Using peer assessment of fieldwork to enhance students' practical training

Ahmed El-Mowafy

Spatial Sciences, Curtin University, Perth, Australia.

Fieldwork training is a key component of several practical disciplines. The paper discusses students' peer assessment of fieldwork as a method to improve their practical training. Peer assessment theories are first discussed. A framework for peer assessment of fieldwork was proposed and steps taken for preparation of students for this task are discussed. A developed marking, feedback and moderation tool of assessment is presented. Application of peer assessment in the field was investigated over a period of two years in one under-graduate unit in the Geospatial discipline as an example. Reliability of peer assessment was estimated by measuring the difference between assessments carried out by groups of peer assessors and its validity was measured by comparing students' marks with those given by tutors. Results show that students have gained from the peer assessment process, mainly as a formative form of assessment, by better understanding and endeavouring to achieve the objectives of field tasks. Tutors use differences among assessments made by groups of students compared to tutors' assessments to identify field components that need better explanation of their content and assessment criteria, tutors' expectations, and how to achieve better marks specific to each practical task.

Keywords: peer review; assessment for learning; moderation; student engagement, surveying.

Introduction

Practical disciplines, including geospatial sciences, surveying, mining engineering, agriculture, environmental, health and sports science require highly professional fieldwork training. Fieldwork components as defined in psychomotor domain taxonomy model presented by Dave, 1975, includes *imitation* (Observing and patterning behaviour after someone else.), *manipulation* (being able to perform certain actions by memory or following instructions), *precision* (performing a skill within a high degree of precision), *articulation* (coordinating and adapting a series of actions to achieve internal consistency) and *naturalization* (mastering a high level performance until it become natural).

In surveying as an example, students typically exercise each practical skill in only one session (El-Mowafy et al., 2013). As a result, their practical experience is limited to their own work. One effective way of improving students' experience is by involving them in peer assessment of other groups work (Falchikov, 2005). This has several advantages, such as:

- The peer assessment process, including teacher feedback, is a valuable way of improving students' approaches to problem solving by learning from mistakes and innovations of others.
- It supports early learning of professional methods including provision of constructive criticism to peers. This is needed in today's highly professional and competitive work environment.

- In order to assess others' work, students must be fully aware of and follow key learning objectives and proper methodology for the fieldwork and appreciate the importance of coming to the final correct solution.

In this paper, the term peer assessment will consistently refer to students assessing the work of their peers.

Fieldwork has two assessable components; the first is fieldwork activities, and the second is office work. The two components are different in nature. Assessment of office work is carried out after completing the fieldwork, finalising computations, reporting conclusions, and developing products (e.g. maps). The majority of assessors use traditional paper assessment forms and some use online tools. For instance, a number of computerised, office-based peer assessment tool kits are available, such as TeCTra (Raban et al. 2009) and SPARK[PLUS] (Freeman & McKenzie 2002). Some researchers, e.g. Campbell & Wren (2012) presented a useful feedback tool using iPads to assess live performance in music and drama which have applications for mobile assessment away from the classroom. However, assessment of fieldwork activities of applied sciences *in the field* involves specific preparation, instantaneous fieldwork interaction, practice, and the use of various tools. For instance, it includes:

- equipment checks;
- checking safety policies and procedures prior to commencing fieldwork;
- professional field booking, recording of data and notes;
- field approach;
- problem solving;
- interpretation of results;
- practical checks at different phases of the work including verification of final results; and
- compliance with professional practice in all of the above.

At present, there is a shortage in studies that discuss peer assessment of fieldwork *in the field.* In this paper, such field peer assessment is discussed and a case study in the area of Geospatial sciences is presented. The focus will be given to three research questions related to the importance of coaching students to carry out peer assessment and its methodology to achieve successful performance, the need for well-structured interactive digital rubrics and its design to provide clear assessment criteria and clear marking scales, and studying reliability and validity of the peer assessment results.

Literature Review

In this section, a brief summary of the relevant literature is given where key points and findings are presented. Most literature focuses on students' peer assessment of exam papers or assignments and discussion of fieldwork assessment is limited. Nevertheless, past research determines essential concepts that are needed to build on in our research related to peer assessment of fieldwork in the field.

Peer assessment as a learning strategy

Learning is an inherently social process, where different strategies for effective learning can be implemented (Strobl, 2007). Peer assessment is one of these strategies that can be used to help develop professional skills in the course of student education as well as

help to promote skill transfer to the workplace (Luca & McLoughlin, 2002). These include team skills, problem solving skills, decision making skills, communication skills, information literacy skills, and time management skills. In peer assessment of individual contributions to group projects, Falchikov (2007) cited evidence that students view peer assessment of group work to have a role in promoting life-long learning skills, including reflection, autonomy, self-efficacy, diplomacy, problem solving and responsibility. In addition, peers work closely together and may therefore have a greater number of accurate behavioural observations of each other (Greguras at al., 2001).

Assessment has a focus on future learning that reportedly improves both short- and long-term outcomes by helping students to make increasingly sophisticated judgments about their learning (Thomas et al., 2011). Part of this preparation for life-long learning requires helping students to learn to continuously monitor the quality of their work during the act of production itself, so they can make improvements in real time (Montgomery, 2000; Sadler, 1989). This can further be enhanced when students use criteria and apply standards when they make judgment during the peer assessment process (Falchikov, 2005). Students must directly demonstrate through assessment and evaluation that they are reaching the desired outcomes. Brew (1999) mentioned that when teachers share with their students the process of assessment, giving up control, sharing power and leading students to take on the authority to assess themselves, the professional judgment of both is enhanced. Assessment becomes not something done to students; it becomes an activity done with students. Thus, the aim of the peer assessment task is to allow students to exercise their judgment in assessing their peers and to apply what they have learnt in a self-reflective confirmatory exercise. The ability to make judgements about whether a performance or a product that meets given criteria is vital for effective professional action in any field (Biggs and Tang, 2007).

