance of this Major Project

By examining the literature, this project has the potential to help educators evaluate the

pros and cons of having students assess the contributions of their peers to group work. As a

result, educators may be better equipped to make an informed decision as to whether or not to

utilize PA. For those educators who do choose to use PA in their classrooms, the literature may also provide guidance on specific strategies to maximize the benefits of PA while minimizing any disadvantages. Finally, by exploring the best evidence on how students assess the contributions of their peers during teamwork activities, this project may provide guidance not only for educators using TBL, but it may also prove useful to educators who employ other forms of group work in their classrooms, such as problem-based learning, case-based learning, or group projects.

METHODS

A number of criteria were used to select appropriate articles for this literature review. First, the article needed to describe a study or educational evaluation in which data were collected regarding PA implementation in the classroom. Acceptable articles could include either quantitative or qualitative methodologies, case studies, meta-analyses, or literature reviews. Second, the article must have been written since the year 2000. Third, the article needed to focus specifically on PA of students' contributions to a group in a university or college classroom. Fourth, the article had to be written in English. Fifth, the article had to focus on the implementation of PA – not merely a description of a computer program used for PA, or the application of a mathematical formula to calculate PA grades. Initially, a sixth criteria required articles to describe PA in a TBL context. However, since only one article meeting all of these criteria could be located, this last criterion was discarded.

I conducted my initial search in the PubMed database using the MeSH terms "educational measurement" and "peer review." Next, I searched PubMed, all of the available EBSCO databases, and Google Scholar using the terms "peer" and "assessment" or "evaluation." Combinations of "individual contribution" or "individual effort" and "group" or "team" were

also used. Once potentially acceptable articles were identified, I used the database link for

'relevant articles' and 'cited by' to search for more articles. I examined the abstract of each of

these articles to find those that appeared to meet all of the search criteria, yielding a total of 66

articles. However, many of the journal abstracts did not clearly indicate the object of the PA,

such as whether the students were assessing an assignment or essay completed by a single peer, a

project completed by a group of peers, the contributions of peers to a group assignment/project,

or some combination of these three. Therefore, the next step was to skim through the articles to

ensure that each article met my search criteria (above). During this step, I eliminated many

articles, reducing my total to 38. At this point, I scanned the reference lists for additional articles,

repeating the process of reading the abstracts and skimming the articles to ensure they were

relevant. Once a total of 45 high quality articles were identified, no more new articles were

sought out (in order to keep this review within manageable limits).

In order to systematically examine each of the articles, a table was constructed to organize

the data. As the information was entered into the table, the number and names of the various

headings evolved. The final version of the table included the following column headings: study;

type of research; key findings; online vs. paper PA; confidential vs. open PA; type of PA

instrument (categorical vs. holistic); including self assessment (SA) or not; narrative comments

vs. no narrative comments; formative vs. summative assessment; how did students learn PA;

overall PA process; how were grades calculated from PA scores; value of grade for PA; and

formula used to calculate PA grade. The full version of this table was printed and displayed, so

that the data could be examined for patterns or trends. Themes and commonalities were

identified using different coloured highlighter pens. The data from these studies is shown in an

abridged version of this table in Appendix B. The following questions guided the analysis of the studies:

1.   What are the potential benefits or drawbacks of PA?

Primarily qualitative data were used to answer this question, including information from student surveys, student comments on PA instruments, and reports from faculty, staff, and tutors.

2.   How valid and reliable is PA?

Answering this question required a review of mainly quantitative data. To examine the validity of PA scores, it was necessary to find a measure that indicated the extent to which the PA score actually measured a student's contribution to the group's work (e.g. that a student who received an above-average PA mark actually contributed an above-average amount of work – and not that the student was merely popular). Since in some studies, student groups were monitored by tutors or faculty, it was possible to compare tutor marks (of student participation or contribution) with student PA scores. Also, in some studies researchers compared student PA scores to group test or group assignment scores.

Another consideration relating to the validity is the variability of PA scores, since educators often expect that student scores (i.e. assignment marks, test scores, or PA scores) will demonstrate a normal distribution. The variability of students' PA scores in several studies was reported on.

The reliability of PA scores was examined using two measures. In some studies, the agreement between PA scores for a single student was reported (inter-rater reliability). In other studies, the ratings for a cohort of students were examined over time (test-retest reliability).

3.   What PA process and instrument constitute the best practice?

First, one must define how "best practice" will be assessed. From one perspective, an effective PA process and instrument could be identified by reviewing staff and student perceptions of PA. From another perspective, an effective PA process and instrument should produce more reliable or valid results. In either case, it can be seen that this question is closely related to the previous two.

Several variables, or choices relating to how PA could be implemented, were examined in this literature review to determine best practice:

- Online versus paper-based PA instruments

- Holistic versus categorical PA

- Confidential versus open PA

- Narrative comments versus no narrative comments

- Rating versus ranking peers

- Student training and preparation for PA

- Student participation in designing the PA instrument

- Inclusion of SA with PA

4.   What is the most appropriate method for incorporating PA in the calculation of a student's final grade?

Peer assessment scores are often used to calculate (or modify) a student grade. The various methods for using PA to calculate a student grade were described and the strengths and limitations of each method were examined.

RESULTS

The purpose of this literature review was to examine the available research literature relating to the PA of individual students' contributions to group work, with the intention of using

this evidence to revise the PA process currently used by this author. Below, the evidence will be used to identify the potential benefits and drawbacks of PA, examine the validity and reliability of PA marks, investigate which PA assessment process and instrument constitute the best practice, and ascertain the most appropriate method for incorporating PA in the calculation of a student's final grade.

Potential Benefits and Drawbacks of Peer Assessment

The literature on PA reports on a variety of benefits for students as well as a few potential problems. Many of these reports are based on qualitative research methods, including student surveys and comments that students recorded on PA instruments. Some of the reported benefits of PA include that: PAs are a fair way to assign individual marks for a group project; there is a reduction in social loafing; students are encouraged to evaluate their own work (or contributions) more critically; students develop interpersonal skills; students learn and grow from the feedback that they receive; and students are better prepared for the workplace. On the other hand, even though most reports of PA are positive, there are still a number of concerns raised, including: students may feel uncomfortable performing PA; some students may not take PA seriously; some students are worried that their assessments may cause harm to their peers; and in some circumstances, PA has led to dysfunctional student group behaviours.

A fair way to assign individual marks for a group project. The majority of researchers report that students believe PA to be a fair method to assign marks for individual contributions to a group project (Carson & Glaser, 2010; Divaharan & Atputhasamy, 2002; Elliot & Higgins, 2005; Jin, 2011; Kench, Field, Agudera, & Gill, 2009, Shiu, Chan, Lam, Lee, & Kwong, 2011; Sluijsmans, Moerkerke, Merriënboer, & Dochy, 2001; Steensels, Leemans, Buelens, Laga, Lecoutere, Laekeman, & Simoens, 2006; Willey & Gardner, 2009). However, two studies

(Kennedy, 2005; Malcolmson & Shaw, 2005) described concerns by either students or teachers that the PA process was not fair. Malcolmson and Shaw suggested a few reasons that students might have perceived PA to be unfair, including that (1) the PA mark was only worth a small portion of the final grade (i.e. that the effort put into PA was not worth the results), (2) there was little discrimination among the PA scores (so there was little if any impact on student grades), and (3) there seemed to be a cultural influence in how some students perceived PA. For example, Malcolmson and Shaw's study was conducted at the University of Aukland's School of Pharmacy. In a focus group, they found that several of the students who had not been born in New Zealand seemed to dislike assessing each other more than the students who had been born in New Zealand.

Sivan (2000) found that student perceptions of PA's fairness related to their previous experience with PA. Sivan found that although all students recognized the value of PA, the inexperienced students had more concerns about whether the process was fair. The more experienced students remarked that they felt more confident in their ability to assess their peers.

Reduction in social loafing. Many students and teachers are wary of group-work assignments, because of concerns that some students do not do their fair share of the work. This phenomenon is known by a variety of names, including *social loafing*, *free-riding*, and *free-loading*. In their meta-analytic review of the literature, Karau and Williams (1993) define social loafing as "the reduction in motivation and effort when individuals work collectively compared with when they work individually or coactively" (p. 681). They found that people often engaged in some level of social loafing, regardless of their gender, culture, or the nature of the task being performed. The model that they developed suggested that "social loafing occurs because individuals expect their effort to be less likely to lead to valued outcomes when working

collectively" (p. 700). In particular, people are more likely to engage in social loafing when their individual contributions cannot be evaluated, when it is not possible to compare their efforts to those of the other individuals in their group, when they work with strangers, when they expect their peers to perform well, and when the task is not particularly meaningful to them (p. 700). Several of these factors can be mitigated by the manner in which group activities are planned and executed by a teacher. For example, group assignments can be designed in such a way as to be meaningful (e.g. solving realistic problems that are relevant to a student's course of studies). Teachers can also form groups in a manner that will reduce the likelihood that students will work with strangers: either students may be allowed to self-select themselves into groups (so that they may work with acquaintances), or opportunities may be provided for groups to work together for an extended period of time. The process of PA has the potential to influence two of Karau and Williams' factors by providing a method to assess the contributions of individual students within a group and by allowing students to compare their own performance against the performance of the average student in their group.

A number of studies have found that students believe that PA reduces the incidence of social loafing (Brooks & Ammons, 2003; Elliott & Higgins, 2005; Shiu et al., 2011; Sivan, 2000; Weaver & Esposto, 2011; Willey & Gardner, 2009; Willey & Gardner, 2010). However, some of Shiu et al.'s student participants commented that the "peer assessment system has no influence to those lazy students" ("Improving the quality of teamwork," para. 1). Interestingly, Kruck and Reif (2001) found that while about 21% of their student participants reported that the PA did not motivate them to contribute more to their group, about 36% of their participants believed that PA motivated their peers to contribute more. Friedman, Cox, and Maher (2008) found that those

students who were concerned about potential problems with social loafing were highly motivated to rate their peers.

Brooks and Ammons (2003) found that the more their student participants thought PA reduced the occurrence of social loafing, the more they believed that team projects were a good way to learn. The researchers also found that student participants believed social loafing was reduced when the PA included specific feedback, was performed early in the group project, and was conducted multiple times per term.

Evaluating student work more critically. Several researchers suggest that SAs and PAs that teach students to evaluate their work critically may lead them to become more effective learners (Divaharan & Atputhasamy, 2002; Sivan, 2000; Willey & Gardner, 2009; Willey & Gardner, 2010). Divaharan and Atputhasamy's student participants reported that PA "encouraged them to be more responsible for their own learning, thereby further developing their higher-order thinking skills, by being more critical of themselves and their peers" (p. 77).

Developing interpersonal skills. In some cases, students who engage in PA may be asked to write narrative comments about their peers' contributions. In other cases, students must work together as a group to negotiate, and then agree on, each student's PA score. In both of these situations, students must utilize interpersonal skills to communicate their feedback to their peers. As a result, many students report that PA helps them to improve their interpersonal skills (Divaharan & Atputhasamy, 2002; Shiu et al., 2011; Willey & Gardner, 2010).

Papinczak, Young, and Groves (2007a) found that PA helped their student participants recognize how they depended on each other and were responsible for each other in their teams. Similarly, Pocock, Sanders, and Bundy (2010) found that their student participants learned how to persuade poor contributors to pull more of their weight in the group.

Willey and Gardner (2009) discovered that students may experience the most benefit from PA when there is at least one weak or under-performing student in their group. In this study, student participants submitted SAs and PAs online, and then met together in their tutorial groups to discuss the feedback. They found that student participants in groups with at least one poor team member reported their ability to give and receive feedback improved (compared with student participants who did not have any poor team members in their groups). The student participants in well functioning teams often had little if anything to discuss in the feedback sessions and, therefore, did not receive the same benefit.

Learning from feedback. Some versions of PA provide the opportunity for students to share specific feedback with their peers (either verbally or in writing). This feedback may provide useful suggestions that could help students to become more effective team members or possibly improve the quality of their work (Ferguson & Kreiter, 2007; Papinczak et al., 2007a; Pocock et al., 2010; Willey & Gardner, 2009; Willey & Gardner, 2010).

Preparing for the workplace. Learning to perform PA in school can help to prepare students for a future career (Chaves, Baker, Chaves, & Fisher, 2006), particularly in professional programs, such as nursing, medicine, education, or management. Some of Sivan's (2000) student participants commented that the PA they performed in class was similar to the assessments that they were required to complete in the workplace (in the hotel and tourism industry).

Peer assessment can feel uncomfortable. A number of students report that having to judge the contributions of their peers felt awkward or uncomfortable (Divaharan & Atputhasamy, 2002; Papinczak et al., 2007a; Pocock et al., 2010; Sluijsmans et al., 2001). In some cases, teachers require students to discriminate among their PA scores (e.g. some students must be given a lower score and other students must be given a higher score). Often, students

report that they dislike being *required* to discriminate (Ferguson & Kreiter, 2007; Levine et al., 2007). Sluijsmans et al. (2001) suggested that students who have less experience with group work and PAs are likely to feel more uncomfortable with these unfamiliar strategies than students who have experience with a wide variety of instructional strategies. Willey and Gardner (2010) found that a small number of their student participants believed that students should not be involved in assessing each other's work. Some of the reasons included that the student participants did not believe they were qualified to assess their peers, while others believed it was their tutor's responsibility: "That is what we are paying them to do" (p. 440).

However, Chaves et al. (2006) remark that the student participants in their study (Master of Science in Nursing students) would be working in leadership roles on graduation, and that "as master's prepared nurse managers, providing negative feedback to nursing staff, medical colleagues, and patients is a critical responsibility" (p. 31). Therefore, learning to give negative feedback, although uncomfortable, is an important skill to learn.

Peer assessment may not be taken seriously. In a few studies, students reported not taking PA seriously. For example, Papinczak et al. (2007a) reported that some student participants did not feel that the criteria on the PA instrument were relevant or felt apathetic about the process and simply awarded the same score to all peers.

