

---

# Self and peer-assessment: a learner perspective

**Dobriila Lopez**

Manukau Institute of Technology  
Private Bag 94006  
Manukau 2241  
New Zealand  
mike.lopez@manukau.ac.nz

**Mike Lopez**

Manukau Institute of Technology  
dobriila.lopez@manukau.ac.nz

**Willem Fourie**

Manukau Institute of Technology  
willem.fourie@manukau.ac.nz

**Denise Clarkson**

Manukau Institute of Technology  
denise.clarkson@manukau.ac.nz

**Kirsten Marais**

Manukau Institute of Technology  
kirsten.marais@manukau.ac.nz

---

This quality assured paper appeared at the 2nd annual conference of Computing and Information Technology Research and Education New Zealand (CITREnz2010) incorporating the 24<sup>th</sup> Annual Conference of the National Advisory Committee on Computing Qualifications, Rotorua, New Zealand, July 6-8. Samuel Mann and Michael Verhaart (Eds).

**Abstract**

Self and peer assessment aid learning and build capabilities that support lifelong learning. However, although both involve learner participation and are often discussed together, they are very different in some respects. This study contrasts learners' experiences of peer assessment with that of self-assessment. It used a non-experimental post-test only design in which students enrolled in courses with both self-assessment and peer-assessment components completed a questionnaire on their experiences. Survey questions were formed into four scales: value, learning, contrasting peer with self and concerns. Additional questions captured basic demographics and conceptions of learning. Participants believed that they learned more from peer assessment than from self-assessment. However, some felt it was not right for other students to mark their work and some felt uncomfortable with the responsibility of marking other's work. Learners endorsed the notion that peer-assessment has a greater learning value than self-assessment and that they should be more involved in assessing other students. However, they also had concerns about the wider use of self and peer assessment and a minority was strongly opposed to their use.

**Keywords**

Self-assessment, peer-assessment, lifelong learning, learner perspective, student perspective, ASSIST.

## Introduction

In a comprehensive review of higher education in the UK, the UK National Committee of Inquiry into Higher Education (Dearing Report) expressed the vision that:

Over the next 20 years, the United Kingdom must create a society committed to learning throughout life. That commitment will be required from individuals, the state, employers and providers of education and training. Education is life enriching and desirable in its own right. It is fundamental to the achievement of an improved quality of life in the UK (Dearing, 1997, p8)

In order for learners to become an active part of such a learning society they have to adopt a "learning approach to life" and develop the judgments they make about learning as well as learning tasks (Boud, 2000). To help learners develop their judgemental skill requires educators to go beyond immediate course-related goals and view assessment in wider terms. In particular assessment should have a purpose of preparing students for activities of learning throughout life in a learning society.

To address this need, educators have developed new assessment practices to overcome the limitations of traditional examinations and tests and promote the learning and development of "learning-how-to-learn skills" (Boud & Falchikov, 2005). To be successful, such an approach needs to be incorporated into assessment practices at all levels, courses and programmes within an institution.

Engagement in peer assessment activities can help build a learner's judgement. Such assessment skills

"are needed by lifelong learners, not only in their continuing learning but also in contexts such as performance appraisal, team building, and so on, which need people who have become adept at assessing each other's work and contributing fairly, sensitively and appropriately." (Race 2001)

There are many definitions of what constitutes peer assessment. Topping (1998, p.250) defines it as "an agreement in which individuals consider the amount, level, value worth, quality, or success of products or outcomes of learning of peers of similar status." Van Der Berg, Admiral and Pilot (2006, p.19) propose that "student assessment is understood to be an arrangement with students assessing the quality of their fellow students' writings and giving feedback to each other." Sluijsmans et al, (2004) consider it as a skill that can and should be acquired through training and suggest that many implementations could fail if the training component is neglected.

Several researchers have commented on the advantages of peer assessment in higher education. Ballantyne et al (2002) suggest it "promotes the acquisition of life long skills due to the active involvement of students in the assessment experience" (p428). Cassidy (2006) advocates peer assessment as a means of improving employability skills, especially those related to non-technical aspects, including oral and written communication, reading, learning skills and strategies, problem solving, decision making, dependability, and responsibility. Sivan (2000) suggests that student involvement in setting assessment criteria contributes to their learning. Topping (2005, p.640) suggested that peer assessment

can enhance self-assessment and that both yield metacognitive gains.