For peer assessment to have a positive impact, the peer assessment process should be effective, i.e. leading to outcomes desired by the students and the instructor and provide meaningful ratings, i.e. the student has taken the rating process seriously and spent time considering each rating (Friedman et al., 2008). An individual feels motivated when he or she feels that effort will lead to an acceptable level of performance, performance will lead to some outcomes, and the outcomes are personally valued (Isaac et al., 2001). Harbaugh (2012) noted the effectiveness of structured peer assessments and guided discussion activities as a means to increase motivation levels, autonomous learning skills and a sense of connectedness with other students in the unit. However, to achieve an effective peer assessment process it should be preceded by a stage of preparation of students for this task. This process is discussed in the following section.

Preparation of students for peer assessment

In order to promote the development of reflective, critical and evaluative skills needed for peer assessment, the learning environment should be designed to enable participants to (Luca & McLoughlin, 2002):
• have a clear understanding of the objectives;
• identify valid assessment criteria; and
• accurately and objectively evaluate success or failure on a given task.

Price and O'Donovan (2006) stress the importance of giving students sufficient practice with the assessment criteria and discussion to develop a shared understanding of the explicit and tacit assessment criteria before commencing peer assessment. These

points were echoed in AlFallay (2004) and Campbell et al. (2001) who found that students needed additional training to provide reliable assessment on specific criteria. Brew (1999) advised that more positive responses to the use of assessment are likely when the tutor's expectations are clear and when students have received systematic practice.

Instructors can use three strategies to improve the quality of peer assessment; modelling, scaffolding, and fading (Falchikov, 2007). Before engaging students in peer assessment, tutors can provide examples to students of how they personally use assessment tools and strategies to improve reliability and accuracy. In terms of scaffolding, Falchikov (2007) encouraged tutors to initially start with structured grading schemes (for example, rubrics) before moving to less structured systems where students negotiate the assessment criteria. Therefore, clear definition of the marking scales and the marking instructions is a condition for performing peer assessment. Students can use rubrics to assess the work of their peers and help them to understand the assessment criteria. It may also help students to become realistic judges of their own performance by enabling them to monitor their own learning rather than relying on their tutors for feedback (Crisp, 2007). Students can become realistic judges when expressing an awareness of what can be practically achieved or expected and when applying the assessment criteria in a methodological manner. Students can acquire expectations and standards of assessment tasks through a preparation stage of peer assessment to be offered by the instructors, for instance, as will be discussed in our experiment in a following section. As students achieve greater independence in peer assessment, the amount of direction and level of support offered by the tutor fades over time. However, this should be discussed and negotiated with students.

Some points of concern in peer assessment and proposed solutions

Davies (2003) reported negative consequences of peer assessment if students feel ill-equipped or not capable, if students feel uncomfortable with the responsibility of peer assessment duties, if students perceived that peer assessment were being used as a means of alleviating pressures for tutors, and if tutors have concerns about subjectivity and reliability of assessment. Carless (2007) seconded that opinion urging that teachers can improve the effectiveness of peer assessment by being very clear with the students about how they will benefit from participating. Biggs & Tang (2007) agreed that it is important that educational outcomes are made clear to the students because it is necessary to get the students on side. Bouzidi & Jaillet (2009) mentioned that performing peer assessment for exams helps the teachers identify what is wrong in their exams, thus giving them an opportunity to correct them. However, revision of the peer assessment increases the teacher's workload. Consequently, they suggest supplementing the solution by creating a system that enhances online assessment.

Bostock (2000) acknowledged certain weaknesses of peer assessment, in particular, the over-estimating of friends' work. However, he explains that this can be avoided by setting up a system which would guarantee anonymity, multiple assessments, multiple assessors (in general more than one), and moderation by the teacher. Bostock (2000) mentioned that the formative or summative assessment of other students' work by the students themselves has several advantages for the learning process, both for the assessor and the assessed. He pointed out that peer assessment encourages students to be independent and develop skills in high cognitive areas. However, some academics are cautious about using peer assessment as a form of summative assessment. Therefore

the use of peer assessment as formative or summative assessment varies from case to case according to the fieldwork activity and the learning objective.

Validity and reliability of peer assessment

In practice, peer assessment may pose validity and reliability concerns. This may include the activities to be assessed, the assessment tools and the assessors (Brennan, 2001). For example, the validity of a test is regarded as acceptable if the test comprises only items enabling the assessment of competences corresponding to training objectives. The reliability of an assessment tool is acceptable if it enables the accurate observation of the learning competence and makes it possible to give an opinion on this competence.

Cho et al. (2006) regarded the reliability of peer assessment a variable that can be measured by the difference between the marks given by peers. They consider the validity of peer assessment a variable that can be measured by the difference between the marks attributed by peers and by teachers. Thus, peer assessment reliability and validity depend on several factors such as students' academic level and abilities, the reliability of the marking schemes and the clarity of the marking instructions.

Some studies investigated the reliability of peer assessment (Haaga, 1993; Falchikov & Goldfinch, 2000). Various measurements were used, including comparison of means, Pearson's product moment correlation, concordance percentage, and intra-class correlation. Falchikov & Goldfinch (2000) found that peer assessments resemble more closely teacher assessments when global judgments, based on well understood criteria, are used rather than when marking involves assessing several individual dimensions. Bouzidi & Jaillet (2009) argued that validity comparisons in the literature are contradictory. Sadler & Good (2006) noticed that students slightly underestimate work of their colleagues, whereas they overestimate their own. De Grez et al. (2012) indicated that a comparison of the teacher and peer assessment scores points at a positive relationship, but also at critical differences.

Cho et al (2006), while carrying out an experiment involving 708 students, proved that the overall assessments of at least four peers on a written examination are reliable and just as valid as the assessments made by the teacher. The results suggested that the teachers' concerns regarding the reliability and validity of peer assessments should not prevent a peer assessment from being applied, at least with the appropriate scaffolding.

