

2011

Measuring Student Support for Participative Assessment and Related Strategies: Development and Testing of the Beliefs about Participative Assessment Instrument (BAPAI)

Christine Brew

Nipissing University, Canada, chrisbrew@brbconsulting.ca

Philip Riley

Monash University, philip.riley@monash.edu

Recommended Citation

Brew, C., & Riley, P. (2011). Measuring Student Support for Participative Assessment and Related Strategies: Development and Testing of the Beliefs about Participative Assessment Instrument (BAPAI). *Australian Journal of Teacher Education*, 36(4).
<http://dx.doi.org/10.14221/ajte.2011v36n4.4>

This Journal Article is posted at Research Online.
<http://ro.ecu.edu.au/ajte/vol36/iss4/3>

Measuring Student Support for Participative Assessment and Related Strategies: Development and Testing of the Beliefs about Participative Assessment Instrument (BAPAI)

Christine Brew
Nipissing University, Canada
La Trobe University
christineb@nipissingu.ca
Philip Riley
Monash University

Abstract: Participative assessment is well established in higher education. However, students' concerns about the appropriateness of participative assessment create resistance to successful implementation. Strategies for addressing student concerns are needed because participative assessment practices appear to improve learning outcomes. The literature lacks validated scales to measure the subjective support for participative assessment. Presented are validated scales measuring support for self- and peer-assessment, group assignments and a proposed correlate, support for a discussion-oriented classroom derived from the responses of 213 pre-service teachers (both first year under-graduates and graduates). Graduates were more supportive of self- and peer-assessment than first year undergraduates and level of support for group assignments and self-assessment were the best predictors of support for peer-assessment. The scales provide: a measure to determine student support prior to participative assessment implementation; and, systematically collected data to initiate dialogue with students about their concerns.

Introduction

Participative assessment is increasing across a range of disciplines in both undergraduate and graduate courses (Ballantyne, Hughes, Mylonas, 2002; Brew, Riley & Walta, 2009; Dochy, Segers & Sluijsman, 1999; Pope, 2001). Despite an initially slow start by academics to embrace the use of self- and peer-assessment

(Williams, 1992), the shift away from exclusive instructor assessment practices is believed to improve pedagogy in three important ways: firstly, by encouraging reevaluation of the relationship between learning and assessment in the context of a growing demand for lifelong learners and reflective practitioners (Dochy et al., 1999); secondly by emphasising the dual role of assessment with respect to formative and summative needs (Groundwater-Smith, Cusworth & Dobbins, 1998); and, thirdly by moving assessments away from student reproduction of discipline knowledge and displays of memory rather than learning, to demonstrations of problem-solving, communication and presentation skills (Dochy & McDowell, 1997).

The debate in the literature concerning the implementation of peer- and self-assessments is substantial. A recent search of the ERIC database returned 1169 peer reviewed articles on the descriptor “peer evaluation” alone. The benefits and disadvantages of self- and peer-assessment are discussed in several reviews of the literature (see Falchikov & Boud, 1989; Topping, 1998; Dochy, et al., 1999; Falchikov & Goldfinch, 2000; Hanrahan & Isaacs, 2001). Broadly, these authors describe the current research as having established that students benefit through: becoming more confident, independent and reflective; obtaining a deepened understanding of the learning that is required; reducing the mystery of assessment and getting more immediate feedback.

While the notion of allowing students to assess themselves and each other may seem odd to the novice, the considered perspective emerging from the literature is that the advantages of participative assessment outweigh the shortcomings and concerns (Falchikov & Goldfinch, 2000; Pope, 2005; Ballantyne, et al., 2002). Two examples illustrate this point. In a meta analysis of 48 studies comparing teacher and peer assessments Falchikov and Goldfinch (2000) reported that peer and teacher assessments were closely aligned when they consisted of global judgements based on well understood criteria, or when academic products rather than professional practice were examined. This finding highlights that any assessment practices should be carefully targeted to the requirements of the learning and that no assessment model is universally applicable. In a more recent review of the literature Nulty (2010) found that peer assessments were successful in aiding student learning and argued for their use to be increased. Ballantyne and colleagues (2002) also reported that even in large classes the benefits of peer assessment for student learning outweigh the difficulties of fair implementation. However, Pope (2005) noted that both peer- and self-assessment tended to increase student stress, particularly among females. The authors argued that this may not in fact be a negative outcome. What can be inferred from this is that the issues are complex and the catchall terms peer- and self-assessment cover a wide range of procedures that are more or less applicable

to specific learning situations. Therefore, making a general statement about the value, reliability and validity of self- and peer-assessment is likely to be unwise.