Others have identified some potential disadvantages. Cassidy (2006) expressed concerns about the validity and reliability of the marks given by students due to lack of expertise, potential bias, discomfort with the extra responsibility, lack of formal training, and perception that is the job of the tutor to do assessment. Some of these elements resulted in students giving average marks to all peers regardless of the quality of their work (Topping, 2005). Topping also states that peer assessment is more reliable than self-assessment but it still lacks objectivity. Bloxam and West (2004) found no evidence of a relationship between students' capacity to mark their peers and their own performance on an assessment task. Other studies reported findings that have no difference in performance between two groups of students in the same unit, one adopting peer assessment, and the other not (e.g. Sluijsmans et al, 2004) without any obvious explanations.

Peer assessment requires the active engagement of students. For it to be successful, students must value both the goal of lifelong learning and the learning arising from the process of peer assessment as implemented in courses. Very little has been published in New Zealand from such a student perspective. Our study aimed to explore this student perspective by asking the following research questions:

- Do students see value in self-assessment?
- Does self and peer assessment help learning?

- Do participants learn more from peer-assessment than self-assessment?
- Should we do more self and peer assessment?
- What are the barriers to wider use?

### **The sample**

The sample was a non-random convenience sample in which we invited participation from students at a metropolitan polytechnic who were enrolled in three degree courses that had both self-assessment and peer-assessment components. Students were recruited at the end of a lecture by a staff member who explained the research and allowed sufficient time to consider participation.

Participation involved completing a questionnaire on their experience. Participation was voluntary and consent was implied by completion of the questionnaire. A total of 143 students were enrolled in the courses and 92 of these (64%) chose to participate. Because some of the authors had a teaching relationship with the participants, analysis of the questionnaires was not carried out until all course marks were finalised.

In the courses chosen, self and peer assessment were used primarily for their believed contribution to learning rather than as a means of generating marks. Where course marks were awarded for the assessment activities, they were allocated by the tutors in recognition of the quality of marking demonstrated in the assessment activities, rather than using any marks generated by the students themselves.

The first course was an introductory (level 5) programming course in which 80% of the course marks

were allocated to content and 20% to capabilities which included self and peer assessment. For the peer assessment component, one of the course assignments was marked by students on a double-blind basis. Self-assessment was carried out throughout the course.

The second course was a level 5 media design course in which self and peer assessment was used formatively in all group work and presentations.

The third course was a level 6 project management course in which self-assessment was carried out throughout the course and peer-assessment was used in two presentations.

### **The Instrument**

The instrument was a purpose designed composite questionnaire, with the first part (sections A, B, C), based on the ASSIST questionnaire (Tait, Entwistle & McCune, 1998). This part explored student conceptions and preferences and classified their approach to studying as surface, deep or strategic. The second part (section D) explored the students' experiences with self and peer-assessment and was based on the work of Boud and Holmes (1995). The third part (section E) captured basic demographic data and the final part (F) was an open ended comments question.

The main analysis in this paper is based on the questions in section D; these questions are set out in full in the appendix and are discussed below. The purpose of including the ASSIST questions and basic demographic questions was to investigate whether there were any associations between these and the responses to the questions in section D and thus hopefully gain insights into possible reasons for the

responses. We used the following variables from the ASSIST scales:

- Prefers courses with surface learning (Bsurf)
- Prefers courses with deep learning (Bdeep)
- Takes a surface approach to studying (Csurf)
- Takes a deep approach to studying (Cdeep)
- Takes a strategic approach to studying (Cstra)

#### *Scales and variables used*

The questions in section D were formed into a number of scales to address the key research questions. For consistency of interpretation, all scales and variables were standardised to the range 0 to 10 and centred at 5. A summary of the scales we used is given in the appendix. In this section, we address each of the research questions in turn and set out the rationale for the choice of scales.

#### DO STUDENTS VALUE SELF-ASSESSMENT?

From the perspective of lifelong learning, we believe, as practitioners, that assessment skills are needed throughout a learner's life and that part of our job as educators is to help build this capability. However, we thought it prudent to ask our students whether they agreed. For this scale we asked participants whether professionals need to be able to judge their own work (D24), whether participants should be able to assess their own performance (D01) and whether self-assessment was a good idea (D03).