These conclusions on reliability and validity of peer assessment need to be further examined when performing this type of assessment in the field for fieldwork activities.

Experimenting peer assessment of fieldwork in the field

In this study, we investigate peer assessment of fieldwork in the field taking the discipline of Geospatial Sciences as an example. The experiment was applied for the unit 'GPS Surveying' which is offered to 3rd year under-graduate students at Curtin University, Australia. The experiment was performed for two years (2011 and 2012) and for four sessions in each year. The number of students involved was 30 and 32, respectively. The practical training is arranged such that two similar classes run per week. The class size was 16 students on average, and students carry out one session per week. Fieldwork is performed in groups of 4 students per group. In this experiment, peer assessment marks are taken as a formative assessment method.

Our research in this experiment focuses on two areas:

(1) Preparation of students for peer assessment through
- coaching of students to carry out peer assessment.
- development of well-structured interactive digital rubrics, which provide i) clear marking scale ii) clear instructions iii) constructive feedback to students and, iv) moderation of marking;

(2) Reliability and validity analysis, where marks assigned by different groups of students are compared, and tutors check consistency of grading and competence of reflection (i.e. reliability checking). Student marks are also compared with marks given by the tutors (i.e. availability checking). The differences were statistically analysed.

In the next sections, the methodology used in this study for peer assessment is first presented followed by a discussion on preparation of students for this process including presentation of a developed interactive digital marking tool. Next, reliability and validity analysis of the peer assessment as applied and tested in our study is discussed. Finally, students' involvement in a constructive assessment of peers is addressed.

Methodology

The framework for peer assessment is developed to ensure that the assessment method is constructively aligned to active learning methods. Steps considered are given in a flow chart, which is shown in (Figure 1). A modelling and scaffolding strategy was applied where before engaging students in peer assessment, examples and strategies of tutors' assessment were provided to students in order to improve reliability and accuracy of their assessment. Structured grading schemes with definition of the marking scales and the marking instructions were used in a form of rubrics. Students were involved in setting the assessment criteria during the design stage of the rubrics through an open discussion where generic field tasks were discussed, including their sequence, requirements, field steps, computations, presentation of results and ranking of different possible performance levels. Details of these rubrics are discussed in the next section. The rubrics help students to apply the assessment criteria in an organised way and aims at helping students to become realistic judges. In the field, paper form of the rubrics were used, and next in the office the marks were input into an interactive digital version of the rubrics to automatically assign marks and present outcomes and feedback. As discussed in Yorke (2003), it is not only the quality of the feedback evolving from the assessment that is crucial but what a student does with the feedback. Therefore, in our study, tutors follow up with students their response to the feedback they receive.

To evaluate reliability, each fieldwork assignment was evaluated by two groups and their marks were compared. The tutors on the other hand encourage and made sure that students take the rating process seriously and spend time considering each rating. In consequence, students were informed that quality of their assessment would be marked.
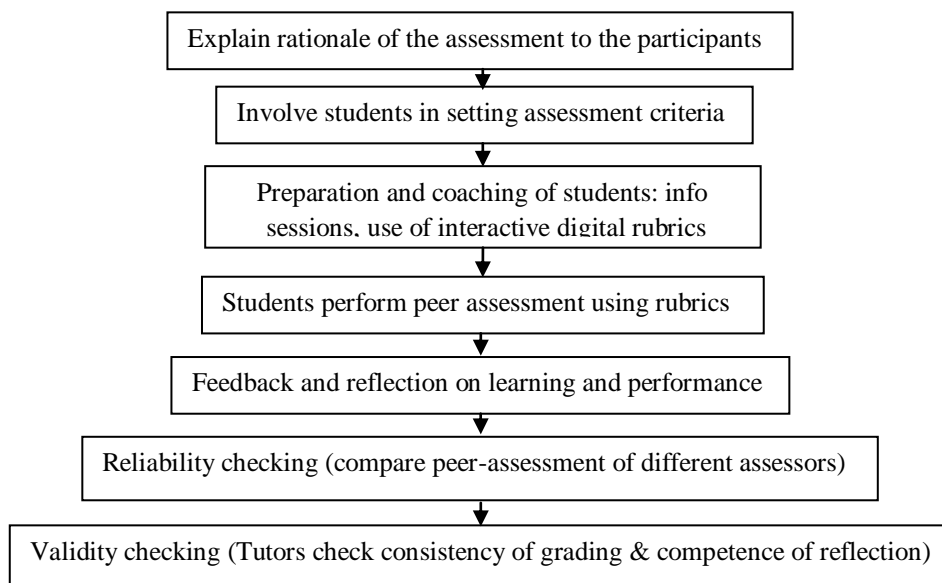
```
┌─────────────────────────────────────────────────────────┐
│     Explain rationale of the assessment to the participants │
└─────────────────────────────────────────────────────────┘
                            ↓
  ┌───────────────────────────────────────────────────┐
  │     Involve students in setting assessment criteria │
  └───────────────────────────────────────────────────┘
                            ↓
  ┌───────────────────────────────────────────────────┐
  │     Preparation and coaching of students: info      │
  │     sessions, use of interactive digital rubrics    │
  └───────────────────────────────────────────────────┘
                            ↓
  ┌───────────────────────────────────────────────────┐
  │     Students perform peer assessment using rubrics  │
  └───────────────────────────────────────────────────┘
                            ↓
 ┌─────────────────────────────────────────────────────┐
 │     Feedback and reflection on learning and performance │
 └─────────────────────────────────────────────────────┘
                            ↓
┌───────────────────────────────────────────────────────────┐
│   Reliability checking (compare peer-assessment of different assessors) │
└───────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────────┐
│ Validity checking (Tutors check consistency of grading & competence of reflection) │
└─────────────────────────────────────────────────────────────────┘
```

Figure 1. Framework of peer assessment of fieldwork in the field

Coaching of students to carry out peer assessment

In one session at the start of the semester, tutors explained the purpose of peer assessment and its value, and discussed with students the general criteria that students should use and standards they need to apply to judge the work of their peers. At the start of each fieldwork session, the tutors explained the objectives of the session, use of equipment (i.e. GPS receivers), field procedure, calculations, and expected outcomes. The tutors made sure that students had a clear understanding of the objectives, and discussed with them the session specific assessment criteria. Next, students carried out the fieldwork session, and during the last quarter of the session, two students from each group visited another group to perform peer assessment. They evaluate, through the use of the allocated rubrics, the ability of their peers in achieving the desired outcomes of the fieldwork session. The students are encouraged to engage in a group discussion to increase objectivity of the assessment.