Peer-assessment may include assessment by peers of students' written work (Pope, 2001), poster presentations (Ballantyne et al., 2002), or oral presentations (Langan et al., 2005), within a range of frameworks including individuals assessing individuals, groups assessing individuals or groups assessing groups (Falchikov & Goldfinch, 2000). Within each of these approaches the training of students with respect to assessment criteria is reportedly haphazard (Sluijsmans, Brand-Gruwel, & van Merriënboer, 2002). Student involvement in setting criteria may be absent (Brew, 2006); shared with the evaluator (McConnell, 2002); or students themselves may define the marking scheme (Stefani, 1994). Grades have also been derived in various ways including: global or holistic measures (Lejk & Wyvill, 2001); students awarding numerical value to set criteria (Falchikov & Goldfinch, 2000); or the rank ordering of students from best to worst (Kane & Lawler, 1978). Within group work, peer-assessment may involve intra-group or inter-group assessment to both processes and outcomes (Hanrahan & Isaacs, 2001; Carter, Howlett, & Daruwalla, 2004). A system of adjusting for bias in peer-assessment to enhance reliability has also been developed (Li, 2001).

In earlier research, Larisey (1994) contends that while there can be considerable student resistance to peer assessment, perhaps partly due to unfamiliarity with the process, the benefits for learning outcomes remain, and hence we need to address student concerns by normalising these practices within our education institutions. Topping (1998, p.269) concurred by summing up, "[O]rganised, delivered and monitored with care, peer assessment can yield gains in the cognitive, social, affective, transferable skill and systematic domains that are at least as good as those from staff assessment".

With respect to pre-service teacher education specifically, pre-service primary teachers in the context of mathematics have reported valuing peer assessment because it provided the opportunity to develop their knowledge of assessment techniques and that it "was useful for learning about other aspects of mathematics education" Zevenbergen (2001, p. 106). Broad support for participative practices based on pedagogical considerations was also reported among teacher educators by Brew, Riley and Walta (2009) though staff support for peer assessment was also linked to time saving in this study. Furthermore staff members were found to be far more supportive of participative assessment practices than their students (Brew et al., 2009).

The reported disadvantages of self- and peer-assessment centre on issues of validity. Threats to validity include the awarding of grades based on friendships, or students not being or feeling capable and it can also be time consuming (e.g., Cheng & Warren, 1997; Searby & Ewers, 1997; Brindley & Schoffield, 1998; Davies, 2000; Topping, Smith, Swanson, & Elliot, 2000). Findings from these

studies and explicit recommendations from others are starting to provide direction for higher educators to develop summative self- and peer-assessment processes that have reasonable validity and valuable learning outcomes for students (Falchikov & Goldfinch, 2000; Pope, 2005). With respect to self-assessment, evidence suggests better agreement with faculty-derived scores for students in advanced compared to introductory courses (Falchikov & Boud, 1989), though this pattern was not evident for peer-assessment (Falchikov & Goldfinch, 2000). Of course reliability and validity issues are central issues with respect to traditional forms of assessment and have been challenged on this basis (Rowntree, 1987; Newstead and Dennis, 1994).

With respect to the main philosophical underpinnings of participative assessment, Topping (1998) proposed that social constructivism was relevant through the notion of co-construction of knowledge through discourse and interactivity. Topping (1998) also suggested that consideration be given to the Vygotskian concept of scaffolding learning in which the learner of another is supported by a more competent other. In this way a greater focus on the capability of the assessor would ensue and is consistent with calls for greater training of students in these approaches before their use (Sluijsmans et al., 2002). Reynolds and Trehan (2000, p.268) argue that the shift away from mainstream assessment practices is more often than not based on “humanistic, student-centred aspirations for social equality, rather than on an analysis of the assessment process in terms of institutional power.” They argue that this is surprising given that one of the central functions of assessment is to provide the basis for granting or withholding of qualifications and thereby, “the most political of all educational processes” (Heron, 1981, p.13). Questions of power and authority within participative assessment are equally problematic. In the following three excerpts drawn from student interviews we can infer that the power of peer-pressure influenced the assessment outcomes:

At the beginning we were careful not to be too critical of other groups because we feared retaliation, i.e. if we gave them a low mark they would repay the compliment. Nothing was stated but it appeared to be an almost unwritten rule. (Reynolds & Trehan, 2000, 273)

While peer-assessment is a valuable learning and teaching experience, I think it can be unfair. Working in small groups means there is no way to ensure balanced marking standards across the classes. I am also aware that some grades have been awarded on the basis of friendship rather than merit which clearly undermine those people that deserve good marks in the first place. (Brew, 2006, p. 150)