#### DOES IT HELP LEARNING?

We also believed that engagement in self and peer assessment activities would directly help the students' learning. In what we termed the *direct* sub-scale, we asked whether they found assessing their own (D07)

and others' work (D08) valuable and whether they learned from their peers' feedback (D13).

In what we termed the *indirect* sub-scale, we asked whether the assessment activities helped participants make a realistic assessment of their abilities (D09), consider their learning progress more closely than usual (D10), focus on the value of the work done (D11) and be more aware of what they needed to know in the subject (D12). We also asked whether they found self-assessment more demanding than having someone else assess them (D02). The rationale for this last question was that we believed that those who took the assessment activities seriously and actively engaged in the tasks were more likely both to learn from the exercise and to find it demanding.

#### DO PARTICIPANTS LEARN MORE FROM PEER?

We partitioned the *direct* sub-scale mentioned above into two micro-scales. Learning from peer assessment was defined as finding assessing others' work (D08) valuable and learning from peers' feedback (D13). Learning from self-assessment was defined as finding assessing their own work valuable (D07). We then defined a variable *LearnedMorePeer* as the difference between these.

#### SHOULD WE DO MORE?

To help judge whether our students believed we should do more self and peer assessment, we defined a *DoMore* scale. We asked participants whether students should be more involved in assessing other students (D06) and whether we should have more opportunities for peer and self-assessment (D23).

To get a more balanced view, we offset the above by trying to identify a number of potential concerns or barriers to wider adoption. We asked participants whether they felt comfortable with the responsibility of marking their own (D21) and others' work (D20) and whether they thought it was right for other students to mark their work (D05). We also asked whether they believed they knew enough to judge whether work was good (D25) and whether they thought that it was the lecturer's job to mark work (D22).

#### BARRIERS TO WIDER USE

We set up the *change* scale to explore whether participants felt that change to our approach was needed before wider adoption. We asked participants whether they would like to see some changes in the procedure (D15), whether they had concerns around workload (D04) and whether they thought the rewards were sufficient for the time spent (D17). We also set up a *reject* scale to identify whether there was any solid opposition to wider adoption. We asked participants whether they found it difficult to follow the marking scheme (D14), thought the assessment scheme was unfair (D18) and thought the whole exercise of self and peer marking was a waste of time (D19).

#### OTHER VARIABLES

As the analysis unfolded, we realised that some students were strongly opposed to the use of self and peer assessment. To explore this further, we added another variable *a-posteriori*. We coded this membership variable (opposed) as 1 if the student agreed (or strongly agreed) either that self and peer was unfair (D18) or that it was a waste of time (D19) and as 0 otherwise.

## Method

The study used a non-experimental post-test only design in which participants completed a questionnaire on their experiences with self-assessment and peer-assessment.

The first four research questions were addressed by comparing the means of the relevant variables to the centres of the corresponding scales. We used bivariate correlations to explore associations between the study scales and demographic and descriptive variables.

All of these techniques require the use of interval level measurement so we began by converting the ordinal questionnaire data to interval level as discussed below.

### Interval level variables

When a scale is used to code responses to questions, we are generally confident in their ordinal properties (i.e. *strongly agree* is a stronger endorsement than *agree* which is, in turn, a stronger endorsement than *neutral*, etc.). When we allocate numeric codes to categories, we have less confidence in the interval properties of the codes should we interpret them as numbers. For example, does the difference between the third and fourth category represent the same difference in strength of endorsement as that between the second and third? However, such interval level measurement is required for valid arithmetic on variables.

We used a polytomous Rasch model (Rasch, 1960; Andrich, 1978) to create interval level variables from the categories; this is a stochastic model that identifies the maximum likelihood estimates of person and item threshold locations by simultaneous modelling of location estimates and the uncertainty in their location.

Figure 1 shows an illustrative probability density map showing the probability of the categories being chosen for any given location (strength of endorsement).