To evaluate the importance of training students before practicing peer assessment, the difference between marks given by trained peer assessors with marks given by a group without training was compared. In general all students were trained except for one group in one class who performed peer assessment for all sessions who has only received basic instructions but without training. In addition, a comparison between peer and self-assessment was performed for all sessions and by all assessors, however this was only performed for the assessment of the report component.

Marking, feedback and moderation tool for peer assessment

The clear definition of the marking scales and the marking instructions is a condition for a successful performance of peer assessment. In addition, due to the fact that different persons are involved in the assessment process, moderation of marking should be

considered, which may vary according to the assessor's background and academic level. Therefore, the use of well-structured and detailed grading schemes (for example, rubrics) serves on these objectives and help in minimization of possible gaps in student grading of their peers.

Three versions of the marking tool were developed to support the three elements; provide clear marking scales, provide feedback of the assessors to their peers, and moderation of marking. To guarantee the best objectivity, the marking scheme was supplemented by clear marking instructions. The rubrics were designed to be adaptable to multiple surveying units. They can also be applied to other disciplines that have a fieldwork component, such as agriculture, physical education and mining with a limited effort for their adaptation into these disciplines. The templates have been designed for individual and group practicals and camp assessments with multiple tasks. The templates were provided to students before commencement of the field sessions. This ensures that students know in advance about how each fieldwork activity will be assessed, and about mark distribution for each task and for each performance level. This helps stimulate student's efforts in addressing all fieldwork tasks and objectives. The developed system was designed to target providing an accurate, fair and consistent moderation approach that narrows down variability in moderation of fieldwork between different assessors.

Four main assessment components were identified in the marking rubrics of surveying; fieldwork, field recording, computation & analysis, and presentation of results. The first two components are related to activities performed in the field whereas the last two are to be carried out in the office after data collection and verification. Each assessment criterion is quantified and varies according to nature of the task of each lab. These four areas were further broken down into four subcategories that are assessed. The component *Fieldwork* has the subcategories *correct instrument setup, maintained instrument setup, correct observation procedure,* and *closing/checking observations taken before leaving site.* *Field recording* has the subcategories *use of paper/digital field notes, observations recorded correctly, clear & complete field notes presented,* and *inclusion of index/cover page*. *Computation & analysis* comprises the tasks *basic calculations & reductions performed, closure and checking calculations carried out,* and *acceptable results achieved.* Finally, *Presentation of results* includes the subcategories *clear & well-structured report elements, description of fieldwork & methodology, results/analysis/conclusions,* and *required plans/maps/tables.*

The activities for each component have been described and linked to different performance levels (i.e. very poor, poor, pass, and distinction). The tasks and performance levels are set to meet the industry standards in fieldwork execution. These levels have been selected according to results of a questionnaire to industry stakeholders and discussion with students. A unified marking scale is linked to each performance level. The assessors need only to tick in the boxes describing performance level and feedback to peers. (Figure 2) shows, as an example, one rubric and its marking scale used for each component and work activity. The peer assessment was conducted in two separate settings by the same assessors, one setting in the field for assessment of fieldwork activities and one assessment after receiving fieldwork reports. Students used a paper hardcopy in the field and later transfer the marked boxes onto a digital copy of the template, which was also used for marking of reports and provide feedback. Preparation of the templates in a digital format serves in streamlining the process and would help in their implementation in a later stage with mobile devices (iPads). This

would help in minimisation of the assessment time spent on transfer of records to computers.

The digital copy of the marking rubric has two interconnected sheets. The first sheet, which is depicted in (Figure 2), is for grading student performance in each sub-task as discussed above. After grading in the first sheet, the assessment outcome is presented in the second marking sheet where a calculator tool is automatically applied (which is prepared via programmed Microsoft Excel worksheet) and assigns marks to each student according to performance in each activity and percentage of his/her contribution. The mark for each of the main components (fieldwork, field recording, computation & analysis and presentation of results) is calculated as the sum of the marks of its subcategories. This sum is presented in a scale from 1 to 10, and it is next weighted according to an assigned weight to each component from the total. This weight may vary from one practical session to another according to the fieldwork task at hand. The final mark is presented in a scale from 1 to 10 as the sum of the weighted marks of all components. (Figure 3) illustrates the calculation sheet component of the developed rubric. Specific feedback for each field task can also be given through this sheet. Furthermore, a field is added to reward innovation made or complexity of the task by scaling up or down the final mark.

Practice of testing the group assignment marking tool in the unit *GPS Surveying* has showed that the tool proved to be very useful as it helped students to focus on the objectives of each activity and assign marks according to achievement. A questionnaire was conducted for students who used the marking tool concerning its usage and required enhancement. In general, students found the marking rubric helpful in assisting their understanding of practical task requirements and in improving their performance and marking capability (Gulland et al., 2012, El-Mowafy et al., 2013). The response of the industry on the design and contents of the marking and feedback rubric received through another questionnaire was good and valuable comments and recommendations were received, which helped in the refinement of the tool.