Negative peer assessments might cause harm. Students who were surveyed about their perceptions of PA sometimes commented that they were concerned that giving negative feedback or a low PA score to their peers might harm their relationships with their teammates (Chaves et al., 2006; Papinczak et al., 2007a; Willey & Gardner, 2010). In a problem-based learning setting, Papinczak et al. reported that both students and tutors were concerned that "the 'family atmosphere' may be compromised by peer evaluation" (p. 180). Student participants were also

worried that a low PA mark might have a negative impact on their peers' grades (Chaves et al. 2006; Kennedy, 2005). Elliott and Higgins (2005) found that while their student participants had no difficulty giving lower PA scores for those students who were perceived to be social loafers, the student participants were "reluctant to down-grade individuals who they knew had personal problems which reduced their contribution to group work" (p. 45).

Jin (2011) found that a minority of their student participants were concerned that their PA scores would not be fair, because "students would mark unfairly due to their like or dislike of the others" (p. 8). Other researchers have reported this same concern among students (Kench et al., 2009; Kennedy, 2005). However, the ability of a minority of students to skew the results of another student's peer rating is not a problem commonly reported in the PA literature. In most cases, educators use a formula to calculate PA scores that averages the ratings awarded by all of the team members (e.g. Kilic & Cakan, 2006), minimizing the effect that a single student can have on another peer's grade. However, Kench et al. suggest developing an appeal process for any students who receive a poor PA score. Magin (2001) examined the reciprocity effects in the PA scores of 169 medical students. He explained that reciprocity involves students who award marks based on their like or dislike of a peer, as well as collusion (in which two or more students may conspire to give each other higher marks). In this study, Magin found that the reciprocity effect was "quite small, accounting for only 1% of the variance in peer scores" (p. 60).

Dysfunctional group behaviours. Kennedy (2005) reported that when he implemented PA in his course, some student participants engaged in dysfunctional behaviours in order to maximize their own PA score at the end of the term. In particular, he was concerned that some student participants would "dominate the group and manipulate tasks to their own advantage" ("Dysfunctional effects of peer assessment", para. 1). Weaker students might be assigned less

important tasks, leading to less learning and lower PA scores at the end of the term. Shiu et al. (2011) also reported that "some 'aggressive' group members would undertake additional tasks in order to obtain higher PA ratings" ("Improving the quality of teamwork", para. 1). In addition, Pocock et al. (2010) noticed that some of their student participants were very strategic in how they participated in group work. By merely attending meetings and contributing minimally to group discussions, they were assured of meeting two of the PA criteria – enough to pass the assessment. As a result, teachers should design PA instruments carefully, with these potential challenges in mind.

In summary, research shows that there are both benefits and potential drawbacks to PA. Therefore, educators should carefully develop a PA process that will achieve intended goals – such as ensuring students are held accountable for their contributions to group activities – while minimizing the potential drawbacks of PA, such as poorly thought-out assessment criteria that may lead some students to engage in dysfunctional group behaviours.

Validity and Reliability of Peer Assessment Marks

Before deciding to invest time and effort into PA, most teachers want to be assured that PAs are a valid and reliable way to assess student contributions to group work. Researchers have examined PA scores in five main ways to determine validity and reliability: (1) comparing the PA ratings for each student, (2) examining how PA scores change over time, (3) comparing PA scores with staff/faculty scores, (4) comparing PA scores with assignment or test grades, and (5) examining the range of PA scores awarded within a class.

One particular challenge is that tests for validity and reliability are typically applied to a single instrument – rather than a whole class of instruments. And since there are literally dozens of different PA instruments described in the literature, it is impossible to state whether all PA

instruments are valid and reliable or not. Therefore, the purpose of this section is merely to investigate whether it is possible for PA to produce results that educators would trust to use for student grades.

Comparison between peer ratings for a single student. If PAs are to be considered reliable, one would expect that students who contribute the most to their groups should consistently get above-average PA scores, and low contributors should consistently get below-average PA scores (inter-rater reliability). If students cannot agree on these high and low-contributing students, then a teacher is likely to doubt the reliability of the PA instrument.

Two studies reported a good level of consistency among peer ratings for each student. Kamp, Dolmans, Van Berkel, and Schmidt (2011) developed a PA instrument (the M-PARS) that was tested and found to have very good inter-rater reliability. They stated that the M-PARS provided reliable results with a minimum of four peer ratings per student. Johnston and Miles (2004) also found a high level of agreement on PA scores within the student groups in their study.

Two studies reported that different student groups demonstrated varying levels of consistency when scoring their peers. Steensels et al. (2006) found that there was a high level of agreement in student PA scores within most student groups. However, one student group showed a much higher standard deviation between the PA scores. In this case, a tutor who had observed the group noticed that two of the students in the group demonstrated dysfunctional behaviours. Those students skewed the PA scores in their group by giving each other high marks while the other group members gave them low marks. Papinczak et al. (2007b) found that some student groups were able to achieve much higher correlations between PA scores than other groups. The

majority of groups who had the highest correlations expressed that they were committed to PA, while the groups with lower correlations in PA scores expressed negative views of PA.

Weaver and Esposto (2011) found that their student participants were very consistent in rating the high-performing students, but there was less consistency when rating the low-performing students.

The research described above shows that it is possible for PAs to demonstrate good inter-rater reliability. Inter-rater reliability may be highest in student groups who strongly believe the PA process is useful and in the scores for high-performing students.

Changes in peer assessment scores over time. Attempts to examine changes in PA scores over time (test-retest reliability) proved challenging, since the use of measurement scales in PA instruments was not consistent across studies. Some researchers (e.g. Divaharan & Atputhsamy, 2002) required student participants to score their peers on a simple scale (e.g. a scale of 0-10), while other researchers (e.g. Carson & Glaser, 2010) required student participants to divide a pool of marks among all the members of a team. Among researchers using the former method, successive episodes of PA might result in different mean PA scores. However, in studies using the latter method, every student group will always have the same mean PA score.

In situations where student participants assessed their peers on a scale, three studies showed that students' mean PA scores increased with successive episodes of PA (Brutus & Donia, 2010; Chaves et al., 2006; Machado, Machado, Grec, Bollela, & Vieira, 2008) while one study (Steensels et al., 2006) showed that the students' mean PA scores did not increase over two episodes of PA.

Comparison between peer and staff assessments. The validity of PA scores is another important consideration for teachers. When assessing student contributions to group work, one

would expect that the PA score should accurately reflect each student's contribution. In the case

of small group work, it is the students who would usually have the best knowledge of their peers'

contributions. Only when there is a tutor for a small group of students (e.g. in a problem-based

learning group) would a staff member have an accurate picture of individual students'

contributions. If PA scores are a valid way to assess student contributions, one would expect that

student PA scores should be very similar to tutor assessment scores – so long as both students

and tutors are using the same instrument to assess students' contributions.

A number of researchers have found significant correlations between student PA scores

and tutor or lecturer scores (Ferguson & Kreiter, 2007; Heyman & Sailors, 2010; Sahin, 2008;

Steensels et al., 2006; Weaver & Esposto, 2011) However, in some studies, students rated their

peers higher than the tutor rating (Chaves et al., 2006; Machado et al., 2008; Papinczak et al.,

2007b), while in other studies, students rated their peers lower than the tutor rating (Sahin,

2008). Ferguson and Kreiter also found that student comments on the PA instrument were very

similar to faculty comments, although the peers often provided an additional perspective. This

similarity between peer and staff/faculty assessments supports the validity of PA.

Comparison between peer and assignment/test grades. Unfortunately, as mentioned

above, tutors are not always available to supervise small group activities – so it is often not

possible to compare student PA scores with tutor scores. As a result, some researchers have

compared student PA scores with other evaluation scores, such as exam, project, and assignment

grades (e.g. Levine et al., 2007). Since exams usually measure knowledge, while (in this context)

PAs measure student contributions, the correlation between these scores may not be as close as

comparing student and tutor assessments. However, it is not unreasonable to assume that many

of the students who do well on exams will also contribute well in groups – so one would think

that there might be a correlation between PA scores and marks from other classroom evaluation methods. In fact, several researchers have found a significant correlation between PA scores and test grades (Johnson & Miles, 2004; Kaufman, Felder, & Fuller, 2000; Levine et al., 2007).

Variability of peer assessment scores. Most educators expect that classroom evaluation methods will result in some sort of distribution of grades, preferably resembling a normal curve. An evaluation tool that does not differentiate among students is of little use in this regard. In the PA literature, the variation in PA scores is often described in terms of whether student participants are able (or willing) to discriminate in their PA scores. Some studies show that the PA scores of student participants show a wide distribution (Kilic & Cakan, 2006; Sahin, 2008). Other studies show that the distribution of PA scores varies from group to group (Saito & Fujita, 2009). Other studies have found that student participants award a very narrow range of PA scores (Drexler, Beehr, & Stetz, 2001)

Many researchers report that a large proportion of their student participants awarded the same PA score to all of their peers. Carson and Glaser (2010) found that about 50% of their student participants awarded all of their peers the same score. In other studies, this proportion has been reported as 35% (Pocock et al., 2010), 40 % and 49% (Kaufman et al., 2000), 44% (Kennedy, 2005), and 66% (Drexler et al., 2001; Malcolmson & Shaw, 2005).

Some researchers have found that the variability in PA scores may be related to the PA method used. When Heyman and Sailors (2010) asked their student participants to award peers a simple score (on a scale of 0-50), they found that the average student PA score varied less than faculty assessment scores. However, when they used a PA instrument that asked student participants to nominate the five highest and five lowest contributors in the class, the range of PA scores was even greater than the faculty assessment scores. Raban and Litchfield (2007) also

found that the variability of PA scores depended on the method used. During the period of time

that student participants relied on their own records to keep track of their peers' contributions,

between 75-90% of groups awarded all members the same PA score. When Raban and Litchfield

implemented an online system that allowed student participants to track the time spent working

on their group project, the number of groups awarding all members the same score dropped to

55%. Then, when Raban and Litchfield had student participants use an online PA instrument that

not only allowed the student participants to track the time spent working on the group project,

but also allowed them to give each other weekly (formative) ratings, they found that the number

of groups awarding all members the same score was further reduced to 20%.

There may also be a difference in students' willingness or ability to differentiate among the

contributions of their peers, based on their own level of performance. Weaver and Esposto

(2011) found that high-performing students awarded a wider range of scores to their peers than

low-performing students.

Additionally, the variability of students' PA scores seems to be interpreted differently by

various researchers. Some researchers are of the opinion that when students award similar PA

scores to all peers, then all students are contributing equally to the group and the group is

functioning well (Brooks & Ammons, 2003, Carson & Glaser, 2010, Russell et al., 2006;

Johnston & Miles, 2004). Other researchers suggest that it is normal and expected that the

contributions of students will vary in quantity and quality. Therefore, students who truthfully and

accurately assess their peers' contributions should award a wide range of PA scores (Raban &

Litchfield, 2007).

In summary, it can be seen that the variability of PA scores may be affected by a number of

factors, including the type of PA instrument used and the ability level of the student. Many

researchers have found that a large proportion of students tend to award their peers similar scores. It is unclear whether students award similar scores because they are unable or unwilling to differentiate between the contributions of their peers – or because they truly believe that all their peers contributed equally to the group work. The meaning and desirability of variability in PA scores is also unclear, with some researchers interpreting low variability as evidence of student cooperation, while other researchers interpret greater variability as increased student competence in PA.

What Peer Assessment Process and Instrument Constitute the Best Practice?

This question first required key variables in the PA process to be identified. These variables included: (1) whether the PA was conducted on paper or online; (2) whether the PA instrument was holistic or categorical; (3) whether the PA was completed confidentially or openly; (4) whether the PA instrument included narrative comments or not; (5) whether students were asked to rank or rate peers; (6) the amount and type of training and preparation for PA; (7) whether students participated in the design of the PA instrument; and (8) whether SA was included in PA.

Second, three criteria were chosen to describe how 'best practice' would be evaluated. Because the first question guiding this literature review identified that there were both benefits and drawbacks to PA, the first criterion focused on which type of PA process/instrument resulted in the most benefits and had the fewest drawbacks. Because validity and reliability were recognized as important concerns (as discussed above), the second criterion examined which PA process/instrument produced the most valid and reliable PA scores. Lastly, since many researchers gathered qualitative data on student and staff perceptions of PA, the third criterion

considered which PA process/instrument resulted in the most positive perceptions among students and staff.

Online versus paper-based peer assessment instruments. Most PAs are completed on paper or online. Paper-based methods are common, as no computer programs or technical support are necessary, but are often reported to require lots of time and effort (e.g. Thompson & McGregor, 2009). Since each student completes an assessment of multiple peers, a single class of 100 students may generate several hundred PA forms. When grading paper-based PAs, the teacher may either need to calculate a number of scores manually or input PA scores into a spreadsheet program (such as Microsoft Excel). On the other hand, online methods of performing PA are becoming more common. Although email may be used to submit PA forms (Sahin, 2008; Steensels et al., 2006), most researchers report using specifically designed computer programs. These programs generally permit SA and/or PA to be submitted confidentially online, calculate PA scores automatically, and allow students to view their own resulting scores. Some of the programs described in the literature include PES (Brutus & Donia, 2010), *CIITN* webtool (Carson & Glaser, 2010), PBL-Evaluator (Chaves et al., 2006), SPARK and SPARK[PLUS] (Thompson & McGregor, 2009; Willey & Gardner, 2009; Willey & Gardner, 2010), and TeCTra (Raban & Litchfield, 2007).

Several studies show that both teachers and students are satisfied with online PA tools. For example, Brutus and Donia (2010) reported that 92% of the 35 professors using PES had positive opinions of this system. Similarly, Raban and Litchfield (2007) reported that the TeCTra program was effective and simple for staff and students to use. Thompson and McGregor (2009) described having much more success with an online PA instrument (using the SPARK program)

than with their previous paper-based methods. The PA process was taken more seriously, and

both staff and students were satisfied with the online method.

When PA is meant to be confidential, some students express concerns with paper-based PA

instruments (Shiu et al., 2011; Sivan, 2000). In Shiu et al.'s study, some student participants

reported concerns that their peers might have seen their PA forms and, as a result, might have

retaliated by awarding them lower PA scores. The student participants highly recommended

using an online method of PA.

In summary, a number of studies reported that both staff and students are satisfied with

online methods of PA. Although paper-based methods can certainly be used effectively (e.g.