Early in the year the focus is getting along with people, no skin off my nose if everyone gets an A. I am happy to give constructive feedback but I want to get along with the group, these are people who I am going to rely on for the rest of the year. (Brew, 2006, p. 150)

In contrast to a plethora of validity and reliability studies, generally speaking evaluations of the experiences of students' views towards participative assessment practices has received far less attention (Ballantyne, et al., 2002), though there are a number of studies that have provided a basis for discussion of issues. For example, following a review of the literature reporting the adoption of peer- and self- assessment in the first year of university Nulty's (2010) found it to be a useful learning tool, not without some problems, in particular the paucity of evidence on the stability of judgements. He came to the view that the use of peer- and self- assessment at first year level should be increased. However, Cheng and Warren (1997) reported that students agree in principle with peer assessment but believe that first years should not be involved. They also noted that some students do not feel qualified to award marks, which was related to a concern about the level of subjectivity in their own and peer' marks and a belief that they should not be given such a responsibility. Kwan and Leung (1996) found a similar concern among some students who thought that giving each other grades was "risky and unfair" and Hanrahan and Isaacs (2001) documented a level of hostility from students towards peer-assessment which Rushton, Ramsey & Rada (1993) noted increased following greater exposure to the practice. Certainly participative assessment does presuppose that students value the opinions of their peers about their work and the learning process and this has been shown to increase throughout a higher education degree as students grapple with and begin to embrace the relativity inherent within all knowledge (Perry, 1970; Baxter-Magolda, 1992). In the context of teacher-education programs, participative assessment through its ability to take the mystery out of the assessment process (Brindley & Scoffield, 1998) can imbue students with a better understanding of the inherent subjectivities in the grading process (Hanrahan & Isaacs, 2001). This awareness is particularly desirable to develop in teachers.

Cooperative learning through group assignments in which three or more students prepare and are evaluated on a joint piece of work is firmly established as a practice in higher education due to its promise of team-building and improved communication skills (Sharp, 2006). Conversely, concerns with group work often centre on assessments where group members receive the same grade regardless of individual input. Diminished effort, sometimes referred to as 'social loafing' through group membership has been described in some detail in the literature (Ingham, Levinger, Graves, & Peckham, 1974; Karau & Williams, 1993; Latane, 1981), as has groupthink which may also be a risk when group work based

assessment is used (Janis & Mann, 1977; Janis, 1982). Many students would rather have an individual grade that fairly reflects their contribution (Habeshaw, 1989; Topping, 1998; Chin & Overton, 2003).

When small group assignments are set, some or all of the assessment is commonly associated with peer evaluation (e.g. Brindley & Scoffield, 1998; Orsmond & Merry, 1996) to take advantage of the effectiveness of learning in groups as well as to maximise fairness of marks among students (Conway et al., 1993; Goldfinch, 1994; Pope, 2001). Given the apparent connection between the group assignments and peer assessment, an investigation of student views towards participative assessment practices warrants consideration of the inter-relationships between support for group assignments, peer-assessment and a discussion-oriented classroom warrants consideration. The latter should reflect a belief in the critical role of peers in learning.

Despite a considerable focus in the empirical literature on self- and peer-assessment that has involved validity studies; e.g., comparing self and peer assigned scores with those assigned by staff for the same assignment (see the meta studies of Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000), there is a lack of reliable scales to systematically measure support for self- and peer-assessment with a focus on students' beliefs. The use of these assessment practices is on the rise in higher education and the level of support is considerably higher among staff compared to students (Brew et al., 2009). Systematically collected information pertaining to these issues with respect to a range of independent variables is likely to support teachers in their quest to respond to specific concerns within their cohorts. Data reported here are based on survey data designed to develop reliable scales for level of support for self- and peer-assessment, group assignments and a proposed correlate, support for a discussion oriented classroom.

Purpose of the study

The study sought to develop a reliable instrument to measure pre-service teachers' beliefs about and support for peer and self-assessment, group assignments and support for a discussion-oriented classroom. Several research questions were also explored in association with the data:

1. What is the level of support among pre-service primary teachers regarding self-, peer-assessment and group assignments?
2. Is support for these practices influenced by level (graduate versus first year undergraduate) and gender?

3. Which constructs (support for self-assessment; support for group assignments; or support for a discussion-oriented classroom) best predict the level of student support for peer-assessment?