Markers: Select ONE box per row

Feedback and Scale for Marks

Component		0-2: Very Poor	3-4: Poor	5-7: Pass	8-10: Distinction
Fieldwork	Correct instrument setup procedures used	☐ Negligence in setting up the instrument	☐ Poor setup, or instrument not centred or levelled correctly	☐ Setup procedure needs improvement	☑ Good setup procedure
	Maintained instrument setup (centring & levelling)	☐ Setup was not maintained	☐ Instrument not centred or levelled all the time	☑ Maintained but requires some improvement	☐ Instrument setup maintained all the time
	Correct observation procedure used	☐ Incorrect, rough with equipment, or poor safety practice	☐ Poor observation practice	☐ Observation practice needs improvement	☑ Good observation practice
	Closing/checking observations taken before leaving site	☐ No closing/checking was made	☐ Missing main checks or additional observations	☑ Some check obs were missed	☐ All check observations were collected
Field Recording	Use of paper/ digital field notes	☐ No evidence of field results collected	☐ Data and observations incomplete	☐ All data & observations evident	☑ All data & obs recorded in systematic manner
	Observations recorded correctly as observed	☐ No evidence of correct observation methods	☐ Large number of obs not in acceptable format	☑ Has most obs, and format acceptable	☐ All obs recorded and in acceptable format
	Clear & complete field notes presented	☐ Field notes ambiguous & largely incomplete	☐ Field notes incomplete & not organised	☑ Field notes complete but need organisation	☐ Notes are complete & well organised
	Inclusion of index/cover page and north point	☐ No index/cover page & no north point	☐ Index/cover page & north point incomplete	☐ Needs improvement	☑ All are well documented
Computation & Analysis	Basic calculations & reductions performed	☐ No calculations	☐ Significant errors in the calculations	☑ Calculations mostly correct	☐ Error-free calculations
	Closure & checking calculations carried out	☐ No calculations for checking or closing	☐ Significant errors in calculations of checks	☑ Minor misclose or check errors	☐ Close & checks are correct
	Analysis of results (and their errors) performed	☐ No analysis	☐ Incomplete or poor analysis	☑ Satisfactory analysis	☐ Complete & effective analysis
	Acceptable results achieved	☐ No results submitted	☐ Significantly wrong results	☑ Results are mostly acceptable	☐ All are complete and correct
Presentation of Results	Clear & well structured report, index, introduction & references	☐ Report not lodged or submitted after penalty period	☐ Poor structure, missing index/ refs, or spelling/ grammar mistakes	☐ Satisfactory structure, index, intro & refs	☑ Excellent structure, index, intro & refs
	Description of fieldwork & methodology	☐ Very poor reporting, or missing description	☐ Missing part(s) of description	☑ Satisfactory description	☐ Excellent description
	Results, analysis & conclusion	☐ Results or analysis are not included	☐ Poor results, no summary or no concl	☑ Satisfactory presentation	☐ Excellent presentation
	Required plans/ maps/ tables	☐ Required plans/ maps/ tables not included	☑ Poorly prepared, or not all are submitted	☐ Satisfactory but need improvement	☐ Excellent plans/ maps/ tables

Figure 2. Example of one marking sheet used (rubric)

Spatial Sciences - Survey Assessment & Feedback

Unit	GPS Surveying 382		
Assessment / Value (%):	Prac 4-Staking out by RTK GPS	5 %	
Student	ID	Name	Amount of work (%) *Must total 100%*
	1		
	2		
	3		
	4		
	5		
Submission Deadline	5:00 pm, October 5, 2012		
Unit Outcomes Assessed	UO1 Use their skills to process GPS data and solve critical GPS problems.		
	UO2 Perform a real time (RTK) and post mission (PPK) positioning.		
	U04 Prepare a GPS Survey report.		

Marking Scheme (For group)

Component	Mark (1-10)*	Feedback
Fieldwork (50%)	0.0	
Field Recording (10%)	0.0	
Computation & Analysis (20%)	0.0	
Presentation of Results (20%)	0.0	
Innovation Scale # (0.8-1.2)	1.0	
Group Mark (%)	0.0	

Individual Marks	Member	Final Mark (5)

Figure 3. Second sheet of the rubric, which includes the digital calculator tool and feedback

Evaluation of reliability and availability of peer assessment

Reliability and validity can be estimated by measuring the difference between assessments carried out by groups of students as peer assessors and by comparing their assessments with assessments done by the tutors. Reliability and validity depend on several factors such as students' academic level and abilities, the reliability of the marking schemes and the clarity of the marking instructions. In our experiment, it is assumed that:

- the assessment items are set up correctly;
- the observation and marking tools are reliable;
- the marking instructions and scale schemes are clear and coherent.

(Table 1) shows the mean (μ) and dispersion (measured by the standard deviation σ) of differences in marks between different groups of assessors when marking same tasks. μ and σ were computed for the four subcategories for each of the four identified main assessment components of the surveying fieldwork tasks, namely *fieldwork*, *field recording*, *computation & analysis*, and *presentation of re*sults. The given values are the absolute average values of μ and σ for all field sessions of the first year of testing (2011). For example, the mean of a main component (e.g. fieldwork) was computed such that, for each of its subcategories, the differences between the marks given by two groups of assessors (each peer-assessor group consists of two assessors who gave one mark for each task) was computed and μ and σ are the overall average and standard deviation of these differences of four groups per class, for two classes per practical

session, and for four sessions per semester. This gives a total sample for each task in a subcategory/semester = 1 difference × 4 groups × 2 classes × 4 practical sessions = 32. The given values of μ and σ are the average values of the tasks in the subcategories of each component. μ and σ were given as percentages of the total mark in each subcategory.