Elliott & Higgins, 2005), some studies reported that teachers found paper-based methods to be a

lot of work (e.g. Thompson & McGregor, 2009), and students were concerned that paper-based

methods might compromise confidentiality (Shiu et al. 2011).

Holistic versus categorical peer assessment. A variety of different PA instruments have

been reported on in the literature. Most of these instruments can be broadly described as either

holistic or categorical. In a holistic PA, students assign a *single* score to a peer (i.e. on a scale of

0-10). There may or may not be a list of criteria to base the assessment on – but students are not

marked on the individual criteria. In contrast, a categorical PA consists of multiple criteria, from

as few as 2, to as many as 55. Students are marked on each criterion, and then the score is

usually totalled.

Several studies have compared holistic and categorical assessments. In their meta-analysis,

Falchikov and Goldfinch (2000) examined the correlations between PAs and faculty

assessments. The highest correlations were found when student participants assessed their peers

using a holistic instrument that listed several dimensions or criteria. Holistic instruments that did

not include any criteria had slightly lower correlations, while categorical instruments had the lowest correlations. Similarly, Ohland, Layton, Loughry, and Yuhasz (2005) compared the inter-rater reliability of three different PA instruments: a one-item assessment (holistic), a one-item assessment with behavioural anchors (holistic assessment with criteria listed), and a categorical assessment. They found that the one-item assessment with behavioural anchors provided the highest level of inter-rater reliability. Lejk and Wyville (2001a) required student participants to complete PA using both categorical and holistic PA instruments. They found that there was greater inter-rater reliability among peer scores with the holistic assessment than with the categorical assessment. They also found that when students' group marks were adjusted by the PA score, the holistic assessment produced a wider range of marks than the categorical assessment.

Two studies found that students prefer holistic assessments, particularly when specific criteria are listed. Lejk and Wyvill (2002) administered surveys to student participants in two different classes. In the first class, students assessed the contributions of their peers using a holistic PA instrument, while in the second class, students assessed the contributions of their peers using a categorical PA instrument. They found that the students who used the holistic PA instrument were more positive about PA. Shiu et al. (2011) reported that their student participants were asked to award a single score (holistic assessment) to peers based on four criteria (corresponding to four project tasks). Afterwards, the majority of student participants recommended keeping the holistic assessment, but adding additional criteria to "better guide students in rating their peers" ("PA criteria", para. 1).

Friedman et al. (2008) examined the motivation of student participants to perform PA in four different experimental conditions: (1) holistic assessment, three times during a course, (2)

holistic assessment, once during a course, (3) categorical assessment, three times during a course, and (4) categorical assessment, once during a course. They found that student participants reported being the most motivated to perform a holistic assessment multiple times (and the least motivated to perform categorical assessments multiple times).

A number of students have reported dissatisfaction with the assessment criteria when using categorical assessments to score peers (Sluijsmans et al., 2001; Papinczak et al., 2007a; Sivan, 2000). In Pocock et al.'s (2010) study, student participants reported that scoring their peers on the criteria in their PA instrument did not appropriately account for the amount of work put into the project by individual group members. Lejk and Wyvill (2001a) suggest that there may be some drawbacks to certain kinds of categorical assessments. In categorical PA, students are often asked to rate their peers on a variety of criteria, and each of the criteria usually contribute equally towards the total PA score. However, it may not be accurate to assume that all criteria are equally important in evaluating a student's overall contribution to the group. In addition, when groups divide the work for a group project, some students may not have an opportunity to contribute equally to each of the criteria listed on the PA instrument, even when those students do their fair share of the work. As a result, Lejk and Wyvill suggest that categorical assessments may be very helpful for providing formative feedback, while holistic assessments may be more appropriate for giving an overall impression of a student's contribution to a group effort.

Overall, the evidence supports the use of holistic PA (particularly holistic PA that includes criteria/behavioural anchors) over categorical PA. When using a holistic PA, student PA scores show a higher level of inter-rater reliability than when using a categorical assessment. Students also prefer holistic PA instruments over categorical PA instruments and are more motivated to perform holistic PA.

Confidential versus open peer assessments. Both confidential and open methods of PA are reported in the literature. Most researchers describe having their student participants submit PAs in a confidential manner (Elliott & Higgins, 2005; Friedman et al., 2008; Heyman & Sailors, 2010; Johnston & Miles, 2004; Kaufman et al., 2000; Kench et al., 2009; Levine, 2007; Malcolmson & Shaw, 2005; Reiter, Eva, Hatala, & Combes, 2002; Saito & Fujita, 2009; Steensels et al., 2006; Weaver & Esposto, 2011). However, a few researchers use a process in which student participants are openly aware of their peers' assessments. In some cases, the whole group must agree on the PA score for each student (Divaharan & Atputhasamy, 2002; Drexler et al., 2001; Raban & Litchfield, 2007), while in other cases, student participants complete a PA for each of their teammates, but share and discuss the feedback openly (Pocock et al., 2010; Willey & Gardner, 2010).

Russell et al. (2006) discussed some of the pros and cons of blind (confidential) versus open PAs before deciding to use a confidential assessment process with their student participants. They suggested that in a confidential assessment, quieter students may have more of a voice, and students might be more likely to give honest assessments when they are not afraid of being confronted about negative feedback. On the other hand, in an open assessment, sharing feedback openly may provide an opportunity for dialogue within a group, and students may have a chance to reply to feedback and explain their actions. Saito and Fujita (2009) recommend keeping PAs confidential in order to avoid disrupting student relationships within the groups.

In some studies, students were asked to complete paper-based PA forms and then hand them in during a class or tutorial. Some student participants expressed concerns with the confidentiality of this method, being worried that the assessments may be seen by their peers (Papinczak et al., 2007a; Shiu et al., 2011). One of the student participants in Sivan's (2000)

class remarked, "To ensure a fair assessment, I think it's better for students to hand in or mark

the peer assessment form at home or anywhere to keep them confidential 'cause it's difficult to

mark down someone who sits next to you" (p. 202).

Lejk and Wyvill (2001b) required student participants to complete a secret (confidential)

PA near the end of a group assignment. The following week, students were required to produce

and submit a grid containing PA scores for all of the group members. One finding from this

study was that there was a greater level of agreement among student scores in the open PA but

students differentiated to a greater extent in the confidential PA. Drexler et al. (2001) also found

that in their open PA, only 34% of groups chose to differentiate among PA scores on either

assignment. Even most of those groups who did differentiate between PA scores did so within a

very narrow range. Interestingly, Drexler et al. found that the groups who did differentiate in PA

scores earned significantly better grades on their second group project (after receiving peer

feedback), while the groups that did not differentiate had more positive attitudes towards their

group and the grading process.

Pocock et al. (2010) employed an open PA process in which student participants completed

PA forms near the end of a group project and then met to discuss these assessments. Students

reported that the process of sharing PA scores and justifying the mark helped them to become

more skilled and confident in presenting a rational argument and in negotiation. Students also

reported that they would not have given different PA scores if the process were confidential.

Overall, Pocock et al. reported that this PA method was popular with both teachers and students.

In summary, the evidence is not clear on whether open or confidential PA is superior.

Although many researchers state that it is important to keep PA confidential (e.g. Kench et al.,

2009), and most research studies describe confidential methods of PA, open methods of PA have been used successfully (e.g. Pocock et al., 2010).

Including narrative comments versus no narrative comments. One of the potential benefits of PA is that it may provide useful feedback to students that could help them to become more effective group members. One of the strengths of an open negotiated PA process is that students engage in a dialogue and verbally share feedback with each other. However, even when PA is submitted confidentially, methods are available (such as computerized PA programs) that allow students to provide anonymous narrative comments. Some researchers specifically note that it may be beneficial for students to provide narrative comments when assessing their peers, so that students can learn from the feedback (Brutus & Donia, 2010). Student participants in Papinczak et al.'s (2007a) study reported valuing the constructive feedback that they received from their peers. Brooks and Ammons (2003) found that their student participants "perceived that the free-rider problems were reduced when evaluations that provided specific feedback were conducted early in a project and several times during that project" (p. 271). Sluijsmans et al.'s (2001) student participants reported being very dissatisfied with only giving a numerical PA score. They wanted an opportunity to give instructional feedback, particularly when giving a peer a below-average score. Ferguson and Kreiter's (2007) study, in which students worked in a group with a tutor, found that the comments of peers were quite similar to the tutor's comments – but the peer comments often provided an additional perspective.

At times, students may be asked to write a narrative comment that is directed to the teacher, rather than to the student's peers. These comments are usually intended to explain or justify the numeric score awarded to the peers. Russell et al. (2006) found that the students who were more committed to the group project wrote longer and more specific justification

statements. However, they do not mention whether there was a correlation between the length of the justification statements and PA marks (or overall grades). Similar to Russell et al., Weaver and Esposto (2011) noticed that high-performing students wrote more detailed comments about their peers than the low-performing students.

Overall, the above studies show that incorporating narrative comments into PA is preferable to not allowing narrative comments. Whether provided in an open or confidential manner, these comments may be used to provide constructive feedback to peers or to justify PA scores to the teacher.

Rating versus ranking peers. When students are asked to assess their peers, either a rating or a ranking system can be used. Rating one's peers (which is much more common) involves assigning a numerical score within a defined set of parameters. For example, students may be asked to rate their peers' contributions on a scale of 0-10. This type of rating could be classified as an *interval* scale, which permits several mathematical operations such as multiplication/division or addition/subtraction. In contrast, ranking involves putting something in order. In some cases, students are asked to rank their peers on a scale from lowest contributor to highest contributor, while in other cases, students are asked to rank several characteristics in each peer. For example, Reiter et al. (2002) asked their student participants to rank seven characteristics (responsibility, facilitation, respect for others, positive participation, knowledge, critical analysis, and communication) in each of their peers. Each peer would have a strongest and weakest characteristic. Ranking scales are therefore *ordinal*, in that although they can be ordered, the difference between each ranked item is not necessarily of the same magnitude. As a result, it can be challenging to use a ranking scale to calculate a student's grade.

Reiter et al. (2002) found that ranking students' characteristics (described above) was not reliable or valid. There was poor inter-rater reliability between the peer rankings, and peer rankings correlated poorly with the tutor's rankings.

Heyman and Sailors (2010) used a different method for student ranking. They asked their student participants to nominate the five classmates who were the highest contributors and the five classmates who were the lowest contributors in the class (of 33 students). Students would receive one point every time a peer rated them as a high contributor, and they would lose one point every time a peer rated them as a low contributor. To calculate each student's PA score, the teacher counted up the number of points that each student received from his/her peers. The PA scores ranged from +16 points to -14 points. The researchers found that this method produced a wide distribution of student rankings (a slightly wider range than the teacher assessments), and the order of the rankings correlated well with the teacher's assessments. Unfortunately however, the researchers did not explain how the PA rankings contributed to the students' overall grades or evaluations.

Overall, PA rating scales are used more commonly than rankings. PA scores obtained from rating scales are versatile in that they may easily be incorporated into a mathematical formula that can be used to calculate a student's group mark. The ranking of student characteristics was not found to be a reliable method of PA. However, asking students to rank the high and low contributors in a group may provide a novel, yet valid and reliable method of PA.

Student training and preparation for peer assessment. A number of researchers strongly recommend that students receive adequate orientation and training prior to performing PA (e.g. Cheng & Warren, 2000). Such training might include a variety of topics, such as the value of feedback in general, instruction on how to provide constructive feedback, education

about PA, training on teamwork and group functioning, practice in completing PAs, and so on.

Shiu et al. (2011) found that student participants doubted their own ability to objectively assess

others, and as a result, they recommended that students should be provided with formal training

on PA.

The quality of PA training is also important in and of itself. Thompson and McGregor

(2009) suggest that the way in which a teacher explains the PA process may make a significant

impact on how students view PA. "Where staff were not confident explaining the self and peer

assessment process student groups tended to treat the system in a 'surface' manner" (p. 444).

Some researchers suggest that student PA scores are more valid and reliable when students

receive adequate training on PA. Sluijsmans et al. (2001) studied the PA scores in two different

classes of students. They found that the reliability of student participants' PA scores was

different in each of the two classes, even though the same PA instrument was used in the two

classes. One of their conclusions was that students needed training in PA to reduce rating errors.

In Sahin's (2008) study, student participants contributed to the development of the PA

instrument and had the opportunity to practice PA once prior to the final summative PA.

Afterwards, Sahin found a high correlation between the lecturer and peer scores. Steensels et al.

(2006) concluded that since student participants showed more ability to discriminate in their PA

scores with practice, PA should be implemented early in an educational program (allowing

students to practice PA over the subsequent years). Falchikov and Goldfinch (2000) found that

there was a higher correlation between peer and faculty assessment scores in well-designed

research studies than in the poorly designed studies. They concluded "if the studies rated as low-

quality also involved less-than-clear implementation, then it is understandable that students may

have been confused about important elements of the exercise" (p. 316). This statement supports

the need to adequately train and orient students to perform PA. Overall, the evidence supports the need to teach students how to assess their peers, as well as the importance of orienting students to the PA process that will be used in class.

Student participation in designing the peer assessment instrument. In some studies, researchers went beyond merely orienting their students to PA and involved their student participants in selecting the criteria for the PA instrument. In their meta-analysis, Falchikov and Goldfinch (2000) found that there was a higher correlation between peer and faculty assessment scores when students were involved in selecting the criteria in the PA instrument. When Elliott and Higgins (2005) involved their student participants in designing the PA instrument, the majority of students reported that the PA process was fair, and that students were more motivated to participate in the group work. However, not all researchers have received positive reviews regarding PA from their student participants. Even after being involved in setting the PA criteria, some student participants report feeling uncomfortable assessing their peers (Sluijsmans et al., 2001) or believing that PA was a lot of effort without much benefit (Malcolmson & Shaw, 2005). Sivan (2000) suggested that students require time and repeated practice with PA to become comfortable and proficient. After studying PA in five separate classes (student participants in each of the classes had varying amounts of experience with PA), Sivan stated "it is not only the experience which contributes to student confidence and acceptance of the [PA] method, but also preparation for the use of the method and students' participation in criteria setting" (p. 200). The students with the most experience of PA made many comments on the value of being allowed to set the PA criteria.

Overall, the evidence indicates that involving students in designing the PA instrument leads to greater student acceptance and motivation. However, research also shows that students

are concerned about spending their class time wisely. Therefore teachers should carefully

balance the time and effort spent on PA with the time and effort to achieve course learning

objectives.