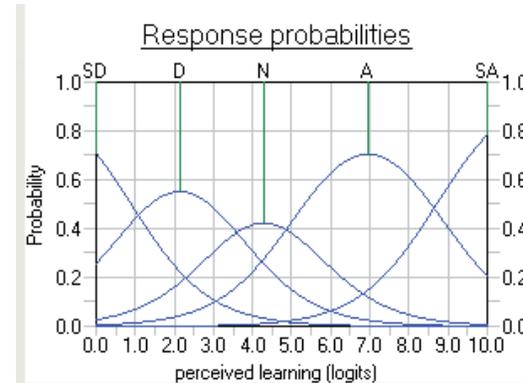


Figure 1: Sample response probabilities for item D08 in the learning scale

From this diagram, we can see that for scale locations between 0 and 1.1, the most likely response category is SD (strongly disagree), from 1.1 to 3.5, the most likely category is D (disagree), from 3.5 to 4.8 it is N (neutral), from 4.8 to 8.6 it is A (agree), and above 8.6 it is SA (strongly agree). The diagram illustrates both the varying intervals associated with the categories, and the clear ordinal pattern. We can also see the uncertainty in the scale (e.g. for imputed endorsement strength 2.3, the most likely category is D, but there is about a 20% chance of either SD or N being chosen).

A key assumption of the Rasch model is that the construct being measured is one-dimensional. We verified this with a principal components analysis as part of the data screening process.

#### Scale endorsement

Section D of the questionnaire used a conventional five point Likert scale. With such a scale, the neutral category is always confounded. For example, participants may have a true neutral position, may be unsure of the question, may feel the question is inappropriate or may feel it does not apply to them.

We did not attempt to force participants to respond to questions since the Rasch model handles missing data simply and naturally. For reporting responses at the individual question level, we wanted a simple measure of endorsement that worked well with missing data and made no distributional assumptions. We calculated endorsement as the proportion of those responding *Agree* or *Strongly Agree*, expressed as a percentage of the total responses, excluding the neutral category. This allowed us to calculate significance with a non-parametric exact binomial test.

#### Results

In this section, we begin by setting out the basic sample characteristics in Table 1. We then describe the screening carried out to verify that the observed dimensionality was consistent with use of the Rasch model and the analysis of response set and bias. After addressing the main research questions, we then explore associations between the study variables and background demographic data. The level of endorsement of each of the questions in section D is set out in the appendix.

#### Dimensionality of scales

For the Rasch model to extract the maximum information from a dataset, the scales must be uni-dimensional. We carried out a principal components

Table 1: Sample characteristics

		Count	Valid	%
Gender	Male	59	92	64%
	Female	33		36%
Age	Under 21	38	92	41%
	21-25	32		35%
	26-30	6		7%
	Over 30	16		17%
First Language	English	47	92	51%
	Other	45		49%
Main focus in prior six months	Work	20	91	22%
	Study	54		59%
	Other	10		11%
	Work and study	7		8%
	Missing	1		1%
First semester of study?	Yes	24	92	26%
	No	68		74%
NQF level of course	5	63	92	68%
	6	29		32%

analysis to verify this. All scales with the exception of the learning scale were uni-dimensional under Kaiser's criterion (Kaiser, 1960), suggesting that use of a Rasch

analysis is appropriate. However, the learning scale showed two distinct components under a variance maximising rotation; these components are shown in Table 2.

Table 2: Components of learning scale

Variable	Communality	F1	F2
D02	47%	<b>0.6576</b>	0.1917
D07	68%	0.0233	<b>0.8254</b>
D08	56%	0.4509	<b>0.5976</b>
D09	68%	<b>0.6455</b>	0.5096
D10	69%	<b>0.7711</b>	0.3069
D11	71%	<b>0.6836</b>	0.4932
D12	64%	<b>0.7802</b>	0.1846
D13	74%	-0.0432	<b>0.8603</b>
<b>Overall</b>	<b>65%</b>	<b>34%</b>	<b>31%</b>

On inspection, the first component seems to relate to the *indirect* sub-scale and the second to the *direct* sub-scale. Although the presence of two components suggests caution in the use of the overall learning scale, it should be noted, that because the *direct* and *indirect* scales are reasonably correlated ( $r_{(90)}=0.5767$ ;  $R^2= 33\%$ ;  $p<.0001$ ), and the un-rotated principal component accounts for 50% of the variability, the Rasch model is likely to extract enough information along this dimension to make use of the overall learning scale meaningful. However, to be prudent, we verified use of the overall *learning* scale in later analysis with tests of each sub-scale. The dimensionality of the scales is set out in table 3.