The differences are first computed among average marks given by different groups of student assessors (representing reliability) and when students self-assess their work against their peers' assessment. The comparison between peer and self-assessment was only performed for the two components *computation & analysis*, and *presentation of results* after students were provided with model reports by the tutors. The last two columns of (Table 1) show statistics of average differences between students' and tutors' assessments (representing validity). μ and σ for these differences were computed as explained above, where the total sample in this case for each task in a subcategory/semester = 2 difference (since we have two group assessors)× 4 groups × 2 classes × 4 practical sessions = 64.

   To estimate the inter-correlations among marks of different assessors, the Intra-class correlation (ICC) coefficient was used as it can be utilised to assess the consistency, or conformity, of measurements made by multiple observers measuring the same quantity and when quantitative measurements are made on units that are organized into groups (Shrout and Fleiss, 1979). The overall averages ICC of marks given by different groups of assessors (peers, tutors, and self) for each fieldwork component in the first year of the study are shown in (Table 2). We assume here that each group have marked the fieldwork independently. We also empirically checked that the marks do not include outliers, i.e. the difference between marks given for each fieldwork component from the mean did not exceed three times the value of the standard deviation computed from the overall marks given to this component (i.e. with a dispersion within 99.73% probability under the standard normal distribution). (Table 2) shows ICC coefficients between marks given by peer assessors and their self-assessment as well as between their marks with marks given by the tutors.

Table 1. Descriptive statistics of peer, self and tutors marks in the 1st year (as a percentage)

Component/score	subtask	Peer Vs. peer mean (μ) %	Peer Vs. peer (σ)%	Self Vs. peer mean (μ) %	Self Vs. peer (σ)%	Peer Vs. Tutors mean (μ) %	Peer Vs. Tutors std. dev. (σ) %
fieldwork	Correct instrument setup procedures used	3.4	3.0	-	-	3.0	3.04
	Maintained instrument setup (centring & levelling)	2.1	5.7	-	-	6.7	6.0
	Correct observation procedure used	3.05	6.8	-	-	8.0	4.7
	Closing/checking observations taken before leaving site	9.29	7.9	-	-	8.70	5.1
field recording	Use of paper/ digital field notes	2.7	2.8	-	-	4.6	4.0

		Peer Vs. peer mean (μ) %	Peer Vs. peer (σ)%	Self Vs. peer mean (μ) %	Self Vs. peer (σ)%	Peer Vs. Tutors mean (μ) %	Peer Vs. Tutors std. dev. (σ) %
	Observations recorded correctly as observed	3.6	2.4	-	-	5.8	5.3
	Clear & complete field notes presented	4.58	4.06	-	-	10.40	5.1
	Inclusion of index/cover page and north point	2.7	2.4	-	-	4.84	4.32
computation & analysis	Basic calculations & reductions performed	2.4	3.1	3.8	3.9	5.4	4.76
	Closure & checking calculations carried out	3.04	3.4	4.5	4.2	7.08	6.2
	Analysis of results (and their errors) performed	6.0	4.76	6.62	6.02	8.1	4.1
	Acceptable results achieved	10.0	6.3	8.2	6.8	8.9	4.1
presentation of results	Clear & well structured report, index, introduction & references	9.0	7.2	6.3	3.16	7.7	4.34
	Description of fieldwork & methodology	2.04	4.2	3.0	3.1	3.9	3.56
	Results, analysis & conclusion	3.8	4.74	7.0	4.1	8.9	3.68
	Required plans/ maps/ tables	8.0	7.1	3.94	3.4	7.18	6.50

Table 2. ICC coefficients among different groups of assessors in the 1st year

Component/score	Peer Vs. peer mean (μ) %	Peer Vs. peer (σ)%	Self Vs. peer mean (μ) %	Self Vs. peer (σ)%	Peer Vs. Tutors mean (μ) %	Peer Vs. Tutors std. dev. (σ) %
fieldwork	4.46	5.85	-	-	6.60	4.71
field recording	3.27	2.90	-	-	6.41	4.68
computation & analysis	5.36	4.39	5.78	5.23	7.37	4.79
presentation of results	5.71	5.81	5.06	3.44	6.92	4.52

Results of (Table 1) show that differences in marks given by different groups of students to their peers for the field components *fieldwork* and *field recording*, were on average limited to 4.46% and 3.27% with a dispersion of 5.85% and 2.74%. Different peer assessor groups had high agreement on field components such as *maintained instrument setup* and *correct observation procedure*, and a small discrepancy of 3% in the *correct instrument setup*. Their main disagreement came in the fieldwork component "*closing/checking observations taken before leaving site*" with 9.29% differences, which shows some biases in judging this task in terms of consideration of all fieldwork checks and their agreement with assessment criteria. From a professional point of view, assessment differences should be minimal for this component. Thus, these results indicate that, under the assumption that in peer-assessment students

accurately and objectively evaluate success or failure on a given task; there is a shortage in the explanation of this component to students.

Comparison between marks given by different groups of peer assessors with those of the tutors for the two fieldwork components *fieldwork* and *field recording* show that they were reasonably close with mean discrepancies of 6.6% and 6.41% and dispersion of 3.71% and 4.68%, which are still higher than those given by peer assessors. In general, students report higher marks compared with teachers. This came in agreement with findings of other studies (e.g. Langan et al., 2008). The fieldwork components that experience the highest differences were *closing/checking observations taken before leaving site*, and *clear & complete field notes presented* where discrepancies between tutors and peer assessors were 8.7% and 10.4% respectively. This indeed indicated that more clarification was needed concerning the assessment of these two components.

In the analysis of the inter-correlation results between different groups of assessors, we assume that ICC values greater than 0.8 would represent a high correlation, and that between 0.5 and 0.8 to represent a moderate correlation whereas a value below 0.5 would indicate a low correlation. (Table 2) shows very high correlations among peer assessment marks for the fieldwork and field recording components, and slightly lower correlation for those marks with the marks given by the tutors. However, in general results indicated that the criteria used for assessment were comparatively applied by different peer assessors.