     Inclusion of self assessments in peer assessments. A number of the educators who

employ PA also require students to assess themselves. Research into combined PA and SA has

shown mixed and contradictory results. Some researchers speculate that these conflicting

findings may be related to gender differences (Chaves et al., 2006), cultural differences, or

age/maturity of the students (Machado et al., 2008).

     Some studies have compared SA scores with assignment grades (awarded by the teacher).

Johnston and Miles (2004) found that while there was a correlation between PA scores and

assignment scores, there was no correlation between SA scores and assignment scores, making

them wonder if SAs are less valid than PAs.

     Johnston and Miles (2004) also found that the overall trend was for student participants to

over-rate their own contributions to the group [SA scores > PA scores]. However, when they

looked at the student participants with the highest and lowest assignment grades, they found that

the student participants in the highest quartile over-rated themselves while the student

participants in the lowest quartile under-rated themselves. This finding is in contrast with both

Thompson and McGregor (2009) and Lejk and Wyvill (2001b), who found that high-performing

student participants tended to under-rate themselves on SA, while the low-performing student

participants tended to over-rate themselves on SA.

     Several other researchers have compared SA scores to PAs and tutor assessments. In some

studies, no significant difference was found between the SA and PA scores [SA score = PA

score] (Kaufman et al., 2000; Machado et al., 2008). In other cases, students rated themselves

higher than the course tutor rated them, but lower than their peers [PA score > SA score > tutor score] (Chaves et al., 2006). Still others have found that students rated themselves lower than the course tutor [tutor score > SA score] (Papinczak et al., 2007b).

Some researchers have found that when SAs are included in the PA process, a small number of student participants may try to *game the system* (Ohland et al., 2005). Ohland et al. reported that "some students attempted to skew the ratings in their favour by rating themselves high and their teammates low" (p. 322). As a result, it is important for teachers to examine student SA and PA scores to ensure that students are not trying to unfairly increase their own grades at the expense of their peers.

While student SA scores may not be as valid or reliable as PA or tutor scores, including SA scores when also utilizing PA is unlikely to have a strong negative impact on grading. Some researchers (Johnston & Miles, 2004; Malcolmson & Shaw, 2005) examined how students' assignment grades would be affected when SA scores were omitted from the calculation of the PA score, and they found that there would be minimal impact on student grades.

The very fact that students have difficulty with accurately assessing their own performance might be a reason to provide opportunities for students to practice SA. As a result, several researchers recommend including SAs with PAs (Elliott & Higgins, 2005; Johnston & Miles, 2004; Papinczak et al., 2007b).

Overall, research shows that SA scores tend to be less reliable than PA scores. SA may also tempt students to attempt to increase their own marks at the expense of their team mates. However, in spite of these issues, since SA and PA scores are usually averaged, the SA score has not been shown to have a significant impact on student grades. Therefore, for teachers who wish

their students to gain insight, experience, and skill with SA (e.g. in professional programs), those

teachers may find the potential benefits of SA to be worth the possible drawbacks.

What is the Most Appropriate Method for Incorporating PA in the Calculation of a

Student's Final Grade?

The literature reveals that the majority of teachers who require students to assess the

contributions of their peers used these PA scores to calculate some sort of grade, such as for a

group project. As reported above, most students believe that using PA scores is a fair and

appropriate method to calculate a student's individual contribution to a group assignment or

project. However, there are a wide variety of ways in which a PA score might count towards a

student's overall grade. Some of the basic methods include: (1) using a simple score, (2)

assigning a mark when a student completes PA, and (3) using the PA score to modify a group

mark.

The simple score. One of the greatest virtues of a simple score is its simplicity. Students

are assessed on a scale of *x* to *y* (where *x* and *y* can denote any parameters the teacher wishes).

For example, a student's contribution to a group project may be assessed holistically on a scale

between 0 and 100 or, alternatively, a student might be marked on six different criteria on a scale

between -1 and 3. In the latter case, a student would receive a score ranging between -6 and 18.

As shown in Table 1, when using a simple score, the student's PA scores are averaged to give a

mark.

Table 1

Example of how a student's PA score is calculated when using the simple scoring method

	If the PA is worth 10% of a student's final grade				
	PA scores awarded by the following students:				
	Peer 1	Peer 2	Peer 3	Peer 4	Peer 5
Peer 6	80/100	70/100	90/100	85/100	80/100
Average PA score:	81/100				
Calculation of student mark:	PA Mark (out of 10) = 81/100 x 0.1 = 8.1				

Researchers report mixed results when using simple scores. Some find that student participants tend to award their peers very similar high scores, inflating the grades of all students (Steensels et al., 2006), while others find that student participants use a wide range of scores when using this method (Sahin, 2008). Thompson and McGregor (2009) found that student participants tended to use the lowest ratings very rarely.

A mark for completing peer assessment. Some researchers who intend student participants to share formative feedback with each other choose not to assign marks based on the PA. In that case, students may simply get points or marks for completing the PA (Ferguson & Kreiter, 2007).

Using the peer assessment score to modify a group mark. The literature describes many different methods that may be used to modify a group mark with a PA score, to create an individual student mark that reflects the student's contribution.

Since different researchers use different terminology and formulas to describe their methods of calculating and using the PA score to modify a group mark, I will use the following standardized abbreviations when the various methods are described (below):

AGM   Average group PA mark (the average PA score assigned to all of the

               students in a group)

ASM   Average student PA mark (the average PA score assigned to a single student

               in a group)

ISM   Individualised student mark (taking PA into account)

N   The number of students in a group

TSM   Total student PA mark (the total of all PA scores assigned to a single student

               in a group)

UGM   Unadjusted group mark (the mark awarded by the teacher to the group – not

               including any PA adjustments)

Dividing a pool of marks. As the name of this method suggests, students are required to

divide a set number of points or marks among all of the members of their group (in some cases,

teachers ask students to include themselves, while in other cases, students do not include

themselves). In its simplest form, the pool of marks is made up of the group mark multiplied by

the number of students in the group:

$$Pool\ of\ marks = N\ x\ UGM$$

Then the students in the group divide this pool of marks among all of their members,

according to each student's contributions to the group. To calculate the student's final mark, the

PA score from each peer is averaged:

$$ISM = \frac{TSM}{N}$$

The weighting factor method. Many researchers describe using the PA scores to calculate a

weighting factor (often referred to as an 'individual weighting factor' or 'contribution factor')

which is calculated as:

$$weighting\ factor\ = \frac{ASM}{AGM}$$

This method has been attributed to Conway, Kember, Sivan, and Wu (as cited in Cheng &

Warren, 2000) as well as Goldfinch and Raeside (as cited in Li, 2001). According to this method,

a student's grade would be calculated by multiplying the student's weighting factor by the group

assignment mark:

$$ISM = UGM\ \times \frac{ASM}{AGM}$$

One of the benefits of this method is that the mean of the group marks is not affected by

student PA scores. So, while individual students' marks are adjusted upwards or downwards,

there is no net change in grades – avoiding the issue of grade inflation. Cheng and Warren (2000)

found that when they applied this method to their students' group marks, 50 out of 53 students'

grades were adjusted, leading to a wider distribution of marks. This wider distribution of marks

led 32% of their students to receive either a higher or lower letter grade than they otherwise

would have received. Another benefit of this method is that this method is fairly easy to calculate

or to program into a spreadsheet.

However, one drawback to this method is that it may produce an unacceptably wide range

of student scores (when students discriminate highly among their peers). For example, Kilic and

Cakan (2006) found that the "impact on most of the students' grades was too high to be

acceptable" (p. 646). In one case, a student's group mark would have been reduced by five letter

grades. So, in order to reduce the impact of the weighting factor, Kilic and Cakan used a Scaling

Factor. Although using this Scaling Factor allowed student's grades to be adjusted within a

reasonable range, this introduces an additional calculation into the PA process – which may

make this method too complicated for some educators to use. Rather than using a Scaling Factor

(which may require several calculations to identify the optimal number), Kaufman et al. (2000)

suggested that the weighting factor may be scaled down by simply taking its square root:

$$ISM = UGM \times \overline{\frac{ASM}{AGM}}$$

Some educators choose to multiply the group mark by the weighting factor only in specific

circumstances. For example, Kench et al. (2009) only used the weighting factor to reduce student

group marks, when their PA scores did not meet a certain minimum standard. They reported that

only 5% of their students had their group marks reduced according to this practice.

Adding/subtracting an individualisation factor from the group mark. Russell et al. (2006)

describe a method in which each student's average PA score is subtracted from the average mark

of all students in that group. If the student receives high PA scores for his/her contributions to the

group, then this individualisation factor will be positive, increasing that student's group

assignment mark. However, if a student receives low PA scores for below-average contributions

to the group, then this individualisation factor will be negative, decreasing that student's group

assignment mark. In this method, a student's individualised group mark is calculated as:

$$ISM = UGM + (ASM - AGM)$$

Like the Scaling Factor (described above), an additional factor may be used to adjust the

magnitude of the individualisation factor. Russell et al. explain that this additional factor (which

they call $C_{S\text{-}G}$) is selected by the teacher. When this factor is greater than 1.0, PA scores have a

larger impact on group marks, while a factor that is less than 1.0 will reduce the impact of PA

scores. However, Russell et al. do not explain how a teacher would select a value for $C_{S\text{-}G}$. Using

this additional factor, a student's individualised group mark would be calculated as:

$$ISM = UGM + C_{S-G} \times (ASM - AGM)$$

Base mark plus a contribution mark. In this method, only a portion of the group mark is adjusted by a PA score (e.g. Pocock et al., 2010). For example, a teacher may decide that 30% of a group's assignment mark will be adjusted by the PA score. In this case, the Base Mark will be the same for every student in the group, while the Contribution Mark is individualised for each student using the PA score. A group mark would be calculated as follows (assuming the PA mark = 30% of a group assignment):

$$ISM = \qquad Base\ Mark + Contribution\ Mark \quad where...$$

$$Base\ Mark = \qquad 70\% \times UGM \qquad\qquad and...$$

$$Contribution\ Mark = \quad 30\% \times UGM \times \frac{ASM}{AGM}$$

Using a computer program. For teachers using computer programs to calculate PA grades, it may be possible to use more complex formulas. For example, the SPARK[PLUS] program, described by Willey and Gardner (2009) allows users to select one of three different methods to calculate the value of PA scores: linear, original, and knee (Willey & Gardner, 2008). After students complete their assessments in the online program, SPARK[PLUS] calculates the student PA scores automatically. The computer program calculates a self and peer assessment score (SPA) which is multiplied by the unadjusted group mark to give an individualised student mark:

$$ISM = UGM \times SPA$$

Using a *linear* scale, the SPA is calculated as:

$$SPA = \frac{ASM}{AGM}$$

Using an *original* scale, the SPA is calculated as:

$$SPA = \frac{\overline{ASM}}{AGM}$$

Using a *knee* scale, the SPA uses this two-part formula (combining the linear and original scales):

$$If \frac{ASM}{AGM} \leq 1, then \ldots \quad SPA = \frac{ASM}{AGM}$$

$$If \frac{ASM}{AGM} \geq 1, then \ldots \quad SPA = \frac{\overline{ASM}}{AGM}$$

Willey and Gardner (2008) show the relationship between the SPA and the weighting factor for
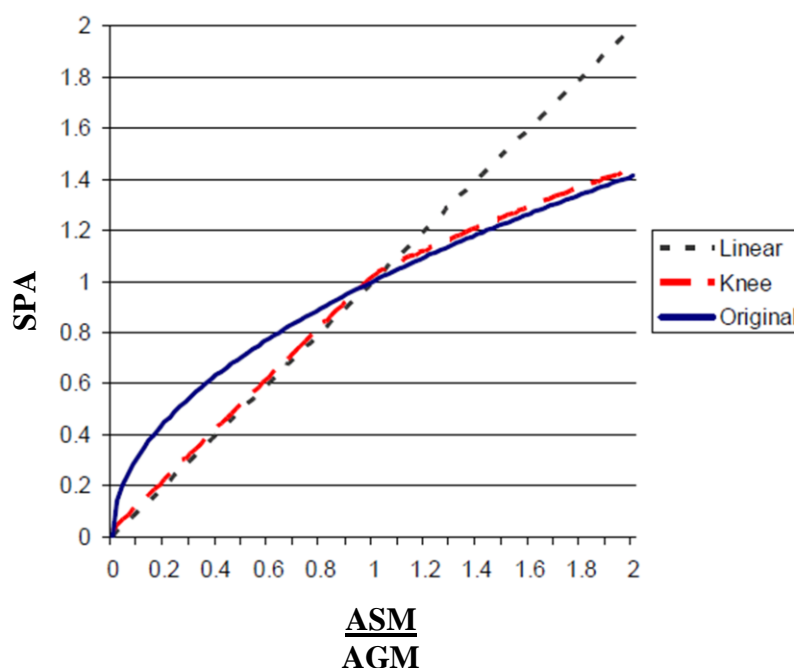
all three scales:

Figure 1. The relationship between the linear, knee, and original scales. Note that the knee plot has been slightly offset to increase readability. From *Improvements in the self and peer assessment tool SPARK: Do they improve learning outcomes?* by K. Willey and A. Gardner, 2008, Paper presented at the ATN Assessment Conference 2008: Engaging Students in Assessment. Reprinted with permission.

Willey and Gardner suggest that teachers may elect to use a particular method/formula for

various reasons. The linear scale can produce a wide distribution of marks (as found by Kilic &

Cakan, 2006), while the original scale brings the weighting factors closer to 1.0 and reduces the

impact of the PA scores. Nepal (2011) suggests that the optimal method to calculate student

grades would award low individual marks for below-average contributions and reward above-

average contributions. However, Nepal also cautions that, "to discourage individualism and to maintain teamwork spirit, a significantly higher individual mark for an above-average contribution than the mark awarded for an average contribution should be avoided" ("Methods used to assign individual marks", para. 3). The two-part method of the knee scale would seem to best achieve these goals.