Table 3: Dimensionality of scales

Variable	Cronbach's Alpha	Variance explained by Principal Component
Valued	0.7028	58%
Learning	0.9013	50% (34% if rotated)
DoMore	0.8282	48%
Change	0.6121	53%
Reject	0.7272	65%
Direct	0.7733	62%
Indirect	0.9349	63%

#### Response set and bias

The ASSIST questionnaire does not use a balanced mix of negatively and positively scored questions. We attempted to get a broad indication of the level of potential bias by balancing positive and negative coding in the last 10 questions of section D and estimating the bias in these questions. Because these questions were towards the end of a relatively long questionnaire, we reasoned that bias, if present, was most likely to be apparent in these later questions.

*Acquiescence bias* is a tendency to agree with a statement rather than to disagree. To test this, we formed a scale by contrast coding agreement and disagreement. A two-tailed one sample t-test was not significant. ( $t_{(86)}=2.0013$ ;  $p=0.0459$ ).

*Centrality bias* is a preference for endorsing inner categories rather than more extreme categories. To test this, we formed a scale by contrast coding the outer "strong" categories with the inner categories. A two-tailed one sample t-test suggests ( $t_{(86)}=8.0912$ ;  $p<.0001$ ) that the mean was significantly greater than 5 with a confidence interval of  $CI_{.99} = (6.2195 \leq$

BiasCentre  $\leq$  7.3897). We concluded that there was significant centrality bias. The Rasch model used is not affected by centrality bias, but because of this bias, the categories were collapsed when forming direct measures of endorsement.

*Desirable response bias* is a tendency to give responses that would be seen as desirable (whether from impression management or self-deception). To test this, we formed a scale, using the difference between preferences for courses with opportunities for deep rather than surface learning as identified by the ASSIST scales Bdeep and Bsurf. Our logic was that because deep learning was an overt objective of the course lecturers, a desirable response bias would show stronger endorsement of the deep scale over the surface scale. A one sample t-test suggests ( $t_{(91)} = -3.5232$ ;  $p = 0.0008$ ) that the mean was significantly less than 5 with a confidence interval of  $CI_{.99} = (4.2850 \leq \text{BiasDesirable} \leq 4.8949)$ . The direction here is interesting; the data show a significant tendency to avoid giving a desirable response.

#### *Research questions*

For each of the first four research questions we used two-tailed one sample t-tests to test the hypotheses that the mean of the corresponding variable was close to the scale centre of 5.

#### DO STUDENTS VALUE SELF-ASSESSMENT?

The mean of the *valued* variable was significantly ( $t_{(91)} = 9.3691$ ;  $p < .0001$ ) greater than 5 with a confidence interval of  $CI_{.99} = (6.4232 \leq \text{Valued} \leq 7.5283)$ . We concluded that participants endorsed the value of self-assessment.

#### DOES IT HELP THEIR LEARNING?

The mean of the *learning* variable was significantly ( $t_{(91)} = 6.6994$ ;  $p < .0001$ ) greater than 5 with a confidence interval of  $CI_{.99} = (5.7427 \leq \text{Learning} \leq 6.6968)$ . Because there were two components in the learning scale, we verified this with separate tests for each sub-scale. Both the direct subscale ( $t_{(91)} = 5.8321$ ;  $p < .0001$ ) and the indirect sub-scale ( $t_{(91)} = 5.6785$ ;  $p < .0001$ ) were also significantly greater than 5; the confidence intervals were  $CI_{.99} = (5.7469 \leq \text{Direct} \leq 6.9655)$  and  $CI_{.99} = (5.6396 \leq \text{Indirect} \leq 6.7357)$ . We concluded that participants endorsed the belief that self and peer assessment helped their learning.