For the two office work components *data analysis* and *presentation of results*, the discrepancies among different peer assessor groups were relatively significant and their dispersion was also large. The components that have the largest differences were *acceptable results achieved, clear & well-structured report elements* and *required plans/maps/tables* with average differences of 10%, 9%, and 8% respectively. Similarly, the differences between the student assessment and tutors were significant. This can be explained by differences in the width and depth of the experience of students and tutors. Also, within the group of peer assessors, not all students could have applied the same criteria in a consistent way. The high discrepancies and low correlations described here suggest that students and tutors still interpret some assessment criteria and indicators of the rubric in a different way. This indicates that more explanation/clarification is needed for the above components and the expectations of the tutors needs to be discussed in more details with students. This shows an important benefit of applying peer assessment, as it helps the tutors to identify the weak areas in the assessment design.

For the same test period, when comparing the marks of trained peer assessors with marks given by another group without training, results showed that their average mark differences were 7% and 10% for the field and office work components. Similarly, the discrepancies between the peer assessor group without training and the tutors ranged between 7% and 12% on average for the field and office work components, respectively. This finding suggests that training of students to carry out the assessment task is important.

When students were asked to self-assess their own work and next compare their marks with the marks given by their peers, we noticed moderate differences and correlations and that students underestimate work of their colleagues and overestimate their own. This is shown in the Tables 1 and 2 by the 5.78% and 5.06% differences in average marks given by self- and peer assessment of the components: *computation & data analysis* and *presentation of results*, with dispersion of 5.23% and 3.44% respectively, and moderate values for correlation coefficients of 0.742 and 0.648.

The lessons learnt from the application of the peer assessment process in the first year helped us to fine tune the process and the marking rubrics and to improve our explanation of the assessment criteria when the process was repeated for a second year (2012). Results of the second year are given in (Tables 3 and 4). The change from the results obtained in the first year was limited but overall positive as it showed less differences and higher correlations among peer, self and tutor assessments when comparing the Tables 1 with 3, and 2 with 4. In summary, the results presented here indicate that the peer assessment process is arguably reliable and valid, and can be considered a useful learning strategy.

Table 3. Descriptive statistics of peer, self and tutors marks in the 2nd year (as a percentage)

Component/score	Peer Vs. peer	Self Vs. peer	Peer Vs. Tutors
fieldwork	0.901	-	0.872
field recording	0.817	-	0.756
computation & analysis	0.722	0.742	0.688
presentation of results	0.668	0.648	0.355

Table 4. ICC coefficients among different groups of assessors in the 2nd year

Component/ score	Peer (without) VS peer with training mean (μ) %	Peer (without) VS peer with training (σ)%	Self (without) Vs. peer with training mean (μ) %	Self (without training) Vs. peer with training (σ)%	Peer (without training) Vs. Tutors mean (μ) %	Peer Vs. (without training) tutors std. dev. (σ) %
fieldwork	7.67	5.90	-	-	7.60	5.66
field recording	6.33	3.82	-	-	6.41	4.83
computation & analysis	9.46	6.10	6.42	4.22	11.41	6.23
presentation of results	10.54	6.89	6.04	4.87	12.58	5.47

Student involvement in constructive assessment of peers

This study serves in part an overarching objective of developing a pedagogical best practice in fieldwork training. Therefore, in our experiment, the following points were considered to support engaging students in a constructive assessment of their peers;

- Students were encouraged to engage in a constructive criticism of their peers, where they discuss their strengths, weaknesses and performance. Students were encouraged to identify how they could improve their own performance and how they could approach the task differently.
- Students were guided to carry out individual thinking and engage with the assessment criteria before meeting in the field. They were advised to discuss with their peers the objectives and requirements of each fieldwork assessment task before commencing it.
- Fieldwork assessment is performed in a small group of students and marks are given after discussion. This helps to strengthen objectivity of the assessment.
- Students openly discuss with their peers their individual contribution within the team work.
- Student assessors were not advised to openly voice their marks to others, as marks were confidential, so only the tutor would know how students had rated their peers.
- We target improving students' aspirations to achieve high standards by gradually introducing a system of "if not of a high standard - redo or fail". This is a more realistic representation of the industry environment students are aiming to join upon graduation.

Concerning students' perception of peer assessment, results of a survey conducted to involved students reflect a very positive attitude and appreciation towards the value of peer assessment, in particular, when applied as formative assessment.

Future work

Our future research work includes development of best-practice guidelines for peer assessment process of fieldwork with emphasis on requirements to achieve industry standards. For example, in the surveying discipline this targets meeting professional surveyors' Code of Practice (WA Dept. of Mines and Petroleum 2011). In addition, the current study remains limited when it comes to sample size, scope of the skills to be mastered and the complexity level of the competencies. Therefore, extension of the study will be performed to include a larger number of units and for a longer period.

We also plan to apply the marking tool of fieldwork activities in the field using mobile technology (iPads). The marking templates on iPads will be generated from a generic form adaptable to include activities for every laboratory and changes in the relative importance of tasks. iPads will be convenient for taking photos or videos for documentation needed in the assessment. The use of iPads with 4G will allow student queries to be sent directly to supervisors for remote discussion and to obtain instantaneous feedback. An example on the use of iPads for assessment is shown by Campbell and Wren, 2012, who presented an App that enabled the markers of live performances to focus their attention on the quality of learning, while the technology streamlined the assessment process.

Conclusions

In this study, peer assessment was demonstrated with focus on fieldwork evaluation. The level of agreement among peer assessments and with tutor assessments was studied for one unit for two years. The lessons learnt from the application of the peer assessment and marking/feedback tool in the first year helped us to enhance this tool and improve explanation of the assessment criteria and tutors' expectations, particularly when large differences were found in the assessment of any component. This resulted in a reduction in the marking gap among peer assessors and with tutors in the second year of the study.