DISCUSSION

After reviewing the literature on PA, it is time to return to the initial question: What is the best evidence relating to PA of individual student's contributions to group work? Under this broad heading, the four specific questions asked were:

1.   What are the potential benefits and drawbacks of PA?

2.   How valid and reliable is PA?

3.   What PA process and instrument constitute the best practice?

4.   What is the most appropriate method for incorporating PA in the calculation of a student's final grade?

Summary of Findings

Peer assessment is reported to provide a number of benefits to students, although some drawbacks or potential problems have also been reported. In most cases, both students and teachers found PA to be a fair way to assign individual students marks for group projects or assignments. Often, educators implement PA for the express purpose of reducing the social loafing that often accompanies group work, and the research shows that PA does help to reduce this widespread problem. PA may also teach students to be more critical of their own and their peers' work, helping them to develop higher-order thinking skills. In addition, requiring students to assess their peers can help students to build interpersonal skills, particularly when students

must share feedback with each other in face-to-face settings. The feedback that students receive

from their peers may help some students to alter their behaviour, leading them to become more

effective learners and team members. For students in professional programs, PA in the classroom

provides an opportunity to learn how to provide feedback to peers – a skill that they will need to

perform after graduation. In spite of these benefits, students often reported feeling uncomfortable

performing PA. Students may believe that it is not a student's role to provide feedback to peers

or that they do not have the ability to accurately assess their peers. Some evidence suggests that

students may become more comfortable with PA as they gain more experience with this skill. In

a few cases, students have been reported to not take PA seriously. A more common concern

reported by students relates to providing negative feedback. Students sometimes worried that

sharing negative feedback would harm their relationships with their peers. Students also worried

that negative feedback (or low PA scores) would have a negative impact on their peers' grades –

particularly when those peers had personal problems that prevented them from contributing fully

to the group work (as opposed to merely being a social loafer). In a small number of cases, the

PA process itself has led some students to develop dysfunctional group behaviours. When the PA

instrument required students to evaluate their peers' performance of a particular task, some

students 'took over' those tasks specifically to increase their PA marks. It is possible that the

type of criteria on the PA instrument might contribute to this particular problem. For example, if

students are required to complete particular aspects of a group project to earn PA marks, then

some aggressive students might be tempted to take over those tasks, simply to maximize their PA

marks. On the other hand, if students are assessed on their overall contributions and/or

interpersonal skills, dysfunctional group behaviours might be minimized.

The second question related to the validity and reliability of PA. Since most educators use their own unique PA instrument, it is impossible to make any sweeping statements about the validity and reliability of PA in general. However, a number of researchers have found some common trends. Students are often consistent in rating the contributions of their peers. This inter-rater reliability is higher in the ratings of high-performing students than in the ratings of low-performing students. On the other hand, PA scores show lack of consistency over time. Some researchers have found that students' PA scores increase, while others have found no significant changes over time. When a tutor is available to supervise group work, it may be possible to compare tutor scores and student PA scores for validity. A large number of studies have found a correlation between these two scores. However, in some studies, the students have consistently marked their peers higher than the tutor, while in other studies, the students have marked their peers lower than the tutor. Correlations have been also found between PA scores and other methods of evaluation (such as tests, projects, or group assignments). Another aspect of students' PA scores that is commonly examined is the variability of the class PA scores. In some cases, students award all of their peers the same (or a very similar) PA score, while in other cases, students award their peers a wide range of PA scores. It is likely that a number of factors impact students' willingness and ability to award their peers a wide range of marks, such as the PA criteria, students' ability to keep track of their peers' contributions, students' level of experience with PA, whether PA is submitted confidentially, and so on.

The third question guiding this literature review inquired about the PA process and instrument that constitute the best practice for educators. In the literature, a wide variety of processes and instruments are described. One variable in the PA process is whether the assessments are completed on paper or submitted online. Teacher workload is often reported to

be the most manageable when assessments are completed using an online computer program

(rather that collating paper PA forms and calculating scores by hand). Also, when the PA process

is meant to be confidential, students sometimes express concern that confidentiality may be

compromised in paper-based assessments – particularly when these assessments are either

completed or submitted in the presence of their peers. Students are more confident that online PA

methods preserve confidentiality. Another variable in PA relates to whether the PA instrument is

holistic, holistic with criteria, or categorical. Several studies have found that a holistic

assessment with criteria tends to produce the most reliable scores, as well as the highest level of

satisfaction among students. In this type of assessment, the PA instrument lists a number of

criteria, but students assign a single score to each peer. Peer assessments may also be completed

confidentially or openly. Although most researchers reported that it was important to keep PA

confidential, a few researchers described situations in which PA was completed openly. In some

cases, students collaborated in their groups to assign a single PA score to each peer, while in

other cases, students completed individual PAs, but submitted their scores openly. Some

researchers have found that students awarded a wider range of PA scores when the PA was

completed confidentially than when it was completed openly. However, an open PA process may

lead to positive outcomes such as improving students' interpersonal skills. Another variable in

the PA process is whether students are required to provide narrative comments on the PA

instrument. Usually these narrative comments are directed at their peers, although at times, the

narrative comments are intended for the teacher, to justify the scores that were awarded to the

peers. Peer comments are often found to be similar to staff comments but greater in both number

and variety. These peer comments may have the potential to provide useful feedback to help

students become more effective learners. High-performing students often provide longer and

more detailed comments than low-performing students. Another variable is that the PA process may involve either rating or ranking one's peers. In almost all cases, students are required to rate their peers (assign them a score). Using rating scales allows teachers to assign a grade in a fairly straightforward manner. Only two studies described using PA to rank peers. In the first study (Reiter et al, 2002), having students rank several characteristics in each peer was not found to be a reliable method of PA. In the second study (Heyman & Sailors, 2010), having students rank the high and low contributors within a class of 30 students was reported to be quite effective. A number of researchers highlighted the importance of adequately training and preparing students for PA. When students are well oriented and have opportunities to practice PA, they are more likely to be confident with the process and produce valid PA scores for their peers. Another variable in the PA process is whether to involve the students in setting the PA criteria or not. In most cases, researchers have reported positive outcomes when involving students in setting the PA criteria, including improved student satisfaction and improved reliability and validity of PA scores. One last variable is whether or not to include SAs in the PA process. Research shows that SA scores are usually less reliable than PA scores. In some cases, students consistently mark themselves higher than their peers or teacher, while in other cases, students consistently mark themselves lower than their peers or teacher. However, including SA scores in a PA is unlikely to make a significant difference in the PA score. Therefore, the decision as to whether or not to include SA will depend on what the teacher wants to achieve. If the goal is merely to produce a PA score that will be used to calculate a student's grade, then incorporating SA will not likely be useful. However, if the teacher intends the students to gain experience and insight with SA, then it could be worthwhile to include SA in the PA process.

The final question in this literature review asked about the most appropriate method for incorporating PA in the calculation of a student's final grade. Three general methods are reported in the literature. First, a teacher can designate a certain portion of the course grade for a PA mark, and then students are awarded a simple score (which may be based on any number of criteria). Second, a teacher can award marks for simply completing the PA. In this case, the numerical score or written comments only provide formative feedback for the peers: it is the act of completing the PA that leads to a grade. Third, the PA score can be used to modify a group mark to create an individualised student mark. It is this third method that is reported most frequently in the literature. There are a number of mathematical formulas that may be used to modify a group mark. Some of the methods are fairly simple while others are quite complex. Depending on the formula used, PA scores may alter the mean of the group scores, possibly inflating all of the student grades. Some of the mathematical formulas compensate for this problem, so that while individual students may have their grades increased or decreased from the original group mark, the mean score remains the same.

Recommendations for Practice

Based upon the findings in this literature review, it is possible to make some recommendations for educators considering using PA to evaluate the contributions of individual students who are working in groups. The initial step in this process begins well before the start of a course, by planning for PA. There are many variables to consider when designing a PA process that meets the needs of both teacher and students.

First, it is important to recognize and clearly communicate the purpose of PA in the classroom. Not only does PA have the potential to reduce social loafing, but it may also help students to become more effective learners, improve their interpersonal skills, and learn to

critically appraise the work of themselves and others. Second, it is important to ensure that both students and teachers are adequately oriented to the PA process. This orientation should include information about how to give feedback, how to complete and submit the PA instrument, when the PA is scheduled to be completed, how the students will receive their peers' feedback, and how (or if) the PA scores will contribute to the students' grades. Since the literature shows that students may feel uncomfortable performing PA, or they may believe it is not appropriate for students to assess their peers, these concerns should be addressed in the orientation. Also, because some students worry that negative PAs may harm relationships with peers or reduce students' grades, it might be advisable to discuss these concerns in the orientation as well.

When deciding how the PA should contribute to a student's final grade, it may be worthwhile to consider using the PA scores to modify a group mark since it is this process that has been studied most extensively. Several researchers report that students may not take PA seriously unless it has the potential to have an appreciable impact on their final grades. Therefore teachers should ensure that PA is appropriately weighted. Many researchers describe using PA scores to modify a group mark that is worth 20-30% of a student's final grade.

Research has not clearly shown one method of calculating grades from PA scores to be superior to another. However, a method that is fairly straightforward and avoids inflating student grades is likely to be most acceptable to teachers. One method that accomplishes these goals is when students divide a pool of marks among their group members. For example, if the teacher gives the students 100 marks to distribute, then all of the students will receive a total score that is centred on 100. This total PA score may be treated as a percentage and used to multiply with the group mark (assigned by the teacher). As a result, some students' group marks may be increased

while others may be decreased. This method is simple to explain, easy to calculate, and transparent to students.

Evidence shows that the PA instrument which is most reliable, and that students are most motivated to complete, is a holistic assessment with specific assessment criteria listed. Therefore, teachers should carefully design the criteria that students will be assessed for. If desired, students may be involved in setting the assessment criteria. However, although students must clearly understand the assessment criteria, a holistic PA instrument only asks the student to provide a single score for their peers (rather than rating each peer on every criterion separately). Depending on the length of time that students will be working in groups, this holistic assessment may be completed only once, or multiple times per term. One additional note about the assessment criteria on the PA instrument: since some teachers have found that a few of their students attempt to take over particular group tasks in order to get a higher score on the PA, teachers may want to ensure that the criteria on the PA instrument centre on fair contributions, good interpersonal skills, and/or teamwork (instead of focusing on specific tasks that must be completed).

Whenever possible, students should have the opportunity to provide narrative comments during PA. Narrative comments may be shared verbally when PA is completed openly, or they may be provided in writing when PA is completed confidentially. These comments can offer useful feedback to the recipient that could be used to improve interpersonal skills and overall learning.

Utilizing a computer program to support PA appears to be a very satisfactory process for both students and teachers, so this option should be seriously considered whenever it is available. Some of the benefits of using a computer program include the ability to keep PA confidential

when desired and managing the work for the teacher (collecting and sorting the PA forms, as well as recording and calculating PA scores).

Although most researchers recommend using a confidential process for PA, both open and confidential methods of PA have been shown to be effective. Computer programs can be very useful for confidential methods of PA. However, if an open method of PA is desired, a process that allows students to track each other's work on the group project as well as providing opportunities to share formative feedback along the way may help students to justify the scores assigned to peers (leading to an appropriate distribution of PA scores within the student group).

Including SA into the PA process has not generally been shown to make a significant difference to students' PA scores. However, some educational programs (such as nursing) aim to encourage students to develop into reflective practitioners. In such a case, the teacher may find it worthwhile to incorporate SA into the PA process.

Areas for Further Research

While this literature review has provided much guidance on how to implement a PA process in which the contributions of individual students within a group are assessed, a number of gaps in knowledge have also been identified. Some interesting areas for future research include:

- Examining PA in the specific context of team-based learning.

- Conducting longitudinal studies of students engaging in PA to discover how students' attitudes toward PA change over time (e.g. if students come to feel more comfortable and skilled with PA, and if they become less worried about how PA may harm their peers). Such longitudinal studies could also examine whether there are any long-term trends in PA scores.

- Exploring what it means when successive PA scores in a group converge or diverge over successive episodes of PA.

- Studying the amount and type of training that students should receive for PA. It would also be helpful to know if this training should take place in the classroom or if students could be assigned readings or activities to complete outside of the classroom.

- Investigating the optimal number of times to perform PA per term, considering both the students' and teachers' points of view.

- Comparing the pros and cons of different methods of using a PA score to produce a student's final grade. Such a study could utilize quantitative methods to examine which formula produces the most valid/reliable results or qualitative methods to discover which method is most satisfactory to students and teachers.

- Examining the pros and cons of involving students in the development of PA criteria (compared with having the teacher provide the PA criteria). Are PA scores more valid or reliable? Are students more satisfied with the PA process?

- Investigating the pros and cons of having students discuss (in a class or tutorial setting) what they learned from their PA feedback. Would such a group discussion lead students to improve their interpersonal skills?

- Since current research shows that students and teachers believe that PA may improve students' interpersonal skills, a future study could attempt to objectively measure whether interpersonal skills actually do improve.

- Although current research shows that SA scores are less accurate than PA scores, it would seem reasonable that SA scores should get closer to PA scores over time, as

students become more proficient at accurately assessing themselves and their peers.

A future study could examine whether this assumption is true.

CONCLUSION

After completing this literature review, I have reflected on what I have learned and have

gained several insights into why my previous experiences with PA may have been disappointing.

As a result, I have collaborated with my current teaching partner to revise our PA process and

instrument in our TBL classroom. We have made the commitment to use the online PA program

available on our campus, in order to make the collection, management, and calculation of PA

scores easier and more efficient. This online method should also help to maintain student

confidentiality throughout the PA process. We have completely revised our PA instrument to

create a tool that is holistic, and students were involved in setting the assessment criteria for their

peers. The PA instrument will require students to provide narrative comments to their peers, and

students were given written instructions on how to provide specific, constructive feedback.

Students will complete PA twice during the term, so that they have the opportunity to learn from

the feedback of their peers to improve their teamwork skills by the end of the term. To help the

nursing students develop insight into SA, there will be a SA component to this PA. Also, since

we are aware of the possibility that a few students may attempt to skew the PA scores by giving

themselves an excessively high score at the expense of their team mates, we will carefully check

the SA and PA scores before releasing the results to the student recipients. As our previous

method of using PA scores to calculate a student grade (a simple score) led to grade inflation, we

have adopted a method that is a slight modification of the 'pool of marks' method, based on the

options available in our online PA program. For each student in the group, the available pool of

marks is increased by 100 points – so, for a group of six students, there will be 600 points to be

divided among the group members. We anticipate that students who are high contributors will receive PA scores above 100, while low contributors will receive PA scores below 100. The computer program calculates each student's average PA score. This average PA score will be treated as a percent, and multiplied by the student's group mark (a sum of all his/her group assignments for the term). This total group mark is worth 28% of each student's final grade, providing an incentive for each student to contribute to his/her group. We hope that these changes will help to create a PA process that is fair and meaningful for ourselves and our students.