Method

Participants

A total of 434 students from three cohorts of Australian pre-service primary teachers were invited to participate in a survey; first years enrolled in a four year Bachelor of Education program (200 students) and graduates enrolled in a one year Diploma of Education from two campuses (234 students). The 213 respondents (49% response rate) reflected the makeup of the larger cohorts being relatively homogeneous with respect to gender (78% women), locality (94% local Australian students) and language background (93% first language English). Eighty percent of students reported their age, of these the majority (63%) were less than 25 years of age, ($M = 25$ years, $SD = 7$). On average graduates were seven years older than undergraduates, and had a higher proportion of men (29% versus 12%). On average men were five years older than the women. Surveys were administered to the students in 2006 at the end of their first semester. Results from one cohort of graduate students have previously been reported (Brew, 2006) but are included to provide a larger sample and to allow for a comparison across levels. Both graduate and under-graduate students had experienced summative peer-assessment in their courses that was associated with group-assignments and had received minimal training with respect to the pre-set criteria. Self-assessment had been experienced by some of the graduate students in connection with peer-assessment and for formative purposes.

Beliefs About Participative Assessment Instrument (BAPAI)

Twenty-two survey items were developed based on issues in the literature and from discussions with academic colleagues engaged in alternative assessment practices to measure students' beliefs about peer-assessment, self-assessment and group-assignments and a proposed correlate, support for a discussion-oriented classroom (Table 1). Nine peer assessment items explored the following themes: perceived validity, preparedness to engage, comfortableness with the process, perceived capability, curiosity regarding peer feedback, and issues regarding perceived responsibility for assessment. Four self-assessment items explored preparedness to engage, perceived validity, capability and learning opportunity. Five group assignment items explored perceived validity, learning opportunities and likelihood of uneven workload distribution. Four items explored level of

support for a classroom environment that encourages peer discussion, debate and sharing of ideas. A six-point scale was adopted: 1 = strongly disagree, 6 = strongly agree.

Data analysis

Analysis of the student data proceeded in six distinct phases. First an exploratory factor analysis (EFA) assessed the presence of latent factors (SPSS 16.0). Although the sample size was less than ideal (Pallant, 2007; Tabachnick & Fidell, 1996) the EFA was used to construct factors that could be assessed using confirmatory methods. The maximum likelihood estimation procedure was selected as it is recommended for use with ordered categorical data (Conroy, Motl, & Hall, 2000). Three methods were adopted to justify the number of latent factors: scree test; parallel analysis (Thompson & Daniel, 1996) and theoretical interpretability. An oblique rotation (Direct Oblimin) assisted in determining the items associated with the factors. Second, the internal reliability of the constructs (Cronbach Alpha scores) was assessed and some item trimming occurred. Third, confirmatory factor analyses (CFA) conducted in AMOS 6.0 assessed goodness of fit indices. There is a range of statistical measures used to assess model fitness and the consensus across the literature is to incorporate a range drawn from the absolute, incremental and parsimonious fit categories (Holmes-Smith et al., 2004). These fit measures along with requirements for good fit included the root mean square residual (RMR $\leq .05$) (Hu & Bentler, 1995), the Chi-square goodness of fit statistic (CFMIN/df, range 1-3), the root mean square error of approximation (RMSEA $\leq .05$) the goodness of fit index (GFI $\geq .90$), the adjusted goodness of fit index (AGFI $\geq .90$) (Kline, 2005) and the comparative fit index (CFI $\geq .95$) (Byrnes, 2001).

The fourth phase examined the inter-correlations between the constructs (both in AMOS where the relative influence of items is accounted for and in SPSS using the mean scores for the contributing items), and a multiple regression analysis in AMOS to identify the strongest predictors for support for peer-assessment. The fifth phase involved a comparison of means to determine the extent to which support for self- and peer-assessment, group assignments and a discussion-oriented classroom varied by gender and level. Effect sizes (partial eta squared) for each are also reported.

Results

Exploratory Factor Analysis

The EFA indicated a factor structure worthy of exploring (Kaiser-Meyer-Olkin measure of sampling adequacy = 0.825, a minimum of 0.6 required; Bartlett's test of Sphericity = 1582.92, $df = 231$, $p < .001$). The scree plot suggested a four factor solution, accounting for 51.2% of the variance. The parallel analysis also indicated a four-factor solution as the eigen value for the fifth factor (1.09), was exceeded by the corresponding random eigen value (1.25). The nine items designed to measure support for peer-assessment associated most strongly with factor 1; the four items designed to measure support for a discussion-oriented classroom strongly associated with factor 2; the four items designed to measure support for self-assessment associated strongly with factor 3 though one item also associated with factor 1; and three of the five items designed to measure support for group assignments associated strongly with factor 4. The group assignment item, *I don't need to work as hard on a group assignment compared to individual assignments*, did not load sufficiently on any factor and the item, *Classroom group work gives me feedback on my strengths as a team player* was associated with factor 2 (Table 1).