#### LEARNED MORE FROM PEER ASSESSMENT?

The mean of the *LearnedMorePeer* variable was significantly ( $t_{(91)} = 5.0687$ ;  $p < .0001$ ) greater than 5 with a confidence interval of  $CI_{.99} = (5.7132 \leq \text{LearnedMorePeer} \leq 7.2397)$ . We concluded that participants endorsed the belief that they learned more from peer assessment than from self-assessment. However, although 90% of participants believed they learnt from the other markers comments about their work (D13) and 86% found assessing another student's work valuable (D08), only 73% believed students should be more involved in assessing other students (D06). It should also be noted that 32% felt it was not right for other students to mark their work and 43% felt uncomfortable with the responsibility of marking others' work

#### SHOULD WE BE DOING MORE?

The mean of the *DoMore* scale was not significantly ( $t_{(91)} = 1.3147$ ;  $p = 0.1887$ ) different from 5. We concluded that participants did not endorse the wider use of self and peer assessment.

The mean of the *reject* variable was significantly ( $t_{(91)}=-5.3677$ ;  $p<.0001$ ) less than 5 with a confidence interval of  $CI_{.99} = (3.0934 \leq \text{Reject} \leq 4.3442)$ . We concluded that participants did not believe that we should reject self and peer assessment. However 14 of the 92 students (15%) were strongly opposed to the use of self and peer-assessment reporting either that it was unfair, or that it was a waste of time.

The mean of the *change* variable was significantly ( $t_{(91)}=3.6505$ ;  $p=0.0005$ ) greater than 5 with a confidence interval of  $CI_{.99} = (5.2112 \leq \text{Change} \leq 6.2854)$ . We concluded that participants felt that changes were needed before wider use of self and peer assessment could be considered.

#### ASSOCIATIONS WITH SAMPLE CHARACTERISTICS

There were no significant associations between the study variables and Gender, English as a first language or whether the study was taken in a student's first semester of study at the institution. There were also no significant associations with the course or the level of course. However, significant ( $p<.01$ ) associations were found with age and the experience of participants in the six months prior to the start of the courses

The need for change was endorsed by the under 21 category with a confidence interval of  $CI_{.99} = (5.7906 \leq \text{change} \leq 6.9478)$ . No other age category significantly endorsed the need to make changes to the approach.

Participants with workforce experience were more likely to express a preference for courses with opportunities for deeper learning ( $r_{(90)}=0.2764$ ;  $p=0.0075$ ) and a preference for a deeper approach to studying ( $\beta=0.2363$ ;  $t=2.8096$ ;  $r_{SP}=0.2135$ ;  $p=.0061$ ). They

were more likely to see value in self-and peer assessment ( $r_{(90)}=0.3032$ ;  $p=0.0034$ ). They were also more likely to endorse wider use of self and peer assessment ( $\beta=0.5077$ ;  $t=6.6894$ ;  $r_{SP}=0.5083$ ;  $p<.0001$ ) and less likely to be opposed to such wider use ( $\beta=0.4083$ ;  $R^2=17\%$ ;  $p<.0001$ ).

#### Barriers

We were concerned at the number of students who were unconvinced of the merits of self and peer-assessment. To quantify the task needed to achieve wider acceptance, we looked at the reject, change and DoMore scales as an indication of students' acceptance of self and peer-assessment. For each of these scales, we ranked responses in order of increasing acceptance of self and peer-assessment and identified the most likely response for each question at the lower quartile, median and upper quartile locations on these scales. Overall, the responses at the median location were favourable to wider use, but those at the lower quartile were not. We reasoned that a comparison of these responses would give us guidance as to what change in attitude would be needed to achieve acceptance by at least 75% of the students.

We identified those items for which there was a difference between the lower quartile and median responses. These are shown in table 4 together with the most likely response (A=agree, N=neutral, D=disagree) for the lower quartile (LQ) and median (M) locations.

Table 4: Target responses for wider adoption

Item	LQ	M
<b>Reject scale</b> • I found it difficult to follow the marking scheme (D16).	A	N
<b>Change scale</b> • I don't know enough to be able to judge whether work is good (D25)	A	N
<b>DoMore scale</b> • It's the lecturer's job to mark my work (D22)	A	N
• It's not right for other students to mark my work (D05)	N	D

### Discussion

Although the participants endorsed the idea that self-assessment was a valuable skill for professionals (and hence for their future careers) and reported that it helped their learning, they did not wholeheartedly embrace the proposition that we should be doing more of it, although neither did they go so far as to reject its use. However, participants who had come from the workforce were more in favour of the wider use of self and peer-assessment.