For teachers who incorporate student group work in their classroom activities, PA has the potential to enhance student learning. Not only might PA increase student engagement, but it may also help students to develop their interpersonal skills and critical thinking. However, teachers who are new to PA may find it challenging to implement. This literature review reveals a myriad of options available to teachers who are thinking of using PA. By reviewing these options, selecting strategies appropriate for their specific classroom setting, and implementing PA thoughtfully and intentionally, teachers should be able to create an experience that will enhance student group work in their classrooms.

REFERENCES

Brooks, C.M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. *Journal of Education for Business, 78*(5), 268-272. doi: 10.1080/08832320309598613

Brutus, S. & Donia, M.B.L. (2010). Improving the effectiveness of student in groups with a centralized peer evaluation system. *Academy of Management Learning and Education, 9*(4), 652-662. doi: 10.1080/02602930701698942

Carson, K. M., & Glaser, R. E. (2010). *Chemistry Is In the News*: Assessing intra-group peer review. *Assessment & Evaluation in Higher Education, 35*(4), 381-402. doi: 10.1080/02602930902862826

Chaves, J. F., Baker, C. M., Chaves, J. A., & Fisher, M. L. (2006). Self, peer, and tutor assessment of MSN competencies using the PBL-Evaluator. *Journal of Nursing Education, 45*(1), 25-31. Retrieved from http://www.journalofnursingeducation.com/

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education, 5*(2), 243-255. doi: 10.1080/135625100114885

Clark, M. C., Nguyen, H. T., Bray, C., & Levine, R. E. (2008). Team-based learning in an undergraduate nursing course. *Journal of Nursing Education, 47*(3), 111-117. Retrieved from http://www.slackjournals.com/jne

Dana, S. W. (2007). Implementing team-based learning in an introduction to law course. *Journal of Legal Studies Education, 24*(1), 59-108. doi:10.1111/j.1744-1722.2007.00034.x

Divaharan, S., & Atputhasamy, L. (2002). An attempt to enhance the quality of cooperative

learning through peer assessment. *Journal of Education Enquiry, 3*(2), 72-78. Retrieved

from http://www.ojs.unisa.edu.au/index.php/EDEQ/index

Drexler, J. A., Jr. Beehr, T.A., & Stetz, T.A. (2001). Peer appraisals: Differentiation of

individual performance on group tasks. *Human Resource Management, 40*(4), 333-345.

doi: 10.1002/hrm.1023

Elliott, N., & Higgins, A. (2005). Self and peer assessment – Does it make a difference to student

group work? *Nurse Education in Practice, 5*(1), 40-48. doi: 10.1016/j.nepr.2004.03.004

Epstein Educational Enterprises (n.d.). What is the IF-AT? Retrieved 15 September, 2010 from

http://www.epsteineducation.com/home/about/default.aspx

Espey, M. (2008) Does space matter? Classroom design and team-based learning. *Review of

Agricultural Economics, 30*(4), 764-775. doi:10.1111/j.1467-9353.2008.00445.x

Fairfield, K. D., & London, M. B. (2003). Tuning into the music of groups: A metaphor for

team-based learning in management education. *Journal of Management Education, 27*(6),

654-672. doi: 10.1177/1052562903257939

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-

analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-

322. doi: 10.3102/00346543070003287

Ferguson, K. J., & Kreiter, C. D. (2007). Assessing the relationship between peer and facilitator

evaluations in case-based learning. *Medical Education, 41*(9), 906-908. doi:

10.1111/j.1365-2923.2007.02824.x

Fink, L. D. & Parmelee, D. X. (2008). Preface. In L. K. Michaelsen, D. X. Parmelee, K. K.

McMahon, & R. E. Levine (Eds.), *Team-based learning for health professions education:*

A guide to using small groups for improving learning (pp. xi-xvi). Sterling, VA: Stylus.

Friedman, B.A., Cox, P. L., & Maher, L. E. (2008). An expectancy theory motivation approach

to peer assessment. *Journal of Management Education, 32*(5), 580-612. doi:

10.1177/1052562907310641

Gielen, S. Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity.

Asessment and Evaluation in Higher Education, 36(2), 137-155. doi:

10.1080/02602930903221444

Heyman, J. E. & Sailors, J. J. (2010). Peer assessment of class participation: Applying peer

nomination to overcome rating inflation. *Assessment and Evaluation in Higher Education,*

Advance online publication. doi: 10.1080/02602931003632365

Jin, X. (2011). A comparative study of effectiveness of peer assessment of individuals'

contributions to group projects in undergraduate construction management core units.

Assessment and Evaluation in Higher Education, Advance online publication. doi:

10.1080/02602938.2011.557147

Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment and*

Evaluation in Higher Education, 29(6), 751-768. doi: 10.1080/0260293042000227272

Kamp, R.J.A., Domnas, D.H.J.M., Van Berkel, H.J.M., & Schmidt, H.G. (2011). Can peers

adequately evaluate the activities of their peers in PBL? *Medical Teacher, 33*(2), 145-150.

doi: 10.3109/0142159X.2010.509766

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681-706. doi: 10.1037/0022-3514.37.6.822

Kaufman, D.B., Felder, R.M., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education, 89*(2), 133-140. Retrieved from http://www.jee.org/

Kench, P.L., Field, N., Agudera, M., & Gill, M. (2009). Peer assessment of individual contributions to a group project: Student perceptions. *Radiography, 15*(2), 158-165. doi: 10.1016/j.radi.2008.04.004

Kennedy, G. J. (2005). Peer-assessment in group projects: Is it worth it? *Proceedings of the 7th Australasian conference on Computing education, 42.* Retrieved from http://delivery.acm.org/10.1145/1090000/1082432/p59-kennedy.pdf?ip=206.87.45.23&CFID=27593237&CFTOKEN=81529548&__acm__=1308 612866_d17a2b6344a5345416b5310700b00cac

Kilic, G.B., & Cakan, M. (2006). The analysis of the impact of individual weighting factor on individual scores. *Assessment and Evaluation in Higher Education, 31*(6), 639-654. doi: 10.1080/02602930600760843

Kruck, S.E., & Reif, H.L. (2001). Assessing individual student performance in collaborative projects: A case study. *Information Technology, Learning, and Performance Journal, 19*(2), 37-47. Retrieved from http://www.osra.org/itlpj/kruckreif.pdf

Lejk, M. and Wyvill, M. (2001a). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education, 26*(1), 61-72. doi: 10.1080/0260293002002229 1

Lejk, M., & Wyvill, M. (2001b). The effect of the inclusion of self-assessment with peer

assessment of contributions to a group project: A quantitative study of secrete and agreed

assessments. *Assessment and Evaluation in Higher Education, 26*(6), 551-561. doi:

10.1080/02602930120093887

Lejk, M., & Wyvill, M. (2002). Peer assessment of contributions to a group project: Student

attitudes to holistic and category-based approaches. *Assessment and Evaluation in Higher

Education, 27*(6), 569-577. doi: 10.1080/0260293022000020327

Levine, R. E. (2008). Peer evaluation in team-based learning. In L. K. Michaelsen, D. X.

Parmelee, K. K. McMahon, & R. E. Levine (Eds.), *Team-based learning for health

professions education: A guide to using small groups for improving learning* (pp. 103-

111). Sterling, VA: Stylus.

Levine, R. E., Kelly, P. A., Karakoc, T., & Haidet, P. (2007). Peer evaluation in a clinical

clerkship: Students' attitudes, experiences, and correlations with traditional assessments.

Academic Psychiatry, 31(1), 19-24. doi: 10.1176/appi.ap.31.1.19

Li, L.K.Y. (2001). Some refinements on peer assessment of group projects. *Assessment and

Evaluation in Higher Education, 26*(1), 5-18. doi: 10.1080/026029300200225 5

Machado, J. L. M., Machado, V. M. P., Grec, W., Bollela, V. R., & Vieira, J. E. (2008). Self- and

peer assessment may not be an accurate measure of PBL tutorial process. *BMC Medical

Education, 8*(1), 1-6. doi: 10.1186/1472-6920-8-55

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work.

Studies in Higher Education, 26(1), 53-63. doi: 10.1080/03075070020030715

Malcolmson, C., & Shaw, J. (2005). The use of self- and peer-contribution assessments within a

    final year pharmaceutics assignment. *Pharmacy Education, 5*(3/4), 169-174. doi:

    10.1080/15602210500293142

Michaelsen, L. K. & Sweet, M. (2008). Fundamental principles and practices of team-based

    learning. In L. K. Michaelsen, D. X. Parmelee, K. K. McMahon, & R. E. Levine (Eds.),

    Team-based learning for health professions education: A guide to using small groups for

    improving learning (pp. 9-34). Sterling, VA: Stylus.

Michaelsen, L. K. Parmelee, D. X., McMahon, K.K., & Levine, R. E. (2008). *Team-based*

    learning for health professions education: A guide to using small groups for improving

    learning. Sterling, VA: Stylus.

Nepal, K. P. (2011). An approach to assign individual marks from a team mark: The case of

    Australian grading system at universities. *Assessment and Evaluation in Higher Education*,

    Advance online publication. doi: 10.1080/02602938.2011.555815

Ohland, M.W., Layton, R.A., Loughry, M.L., & Yuhasz, A.G. (2005). Effects of behavioural

    anchors on peer evaluation reliability. *Journal of Engineering Education, 94*(3), 319-326.

    Retrieved from http://www.jee.org/about-jee

Papinczak, T., Young, L., & Groves, M. (2007a). Peer assessment in problem-based learning: A

    qualitative study. *Advances in Health Sciences Education, 12*(2), 169-186. doi:

    10.1007/s10459-005-5046-6

Papinczak, T., Young, L., Groves, M., & Haynes, M. (2007b). An analysis of peer, self, and tutor

    assessment in problem-based learning tutorials. *Medical Teacher, 29*(5), e122-e132. doi:

    10.1080/01421590701294323

Parker, N. R. (2007). *A team-based learning model to improve sight-singing in the choral music classroom* (doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database.

Parmelee, D. X. (2008). Team-based learning in health professions education: Why is it a good fit? In L. K. Michaelsen, D. X. Parmelee, K. K. McMahon, & R. E. Levine (Eds.), *Team-based learning for health professions education: A guide to using small groups for improving learning* (pp. 3-8). Sterling, VA: Stylus.

Pileggi, R. & O'Neill, P. N. (2008). Team-based learning using an audience response system; An innovative method of teaching diagnosis to undergraduate dental students. *Journal of Dental Education, 72*(10), 1182-1188. Retrieved from http://www.jdentaled.org/

Pocock, T.M., Sanders,T., & Bundy,C. (2010). The impact of teamwork in peer assessment: A qualitative analysis of a group exercise at a UK medical school. *Bioscience Education, 15*(3). doi: 10.3108/beej.15.3

Raban, R., & Litchfield, A. (2007). Supporting peer assessment of individual contributions in groupwork. *Australasian Journal of Educational Technology, 23*(1), 34-47. Retrieved from http://www.ascilite.org.au/ajet/ajet.html

Reiter, H. I., Eva, K. W., Hatala, R. M., & Norman, G. R. (2002). Self and peer assessment in tutorials: Application of a relative ranking model. *Academic Medicine, 77*(11), 1134-1139. Retrieved from http://journals.lww.com/academicmedicine/pages/default.aspx

Russell, M., Haritos, G., & Combes, A. (2006). Individualising students' scores using blind and holistic peer assessment. *Engineering Education, 1*(1), 50-59. Retrieved from http://www.engsc.ac.uk/journal/index.php/ee/index

Sahin, S. (2008). An application of peer assessment in higher education. *The Turkish Online*

> *Journal of Educational Technology, 7*(2), 5-10. Retrieved from http://www.tojet.net/

Saito, H., & Fujita, T. (2009). Peer-assessing peers' contributions to EFL group presentations.

> *RELC Journal, 40*(2), 149-171. doi: 10.1177/0033688209105868

Shiu, A.T.Y., Chan, C.W.H., Lam, P., Lee, J., & Kwong, A.N.L. (2011). Baccalaureate nursing

> students' perceptions of peer assessment of individual contributions to a group project: A

> case study. *Nurse Education Today,* Advance online publication. doi:

> 10.1016/j.nedt.2011.03.008

Sivan, A. (2000). The implementation of peer assessment: An action research approach.

> *Assessment in Education, 7*(2), 193-213. doi: 10.1080/713613328

Sluijsmans, D.M.A., Moerkerke, G., van Merriënboer, J.J.G., & Dochy, F.J.R.C. (2001). Peer

> assessment in problem based learning. *Studies in Educational Evaluation, 27*(2), 153-173.

> doi: 10.1016/S0191-491X(01)00019-0

Steensels, C., Leemans, L., Buelens, H., Laga, E., Lecoutere, A., Laekeman, G., & Simoens, S.

> (2006). Peer assessment: A valuable tool to differentiate between student contributions to

> group work? *Pharmacy Education, 6*(2), 111-118. doi: 10.1080/15602210600662279

Thompson, D., & McGregor, I. (2009). Online self- and peer assessment for groupwork.