Comparison among marks given by different peer assessor groups and with those of the tutors for evaluation of the field component showed that they were reasonably close. However, for the office work component, the discrepancies were relatively significant and their dispersion was also large. This can be explained by differences in the width and depth of the experience of students and tutors.

When comparing results of peer assessors who are trained with another group without training, the first group had more consistent results. The differences in marking between the two groups suggest that training of students to carry out assessment is important. On the other hand, when students are asked to self-assess their own work and compare their marks with the marks given by their peers, students moderately underestimate work of their colleagues and overestimate their own.

Acknowledgments

References

AlFallay, I. 2004. The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*: 32(3), 407–425.

Biggs, J. and Tang, C. 2007. *Teaching for quality learning at university: What the student does* (3rd ed.) Maidenhead, Berkshire: Open University Press.

Bostock, S. 2000. Student Peer Assessment. *The Higher Education Academy*, 1-4.

Brew, A. 1999. Towards autonomous assessment: Using self-assessment and peer assessment. In S. Brown and A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches:* 159-171. Buckingham, UK: Open University Press.

Bouzidi, L., and Jaillet, A. 2009. Can Online Peer Assessment be Trusted?. *Educational Technology & Society*, 12 (4), 257–268.

Brennan, R. 2001. An Essay on the History and Future of Reliability from the Perspective of Replications. *Journal of Educational Measurement*: 38(4), 295-317.

Campbell, K., Mothersbaugh, D., Brammer, C. and Taylor, T. 2001. Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly*: 64(3), 23–42.

Campbell, A. and Wren, J. 2012. Feedback on group performance using an iPad app. *Proceedings of the 21st Annual Teaching and Learning Forum, 2-3 February 2012. Perth: Murdoch University.*

Carless, D. 2007. Learning-oriented assessment: conceptual bases and practical implications. *Innovations in Education and Teaching International*: 44(1), 57–66.

Cho, K., Schunn, C., and Wilson, R. 2006). Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology*: 98(4), 891-901.

Crisp, G. 2007. *The e-assessment handbook*. London: Continuum.

Dave, R. H. 1975. *Developing and Writing Behavioral Objectives*. (R. J. Armstrong, ed.). Tucson, Arizona: Educational Innovators Press.

Davies, P. 2003. Closing the communication loop on the computerized peer-assessment of essays. *ALT-J Research in Learning Technology*: 11(1), 41-54.

De Grez, L., Valcke, M. and Roozen, I. 2012. How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments?. *Active Learning in Higher Education*, 13(2), 129–142.

El-Mowafy A., Kuhn M., and T. Snow. 2013. Blended Learning in Higher Education – Facing Current and Future Challenges in Surveying Education. *Issues in Educational Research,* 23(2), 132-150.

Falchikov, N., and 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.

Falchikov, N. 2005. *Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. New York: Routledge Falmer.

Falchikov, N. 2007. The place of peers in learning and assessment. *In D. Boud and N. Falchikov (Eds.), Rethinking assessment in higher education: Learning for the longer term: 128-143. London: Routledge.*

Freeman, M. and McKenzie, J. 2002. SPARK, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects. *British J. of Educational Tech.*: 33(5), 551-569.

Friedman, B.F., Cox, P.L. and Maher, L. E. 2008. An Expectancy Theory Motivation Approach to Peer Assessment. *Journal of Management Education*: 32(5), 580-612.

Greguras, G. J., Robie, C. and Born, M. P. 2001. Applying the social relations model to self and peer evaluations. *The Journal of Management Development*: 20(6), 508-525.

Gulland, E. K., El-Mowafy, A. and Snow, T. 2012. Marking moderation in land surveying units. *Proceedings of the 21st Annual Teaching Learning Forum, 2-3 February 2012*. Perth: Murdoch University.

Haaga, D. 1993. Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, 20(1), 28–32.

Harbaugh, A. 2012. Effective strategies to support online learning and student engagement. *Proceedings of the 21st Annual Teaching Learning Forum, 2-3 February 2012. Perth: Murdoch University.*

Isaac, R. Z., Wilfred, J. P. and Douglas, C. 2001. Leadership and motivation: The effective application of expectancy theory. *Journal of Managerial Issues*: 13(2), 3045-3695.

Langan, M., Shuker D., Cullen, R., Penney, D., Preziosi, R. and Wheater, P. 2008. Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment and Evaluation in Higher Education*: 33(2): 179–190.

Luca, J. and McLoughlin, C. 2002. A Question of Balance: Using Self and Peer Assessment Effectively in Teamwork. *Proceedings ASCILITE 2002, 1-4*.

Montgomery, K. 2000. Classroom rubrics: Systematizing what teachers do naturally. *Clearing House*: 73(6), 324-328.

Price, M. and O'Donovan, B. 2006. Improving performance through enhancing student understanding of criteria and feedback. *In: Bryan C and Clegg K (Eds) Innovative Assessment in Higher Education:* 100–109. *London: Routledge*.

Raban, R., Litchfield, A., Willey, K., Hermans A., Murray s., Davis D., Dreher H., Harris, N. and Wenham, K. 2009. Supporting student self and peer assessment in large group work projects, Final Report, ALTC, 2009.

Sadler, D. R. 1989. Formative assessment and the design of instructional systems. *Instructional Science*: 18, 119-144.

Sadler, P.M. and Good, E. 2006. The Impact of Self- and Peer-Grading on Student Learning, *Educational Assessment*: 11(1), 1–31.

Shrout, P. E. and Fleiss, J. L. 1979. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*: 86 (2), 420–428.

Strobl, J. 2007. Geographic Learning. *Geoconnexion International Magazine*: 6(5), 46-47.

Thomas., G., Martin, D. and Pleasants, K. 2011. Using self- and peer-assessment to enhance students' future-learning in higher education. *Journal of University Teaching & Learning Practice*: 8(1), 1-17.

Yorke, M. 2003. Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*: 45, 47–501.