Overall, participants supported the notion that changes were needed before self and peer assessment should be used more widely, but on elaboration, we found that the perceived need for change was associated with age, with only younger participants believing change was needed. Perhaps a class discussion of the value of self and peer assessment that drew on the experience of older students, especially those with recent experience in the workforce, would help foster a change of attitude?

Before we can recommend wider use of self and peer assessment, we need to understand better why 15% of the participants were strongly opposed to its use and to find ways of addressing these students' concerns. We also need to find better ways of engaging with those who, although not strongly opposed, nevertheless remain unconvinced of the benefits. We take note that, although 86% found assessing another student's work valuable and 90% learnt from the other markers' comments about their work, only 73% felt students should be more involved in assessing other students.

In the quartile comparison, we identified four questions as critical indicators of acceptance of wider use of self and peer-assessment and the magnitude of change needed to achieve acceptance of their wider use.

We note that a number of students found it difficult to follow the marking scheme. We used the same marking schemes and criteria that were used for tutor assessment in the courses. These are detailed and specific and are shared with students early in the courses in the hope that students will clearly understand the scope and quality required for success in the assessment. We need to review these marking schemes. If students find them difficult to follow for self and peer-assessment, are they fulfilling their intended role when they are used for tutor assessment?

Many students felt that they did not know enough to judge whether work was good. Is this because they really do not know enough, or because they lack confidence in their judgement? Further investigation is needed to clarify this.

The remaining two questions seem to tap in to a conception of assessment as “something done to learners” rather than as a central part of learning and something that is undertaken as a partnership between tutors, learners and co-learners. For many of the students, the courses related to this study gave them their first encounter with both self and peer assessment. We would expect their confidence in these to grow as they progress through their studies, provided they are used in a consistent manner in successive courses and appropriate reflection and feedback was given to build confidence in the approach.

However, we need to identify strategies that will help mitigate the barriers identified in their first encounter.

### **Conclusion**

We remain convinced of the value of including self and peer assessment activities in our courses. Although the majority of our students agree with us, a minority remain unconvinced of the benefits or appropriateness of such activities. We are reluctant to recommend more widespread use until their concerns can be addressed.

We have gained some insights into why students may be concerned and have quantified the degree of change needed to overcome these barriers. Our next step is to devise strategies to mitigate these concerns and to evaluate their effectiveness.

### **Acknowledgements**

This research was supported in part by Contract no. TLRI-9233 awarded to Victoria University with sub-contracts to Massey University, the Manukau Institute of Technology, and Te Whare Wananga o Awanuiaranga

from the New Zealand Ministry of Education through the New Zealand Council for Educational Research

Portions of the questionnaire were based on the ASSIST questionnaire © 1997 Centre for Research on Learning and Instruction, University of Edinburgh. Some of the questions in section D were based on the work of David Boud and Harvey Holmes as reported in “Enhancing learning through self-assessment” © 1995 David Boud and contributors

### **References and Citations**

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-73.

Ballantyne, R., Hughes, K., & Mylons, A. (2002). Developing Procedures for Implementing Peer Assessment in Large Classes Using Action Research Process, *Assessment & Evaluation in Higher Education*, 27, 5, 427-441

Bera, A., Jarque, C. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6 (3): 255-259. doi:10.1016/0165-1765(80)90024-5.

Bloxham, S. & West, A. (2004). Understanding the Rules of the Game: Marking Peer Assessment as a Medium for Developing Students’ Conceptions of Assessment, *Assessment & Evaluation in Higher Education*, 29, 6, 721-733

Boud, D. and Holmes, H. (1995) Self and peer marking in a large technical subject, 63-78 in Boud, D. *Enhancing Learning through Self Assessment*, London: Kogan Page

Boud, D. (2000). Sustainable assessment: rethinking for the learning society, *Studies in Continuing Education*, 22, 2, 151-167

Boud, D. and Falchikov, N. (2005). Redesigning assessment for learning beyond higher education, In *Research and Development in Higher Education* 28. Brew, A. and Asmar, C. (Eds.), Sydney: HERDSA, 34-41

Cassidy, S., (2006), *Developing Employability Skills: Peer Assessment in Higher Education*, Education + Training, 48, 7, 508-517

Dearing, R. (Chair) (1997). Higher Education in the Learning Society, *The National Committee of Inquiry into Higher Education. Summary Report*.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Race, P. (2001). *A Briefing on Self, Peer and Group Assessment*. LTSN Generic Centre

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Sivan, A. (2000), The implementation of Peer Assessment: an Action Research Approach, *Assessment in Education*, 7.2.193-213

Sluijsmans, D., Brand-Gruwel, S., Van Merriënboer, J. & Martens, R. (2004), Training Teachers in Peer Assessment Skills: Effects on Performance and Perceptions, *Innovation in Education and Teaching International*, 41, 1, 59-78.