> *Education + Training, 51*(5/6), 434-447. doi: 10.1108/00400910910987237

Weaver D., & Esposto, A. (2011). Peer assessment as a method of improving student

> engagement. *Assessment and Evaluation in Higher Education,* Advance online publication,

> 1-12. doi: 10.1080/02602938.2011.576309

Willey, K., & Gardner, A. (2008). *Improvements in the self and peer assessment tool SPARK:*

> *Do they improve learning outcomes?* Paper presented at the ATN Assessment Conference

2008: Engaging Students in Assessment, Adelaide, South Australia. Retrieved from

http://www.ojs.unisa.edu.au/index.php/atna/article/view/343/258

Willey, K., & Gardner, A. (2009). Developing team skills with self- and peer assessment: Are

benefits inversely related to team function? *Campus-Wide Information Systems, 26*(5),

365-378. doi: 10.1108/10650740911004796

Willey, K., & Gardner, A. (2010). Investigating the capacity of self and peer assessment

activities to engage students and promote learning. *European Journal of Engineering

Education, 35*(4), 429-443. doi: 10.1080/03043797.2010.490577

APPENDIX A: PEER ASSESSMENT FORM (WINTER TERM 1 - 2009)

Instructions for completing your peer evaluations:

1. First, read through this evaluation form and think about what scores you should give your teammates. You may find it helpful to write the scores and your comments on a scrap piece of paper, first.
2. Remember that this evaluation is meant to be truthful and honest. This is your opportunity to reward your teammates who put in the work, and to provide some constructive feedback for other teammates who could stand a bit of improvement.
3. You are required to show some discrimination in your scoring. (That means that you may not assign 10/10 to all of your teammates)
4. Also remember that the evaluation form will require you to enter some comments (a comment about how each teammate contributes to the group; and a comment giving a constructive suggestion of how each teammate can improve). Please be respectful when making suggestions for improvement.
5. When you are ready…
6. Sign in to Vista and go to the Seminar class site.
7. Click on the link to go to iPeer
8. Use your student number as both your UserID and your password. Each student number is preceded by an "s". For example: s12345678
9. When you get in to iPeer, it may ask you to update your email address.
10. After that, click on "Home"
11. There should be a link to the "First Peer Evaluation for Seminar Class"
12. When you go to the evaluation, first start by scrolling up and down the page, to see how the evaluation form looks. It may be a bit confusing to scroll at first, because there are 2 scroll bars (one in your internet browser, and the other in iPeer itself)
13. The names of each of your teammates are listed on one of the dark blue heading bars.
14. First, look at the name of the student in the first menu bar. This is your first teammate that you will evaluate. Make sure you enter a score for each of the 10 numbered items, as well as the 2 comment boxes. You will need to use the inner scroll bar to get down to the bottom of the evaluation form.
15. When you have finished entering all of the comments for your 1st teammate, click on:

Save this Section (Click on this button to save now or you may lose your input)

16. Now look for the next dark blue menu bar, containing the name of your 2nd teammate. Repeat the process for the second, and all of your other teammates.
17. When you have saved all of your peer evaluations, click on:

Submit to complete the Evaluation

If you make a mistake:

Don't panic. Email Charlene at charlene.strumpel@ubc.ca … it is possible to delete your evaluations so that you can start over.

NRSG 214 Seminar Peer Evaluation

Section One (out of 10 marks)	Scale					
	Never (0)	Rarely (0.2)	Sometimes (0.4)	About ½ the time (0.6)	Most of the time (0.8)	All of the time (1)
1. Participates fully in team activities						
2. Comes to class well-prepared for team activities						
3. Communicates effectively with team members: • Expresses opinions respectfully and with clarity • Listens to the perspectives and contributions of others • Collaborates effectively with team members to make decisions and resolve conflicts						
4. Attendance: • Is present for team activities • On time/punctual						
5. Uses time efficiently and keeps the group focused on agreed-upon goals						
6. Takes responsibility for his/her own part of team work and decision-making						
7. Takes part in organizing team roles and responsibilities						
8. Is a good team player: • Encourages others to share their opinions • Is responsible for doing his/her own part of the work • Is not a "control freak"						
9. Shares accurate information with the team						
10. Open to change: • Willing to re-evaluate his/her own position in light of new information from others						
Total	a perfect score = 10 points					

Section Two: Comments (No marks for this section)
11. Please describe one thing that this team member does well, which helps to make your team more effective.
12. Please give one constructive suggestion, to help this team member become a more effective part of the team.

APPENDIX B: SUMMARY OF LITERATURE REVIEW

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Brooks & Ammons (2003)	Comparison of student scores, in 3 successive rounds of PA Student perceptions of PA (questionnaire)	Paper	Confidential	BOTH Holistic and Categorical	Yes	Group mark = 31.25% of final grade Divide a pool of marks	PA scores became closer together after the 2nd peer assessment. Students perceived that the free-rider problem was reduced by: • PA including specific comments • PA was conducted early • PA was conducted several times during the team project Students who believed the free-rider problem was reduced were more likely to: • believe that team projects were a good way to learn • believe that the group worked well together
Brutus & Donia (2010)	Compared student PA scores in multiple sections Compared scores of students who did PA once (2nd semester only) vs. students who PA twice (both 1st and 2nd semester)	Online	Confidential	Categorical	No	Survey of 35 professors who used the PES system: • 59% of professors used PA to modify a group mark • 10% used a simple score • 5% used PA to modify a participation mark	Students who performed PA twice received higher scores that students who had only performed PA once. Students who performed PA twice: The 2nd PA scores were higher than the 1st PA scores.
Carson & Glaser (2010)	Comparison of student PA scores (central tendencies and variability) Comparison of student perceptions (pre-course and post-course surveys)	Online	Confidential	Holistic with criteria	No	Group mark = 12% of final grade Divide a pool of marks	About ½ of the students' total PA scores = the average score (100). The students did discriminate when assigning PA scores to their peers. Students perceived the PA to be reasonably fair and accurate.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Chaves, Baker, Chaves, & Fisher (2006)	Comparison of SA scores, PA scores, tutor scores	Online	Not specified	Categorical	Yes	Not for marks	Numeric scores: PA score > SA score > tutor score. Comments: peers gave more positive comments than tutors or self assessments. Students rarely identified areas for improvement in peers (comments were optional).
Cheng & Warren (2000)	Examined the impact of the PA on student grades for a group project		Confidential	Categorical	No	Group mark = 40% of final grade Group mark multiplied by a weighting factor	94% of students' grades were altered by the PA (increased or decreased).
Divaharan & Atputhasamy (2002)	Student perceptions of PA	Verbal/ Paper	Open, consensus	Holistic with criteria		Group mark = 25% of final grade Not mentioned how students' grades were calculated	Students had generally positive perceptions of peer assessment: it was fair, it motivated them to work better in groups, it helped them to improve their interpersonal skills, it encouraged them to be more critical of themselves and their peers. Students felt awkward in participating in face-to-face PA.
Drexler, Beehr, & Stetz (2001)	Examined PA scores on 1st and 2nd group project Compared PA scores between the 2 projects Examined student perceptions of PA (survey)	Verbal	Open, consensus	Holistic with no criteria		Group mark = 40% of final grade Students awarded peers a score between 80-100% Group mark multiplied by PA score Students who did not do at least 80% of their share of the work got zero (in consultation with the teacher)	34% of groups differentiated on the PA score (increased or decreased). The amount of differentiation was very small (mostly between 98%-102%). Members of non-differentiating teams (all students had equal scores) had more positive attitudes towards their groups than students in the differentiating teams. The groups who did choose to differentiate their grades scored significantly higher grades on the 2nd project than the groups who assigned equal scores to everyone.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Elliott & Higgins (2005)	Examined student perceptions of PA (questionnaire)	Paper	Confidential	Categorical	Yes	PA and SA = 10% of the group project mark Simple Score	Most students found the SA and PA fair. Students reported an increase in motivation to participate meaningfully and produce a high standard of work. Students didn't have any trouble giving low marks to 'free riders', but did have trouble giving lower marks to peers with personal problems that reduced their contribution to group work.

Falchikov & Goldfinch (2000)	Meta-analysis: Examined the correlation between self and peer assessments and faculty-derived assessments				Yes		The greater correlations were found in:
							• PA requiring students to make a global judgement (holistic score) with several specific criteria or guidelines given
							• PA of group products and group process (rather than 'professionalism')
							• Situations where the students rated less than 20 peers
							• PA in which students had input into the criteria
							• There was no particular correlation found in:
							  ○ PA, depending on the subject being taught
							  ○ PA, depending on the year/level of the students
							Also, well-designed studies had better correlations than poorly-designed studies – perhaps because the PA process was more clearly defined in those well-designed studies. They recommend that when implementing PA, ensure that it is clearly explained, so that students are not confused about the process.

Ferguson & Kreiter (2007)	Examined correlation between peer and faculty marks Examined student comments in PA Examined student and faculty perceptions of PA	Year 1: Paper Year 2: Paper Year 3: Online		Categorical	No	Mark to complete the PA	The peer and faculty scores had a significant correlation. Students felt comfortable evaluating peers, and found the feedback they received to be useful. Peer feedback was often similar to faculty feedback, but often provided an additional perspective. A few students complained about being *required* to differentiate in PA scores.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Friedman, Cox, & Maher (2008)	Compared: Holistic vs. Categorical PA, and Single vs. Multiple PA per course (4 conditions) Examined students' motivation to perform PA and students' satisfaction with PA in these 4 conditions (2 questionnaires)	Paper	Confidential	Holistic	No	Written team project = 30% of final grade Team presentation = 10% of final grade Group mark multiplied by a weighting factor	Students had the highest motivation to complete holistic PA multiple times per term and the lowest motivation to complete categorical PA multiple times per term. In all conditions, students were less motivated to perform PA at the end of the term. Students who were concerned with group issues (e.g. free-riding) were more motivated to rate their peers.
Heyman & Sailors (2010)	Compared: Teacher assessment of student participation in class to: Peer assessment of student participation using peer rating and Peer nominations (ranking)	Paper	Implied confidential	Compared: Rating (holistic) vs. Ranking	No	Class participation = 25% of final grade Holistic method used simple score Not stated how the peer ranking contributed to student grades	Rating: • Students discriminated less than the teacher (student PA scores had a range of 8, while teacher scores had a range of 15 marks) • Students marked peers higher than the teacher Nominations/ranking: • Instructor's rating of class participation correlated well with the PA rankings • Student rankings discriminated among high and low contributors (student PA scores had a range of 30, while teacher scores had a range of 20 marks)
Jin (2011)	Examined student perceptions of PA (questionnaire) in 2 classes			Categorical	No	1st class (group mark = 30% of grade): 2nd class (group mark = 5% of grade): Group mark multiplied by a weighting factor In 2nd class, PA only used to reduce a student's group mark	The majority of students in both classes had positive responses to the PA (that it was appropriate, fair, and clearly explained). Students in 2nd class had more positive responses and more students believed PA was fair.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Johnston & Miles (2004)	Examined and compared student SA and PA scores		Confidential	Categorical	Yes	Group mark = 20% of course grade Group mark multiplied by a weighting factor	Numeric scores: SA score > PA score. High performers over-rated self; low performers under-rated self. There was no correlation between students' self ratings $\leftrightarrow$ how their peers rated them. No correlation between the SA score $\leftrightarrow$ group assignment score. Significant correlation between the PA score $\leftrightarrow$ group assignment score. The groups had relatively strong agreement as to who were the greater and lesser contributors in each group. 26% of students did not differentiate in PA scores.
Kamp, Domnas, Van Berkel, & Schmidt (2011)	Examined validity and reliability of PA using their online M-PARS tool	Paper		BOTH Holistic and Categorical	No	Not for marks	The initial number of criteria (34) were reduced to 14, to provide a reliable PA instrument. Student ratings using this tool were found to be reliable. There was a high inter-rater agreement (it only took 4 peers to give a reliable evaluation).
Kaufman, Felder, & Fuller (2000)	Examined student SA and PA scores Compared SA and PA scores to student test scores	Paper	Midterm: Open Final: Confidential	Holistic	1st class: No 2nd class: Yes	Team homework mark = 15% of final grade Group mark multiplied by a weighting factor	Peer ratings correlated significantly with test scores. Numeric scores: SA score = PA score (no significant difference). No significant differences for different ethnic groups/minorities. Only 2 out of 39 teams submitted identical scores for all students. It was more common for students to rate themselves low (rather than inflating their grades).

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Kench, Field, Agudera, & Gill (2009)	Examined student perceptions of PA (questionnaire) Examined how student grades were affected by PA	Online	Confidential	BOTH Holistic and Categorical	No	PA was only used to <u>reduce</u> student marks (for unsatisfactory group contributions) Group mark multiplied by a weighting factor	Only 4 of the 169 students got their group scores decreased at mid-term, and 4 students got their group scores decreased at the end of term (not the same 4 students). Overall, student PA scores increased from mid-term to the end of term. In the questionnaire, students generally gave positive responses. • Understanding the PA process = most positive • Fairness and appropriateness of PA = somewhat positive • Impact of PA on student motivation & satisfaction = somewhat positive
Kennedy (2005)	Compared student PA scores Examined student comments on PA forms		Implied Confidential	Holistic with no criteria	No	Dividing a pool of marks	Some students were reluctant to give their peers low marks. Some students awarded all their peers 100 points, regardless of their contributions. Some students dominated the groups and took over the group tasks. Weaker students sometimes had less opportunity to contribute to the group tasks. Some students doubted that the PA process was fair. Student PA scores were not consistent (some groups had very large standard deviations). The PA process led to conflict between group members. After average PA scores were calculated, 72% of student grades were tightly clustered around the mean score (100).
Kilic & Cakan (2006)	Examined PA scores Compared their PA scores to Cheng & Warren's (2000) student scores	Paper	Confidential	Categorical	Yes	Group mark = 100% of final grade (2 projects) Group mark multiplied by a weighting factor AND a scaling factor	Their students had a much wider range of scores after both the first project and the 2nd project than Cheng & Warren (2000) AND the IWF had a larger range than Cheng & Warren. There was a very wide range in marks after applying the IWF in this study. Some students' grades changed by up to 4 letter grades after applying the IWF.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Kruck & Reif (2001)	Examined student perceptions of PA (survey)	Online	Confidential	Not mentioned		PA = 20% of the group project mark Dividing a pool of marks	20% of students indicated that the PA motivated them to perform differently, while 36% thought the PA would motivate their peers to perform differently. 41% thought that current value of PA was just right. Students with high GPAs tended to choose each other to work in teams.
Lejk & Wyvill (2001a)	Compared scores in: Holistic PA vs. Categorical PA	Paper	Confidential	BOTH Holistic and Categorical	Yes	The holistic score was used to determine student grades Group mark multiplied by a weighting factor	There was greater agreement among PA scores in the holistic assessment than the categorical assessment. The holistic assessment produced more students (than the category-based assessment) whose grades were increased or decreased more than 5 or 10% by the assessment process.
Lejk & Wyvill (2001b)	Compared 'secret' PA vs. open, agreed PA Compared scores in SA vs. PA: • Agreement of scores in each category • Agreement of SA with PA • Examined students who scored themselves higher or lower than peers had scored them	Paper	Confidential AND Open	Categorical	Yes	The 'open' PA score was used to determine student grades Group mark multiplied by a weighting factor	There was more agreement between the PA and SA scores in the 'open' assessment than in the 'secret' assessment. Stronger students tended to under-rate themselves and weaker students tended to over-rate themselves The 'secret' assessment provided a greater range of grades than the 'open' agreed assessment.
Lejk & Wyvill (2002)	Examined student perceptions of SA and PA (total of 4 surveys) • Before and after completing SA and PA in each course • Holistic vs. categorical PA	Paper	See above 2 studies	See above 2 studies	Yes	See above 2 studies	The students who experienced the holistic PA: • Appeared to be less negative about PA than the those using categorical PA • Were generally happier with their group process than those using categorical PA • After experiencing PA, many students reported finding a fair method of PA