## APPENDIX – SECTION D OF THE QUESTIONNAIRE

Table 5: Endorsement of self and peer assessment questions

Variable	Endorsement	Margin of error	Sig.	
D01	The ability to assess my own performance is important	<b>97%</b>	± 4.44%	0.0000
D02	Self-assessment is more demanding than having someone else assess me	<b>88%</b>	± 9.09%	0.0000
D03	The idea of self-assessment is a good one	<b>89%</b>	± 8.85%	0.0000
D04	Self-assessment would be a good idea if we didn't have such a high workload	<b>88%</b>	± 9.62%	0.0000
D05**	It's not right for other students to mark my work	<b>32%</b>	± 13.34%	0.0043
D06**	Students should be more involved in assessing other students	<b>73%</b>	± 13.46%	0.0006
D07	I found assessing my own work to be valuable	<b>81%</b>	± 11.51%	0.0000
D08**	I found assessing another student's work valuable	<b>86%</b>	± 9.86%	0.0000
D09	This exercise helped me make a realistic assessment of my own abilities	<b>93%</b>	± 7.16%	0.0000
D10	I had to consider my own learning progress more closely than usual	<b>96%</b>	± 5.63%	0.0000
D11	Having to justify my mark helped me focus on the value of the work done	<b>94%</b>	± 6.83%	0.0000
D12	This exercise made me more aware of what I need to know in this subject	<b>97%</b>	± 4.70%	0.0000
D13**	I learnt from the other markers comments about my work	<b>90%</b>	± 8.73%	0.0000
D14	I found it difficult to follow the marking scheme	39%	± 14.55%	0.1237
D15	I would like to see some changes in the procedure	59%	± 17.24%	0.2912
D16	A similar (but improved) scheme should be used in other courses	<b>84%</b>	± 11.42%	0.0000
D17	I don't think the rewards were sufficient for the amount of time I spent	53%	± 15.39%	0.7914
D18	The assessment scheme was unfair	<b>14%</b>	± 10.88%	0.0000
D19	The whole exercise of self and peer marking was a waste of time	<b>17%</b>	± 10.48%	0.0000
D20**	I did not feel comfortable with the responsibility of marking others' work	43%	± 13.76%	0.2820

D21	I did not feel comfortable with the responsibility of marking my own work	44%	± 13.81%	0.4030
D22	It's the lecturer's job to mark my work.	60%	± 13.80%	0.1143
D23	We should have more opportunities for peer and self-assessment	<b>78%</b>	± 12.74%	0.0000
D24	Most professionals need the ability to judge their own work.	<b>92%</b>	± 7.11%	0.0000
D25	I don't know enough to be able to judge whether work is good.	65%	± 13.34%	0.0154

Questions marked \*\* relate expressly to peer assessment. The margin of error is calculated from the standard error of the binomial at a 99% confidence level. Response endorsement that is significant at the  $p < .01$  level is shown in bold. (Non-significant endorsement is offset to the right.)

Scales defined

Variable	Concept	Definition
Valued	The concept of self-assessment is valued	+D01 +D03 +D24
Learning	Self and peer assessment helped learning	+D02 +D07 +D08 +D09 +D10 +D11 +D12 +D13
DoMore	We should do more self and peer assessment	-D05 +D06 -D20 -D21 -D22 +D23 -D25
Change	We need to change the approach before we do more	+D04 +D15 +D17
Reject	We should not use self and peer assessment	+D14 +D18 +D19

Sub-scales

Direct	Learning happened directly from assessment activity	+D07 +D08 +D13
Indirect	Assessment activity indirectly helped learning	+D02 +D09 +D10 +D11 +D12
LearningPeer	Learning happened directly from peer assessment activity	+D08 +D13
LearningSelf	Learning happened directly from self-assessment	+D07