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Levine, Kelly, Karakoc, & Haidet (2007)	Examined: Student PA scores vs. Knowledge-based scores Examined student comments on PA forms	Paper	Confidential	Holistic	Yes	Students allowed to choose if the PA would count for marks or not. Out of 8 cohorts of students, only 2 decided to use the PA for marks. Not mentioned how students' grades were calculated	Modest correlation between PA and knowledge-based scores. Many students spontaneously wrote comments about disliking peer evaluation (which led the authors to make the decision to eliminate the requirement to discriminate in the PA scores... which has since led to higher levels of student satisfaction).
Li (2001)	Describes a method to "normalize" peer assessments using an IWF			Categorical	No	Group mark multiplied by a weighting factor AND a normalisation factor	When the IWF method of calculating student grades from PA produces too wide a range of grades, the normalisation factor may be used to reduce the deviation in grades.
Machado, Machado, Grec, Bollela, & Vieira (2008)	Compared: Student SA scores vs. Student PA scores vs. Tutor scores			Categorical	Yes	PA = 10% of final grade Not mentioned how students' grades were calculated	Numeric scores: SA score = PA score > tutor score SA and PA scores increased from 1st → 2nd assessment. Tutor scores did not increase.
Magin (2001)	Examined: SA and PA scores vs. Tutor scores			Categorical	Yes	PA and tutor grade = 15% of final grade Not mentioned how students' grades were calculated	Reciprocity and collusion accounted for only about 1% of the variance in scores (minimal effect). It is possible for peer assessments to be relatively free of bias.
Malcolmson & Shaw (2005)	Examined PA and SA scores to see if: • Students discriminate with PA • The PA mark would change if the SA mark was added/ removed • Li's (2001) normalisation factor would affect PA/SA marks	Paper	Confidential	Categorical	Yes	PA = 7.5% of final grade Group mark multiplied by a weighting factor (and possibly a normalisation factor)	There was only a small amount of discrimination in the PA/SA grade. Removing the SA mark did not make a significant change in the PA mark. Using the normalization factor did not make a significant impact in the PA mark. PA was not seen as useful or fair (students did not perceive that it made a difference). Students from certain cultural groups seemed to dislike assessing each other.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Ohland, Layton, Loughry, & Yuhasz (2005)	Compared student scores on: • Holistic PA vs. • Categorical PA vs. • Holistic PA with behavioural anchors (descriptive criteria)		Confidential	Compared: Holistic vs. Categorical vs. Holistic with criteria	Yes	Group work (2 projects) = 20% of final grade Group mark multiplied by a weighting factor	The holistic PA with criteria was the instrument with the highest reliability. A few students attempted to skew the ratings (by scoring themselves high and the rest of the group low). Teachers used their discretion to remove the skewed ratings from the PA calculations.
Papinczak, Young, & Groves (2007)	Examined student and PBL tutor perceptions of PA and SA (direct observation and focus groups)	Paper	Somewhat confidential (completed in room with peers)	Categorical	Yes	Not mentioned	Students expressed both positive and negative perceptions of PA/SA. Two positive themes were: (1) increased responsibility for others, and (2) improved learning. Four negative themes were: (1) lack of relevancy, (2) challenges (e.g. difficulty with scoring system; apathy with process), (3) Discomfort (e.g. discomfort with giving lower scores to peers, wanting it to be more anonymous), and (4) Effects on the PBL process (e.g. relationships could be disrupted by ill-feeling from negative evaluations).
Papinczak, Young, Groves, & Haynes (2007)	Compared PA, SA, and tutor scores Examined student scores on self-efficacy instrument Examined student feedback (mainly reported in paper above)	Paper	Somewhat confidential (completed in room with peers)	Categorical	Yes	Not mentioned	When compared to tutor marks: • Some groups/individuals were more accurate than others • PA more accurate than SA • Correlation between PA and tutor scores improved as students got more practice Numeric scores: PA score > tutor score > SA score. Some students deliberately gave peers 100% irrespective of performance. Removal of highly skewed scores improved the correlation between PA and tutor scores. Students were sceptical about accuracy of PA (resulting from friendships, tit-for-tat, or dishonesty).

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Pocock, Sanders, & Bundy (2010)	Compared PA scores Examined student perceptions of PA (focus group or discussion with faculty member)		Open	Categorical		Base mark + contribution mark	65% of the groups discriminated between their individual members' contributions. Most students were positive about the overall experience & reported learning how to work more effectively as a team. Some students had negative perceptions of PA: (1) feeling uncomfortable with having to assign contribution marks, (2) feeling pressured to give higher contribution marks to peers than they thought the peers deserved, and (3) the contribution criteria did not adequately account for different work patterns.
Raban & Litchfield (2007)	Distribution of PA scores in 3 different conditions: • No online support • Time recording of contributions • Weekly time recording, rating and comments with TeCTra program	Year 1: Verbal Year 2: Verbal and online Year 3: Verbal and online	Open, consensus	Holistic		Dividing a pool of marks	During the years that there was no online support, between 75-90% of student groups awarded all students similar grades (0-5% differentiation in grades). During the years that students recorded the amount of time spent on the project each week, 55% of student groups awarded all students similar grades (0-5% differentiation in grades). During the years that students used the online TeCTra tool to record time spent on the project and rated their peers contributions weekly, at the end of the project, only about 20% of student groups awarded all students similar grades (0-5% differentiation in grades).
Reiter, Eva, Hatala, & Norman (2002)	Examined Student's relative rankings (SA and PA) vs. Tutor rankings	Paper	Confidential	Ranking	Yes	Not for marks	Students needed to rank 7 characteristics in each of their peers: • Use of relative ranking was not reliable (poor correlations with tutor rankings). • Relative rankings of strongest/weakest traits were not consistent over the 3 assessments.

Study	What was studied?	Online/paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Russell, Haritos, & Combes (2006)	Compared student PA scores Examined student comments on PA (justifying why they awarded a particular score)		Confidential	Holistic	Yes	Group mark = 100% of final grade Adding/subtraction the PA score from the group mark	Students did differentiate amongst their peers. PA scores changed with each task the group completed. Of the 7 groups, 2 groups' PA scores got closer to the mean with each successive PA, while 5 groups' PA scores were more differentiated with each successive PA. Higher achieving students tended to write more lengthy and detailed comments than the lower-achieving students. Students believed they were performing at a higher level than they actually were. Student comments indicated that they viewed PA as authentic.
Sahin (2008)	Compared: Student PA scores vs. Teacher scores	Online		Categorical	Yes	Not mentioned	There was a high correlation between the scores awarded by the teacher and the peer assessments. Students differentiated in PA scores.
Saito & Fujita (2009)	Compared students' PA scores	Paper	Confidential	Categorical	No	Not mentioned	Many (but not all) of the groups distinguished between students who were high or low co-operators.
Shiu, Chan, Lam, Lee, & Kwong (2011)	Student perceptions of PA (questionnaire and focus groups)	Paper	Somewhat confidential (student team leader collects PA forms)	Holistic with criteria	No	PA = 10% of the group project mark Base mark + contribution mark	Researchers found 7 themes: (1) **Satisfaction**, (2) **Free-riders**: PA only slightly helped to avoid free-riders , (3) **Fairness**: responses were mixed, (4) **Improving the quality of teamwork**: responses were mixed – students said that PA wouldn't influence lazy students (PA wouldn't change human nature), (5) **Weighting percentage of PA** – several students wanted the PA to be worth more than 10%, but in the focus groups, most agreed that it would be more appropriate to keep the weighting at 10%, (6) **Submission of PA**: students were concerned about confidentiality with the paper submissions, and (7) **PA criteria**: mixed reviews – most students wanted to keep the holistic PA, but wanted to include more criteria to help students decide on a mark.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Sivan (2000)	3 cycles of action research: Student perceptions of PA (questionnaire and semi-structured interviews)	Paper	1st and 2nd research cycles used Open PA 3rd research cycle used Confidential PA	Categorical	No	Group mark = 70% of final grade Group mark multiplied by a weighting factor	Greater exposure to PA leads to greater acceptance and confidence in PA. Better preparation of the students for PA leads to more acceptance and confidence in PA. Less experienced students recognized the value of PA but were reserved about the fairness of the process. Students recognized that PA contributed to their learning. PA provided an incentive for students to contribute to the group. When students will need to use PA in their future work as graduates, students said that PA should be part of the learning process in school (to better prepare them for the workplace). Students wanted the PA to be confidential. Students were satisfied with being able to contribute to setting the PA criteria.
Sluijsmans, Moerkerke, van Merriënboer, & Dochy (2001)	Examined student perceptions of PA (questionnaire) and compared responses in the 2 studies. Students in study 1 had not used PBL before Students in study 2 had more experience with different kinds of instruction			Categorical	No	Study 1: PA was part of the grade – not mentioned how students' grades were calculated Study 2: PA was not part of the grade, but if a student didn't get a passing score from his/her peers, the student would need to complete an additional task	The students in Study 2 were more confident in their ability to assess their peers. The students in Study 1 felt very uncomfortable in assessing peers. They also commented on the drawbacks of only giving a numerical score (no comments/feedback). The students in Study 2 also felt uncomfortable giving a negative score, especially if there was no opportunity to give comments/feedback.

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Steensels, Leemans, Buelens, Laga, Lecoutere, Laekeman, & Simoens (2006)	Compared student PA scores (in 2 successive episodes of PA) Examined student perceptions of PA (questionnaire)	Online	Confidential	Categorical	No	PA = 15% of project mark Not mentioned how students' grades were calculated	There was a wider range of PA scores on the 2nd PA (more discrimination in the 2nd PA). There was a significant correlation between PA scores and the ratings from the external tutors Numeric score: PA score > tutor score (but did not use same tool to score students). Most PA scores had a fairly similar, small standard deviation (indicating that there was inter-rater reliability) – but one group had scores with a much higher standard deviation (external tutor reported that there were 2 group members who didn't participate, and awarded skewed scores to peers). Overall, students tended to assign a narrow range of scores at the high end of the possible range. Overall, students reported a positive attitude to PA.
Thompson & McGregor (2009)	Compared student PA scores Examined student and faculty perceptions of PA (in examples A & B, researchers reflected on their experiences with paper-based PA) (in examples C & D, researchers conducted focus groups and interviewed teachers who used online SPARK tool)	A: Paper B: Paper C: Online D: Online	Class A & B were somewhat confidential Class C & D were confidential	A: Holistic with criteria B: Holistic C: Categorical D: Categorical	Yes	A: Group mark = 60% of final grade B: Group mark = 30% of final grade C: Group mark = 20% of final grade D: Group mark = 30% of final grade Exact formula to calculate grade was not given, but the peer mark was used as a multiplication factor	In the paper-based PA systems, students didn't differentiate in PA scores. Reasons given included: (1) Not enough time for students to reflect, (2) Lack of anonymity prevented students from giving higher or lower scores to peers, (3) Teacher's reluctance to implement a system that required complex calculations due to the workload for the large classes, (4) Method wasn't transparent to students, and (5) Lack of explanation to students meant students didn't take it seriously. Teachers found the paper-based system onerous and discouraging. In the on-line system: Numeric score: high-performing students SA > PA low-performing students PA > SA Few students used the lowest rating (they saved it for those students who really didn't contribute).

Study	What was studied?	Online/ paper	Confidential or open	PA Instrument	Include SA	Calculation of student grades	Findings
Weaver & Esposto (2011)	Compared student PA scores Compared student grades with previous cohorts of students (prior to course re-design) Examined student and lecturer perceptions of PA (focus groups)		Confidential	Holistic	No	Group mark multiplied by a conversion factor (The lecturer developed a conversion table, to convert the average PA mark to an 'IMF score')	Students expressed minimal concern about "free-riders". Students were not overly concerned about how the PA scores would impact their grades. Student PA ratings were very consistent with the lecturer's perception of students' contributions within their groups. Students were very consistent when rating high-performers, but were less consistent when rating low performers. High-performing students awarded a wider range of PA scores than low-performing students.
Willey & Gardner (2009)	Compared students' PA scores Examined students' perceptions of PA (questionnaire)	Online	Implied confidential	Categorical	Yes	Not mentioned how students' grades were calculated (implied that the SA/PA marks are used as a multiplication factor) Calculations made by SPARK[PLUS] program	Students who reported being neutral about whether SA or PA had improved their teamwork experience were not necessarily in well-performing teams. More students in teams with at least 1 poor team member reported that SA and PA improved their ability to give and receive feedback. Students in well functioning teams did not have much to discuss in the feedback sessions (where students discussed their feedback in tutorial sessions).
Willey & Gardner (2010)	Examined students' perceptions of PA (3 surveys)	Online and verbal	Implied confidential completion but feedback shared openly afterwards	Categorical	Yes	Not mentioned how students' grades were calculated (implied that the SA/PA marks are used as a multiplication factor) Calculations made by SPARK[PLUS] program	Students reported that the SA and PA were successful in helping them to achieve the desired learning outcomes. The peer feedback students received increased their engagement and supported their learning. Some students felt reluctant to assess their peers. Some students felt that the PA was mainly aimed at making the group project fair (ensuring all students contribute) [but teachers intended that the SA and PA were also to help students learn].