Internal reliability measures

Of the nine peer assessment items seven were retained providing a robust Alpha value of 0.86. For the four discussion items all contributed to the Alpha score of 0.75, as did the four self assessment items: Alpha 0.71. Both Alpha scores are acceptable given the small number of items. The three group assessment items provided a modest Alpha of 0.63. Final items that measured the four constructs are marked in Table 1, as are those that were recoded for consistent directionality.

Item	Assessment Type			
	Peer F1	Discussion* F2	Self F3	Group F4
+1: I am prepared to assess my peers knowing that it would effect their final grade	.77			
+2: I am happy for peer-assessment to count in my final grade	.72			.20
+3: I trust my peers to assess me accurately	.66			.24
+4: Assessment of my work should be solely the responsibility of the lecturer/tutor®	.63		-.20	
+5: Peer assessment is not valid®	.58			
+6 I have enough to learn without having to be involved in the assessment of my peers®	.52	.12	-.24	
+7 Peer assessment is a way for teachers' to reduce their assessment responsibilities®	.40	.10	-.15	.13
+8 I am curious about my peers' perceptions about the quality of my work	.38	.16		.12
+9 I don't know enough to be involved in the assessment of my peers®	.35			
+10 I like classes that allow time for discussion		.89		
+11 I like classes where we debate different perspectives		.65		
+12 I like classes where I am encouraged to express my ideas		.57		
+13 I prefer classes where students don't talk much®		.52		
+14 Classroom group work gives me feedback on my strengths as a team player		.38	.15	.16
+15 I have enough to learn without having to be involved with my own assessment®			-.84	.20
+16 Self assessment is not valid®	.25	-.16	-.58	
+17 It would contribute to my own learning to be given the opportunity to have input into how I am assessed			-.50	
+18 I don't know enough to be involved in my own assessment®	.30		-.37	-.14
+19 I don't need to work as hard on a group assignment compared to individual assignments				
+20 I don't like group assignments because I end up doing most of the work®		.11		.70
+21 Group assignments provide a valuable way to learn from peers				.59
+22 Group assignments are difficult to assess fairly®				.50

Extraction Method: Maximum Likelihood. Rotation Method: Oblimin with Kaiser Normalization. Rotation converged in 8 and 6 iterations for the 3 and 4 factor solutions respectively: Correlation coefficients <0.1 deleted for clarity. + Items retained. ® recoded. * Proposed correlate of assessment types.

Table 1: EFA Pattern Matrix four factor solution

Confirmatory Factor Analysis

The initial CFA provided both poor and fair fit measures (CFMIN/df = 1.878; RMR =.07; GFI = .87; AGI = 0.84; RMSEA = 0.06; CFI = 0.90). To improve the fit measures the standardised residual covariances (SRC) were examined for pairs of items with values >2.58 (Byrne, 2001). One pair (items 14 & 21) had a SRC of

3.307. Given that item 14 was designed to measure support for Group Assessment and not Support for a Discussion- Oriented Classroom it was selected for deletion. The subsequent fit indices were an improvement but still not satisfactory (CFMIN/df = 1.827; RMR =.07; GFI = .88; AGI = 0.85; RMSEA = 0.06; CFI = 0.91). Modification indices provide information to improve goodness of fit through the process of item reduction or scale revision (Vandiver & Worrell, 2002). The modification index for the errors associated with items 4 and 6 was identified as problematic (MI = 18.03) and item 6 was chosen for trimming having the lower coefficient of the two items in the EFA and reasonable fit indices were then obtained (CFMIN/df = 1.59; RMR =.06; GFI = .91; AGI = 0.87; RMSEA = 0.05; CFI = 0.94). Test for normality of items obtained skewness values $\leq|1|$; and kurtosis values $\leq|2|$ indicating the data were not significantly different from normality (Tabachnick & Fidell, 1996).

	Self	Group	Discussion
Peer ¹	.54 ^{***}	.57 ^{***}	.37 ^{***}
Peer ²	.45 ^{***}	.43 ^{***}	.30 ^{***}
Self ¹		.29 ^{**}	.24 ^{**}
Self ²		.17 [*]	.17 [*]
Group ¹			.24 [*]
Group ²			.17 [*]

* p<.05; ** p<.01, *** p<.001.¹correlations calculated from the varying contribution of the items to the latent factor- Method 1; ²correlations calculated from the average of the scores for the contributing items – Method 2. Peer – Support for peer-assessment; Self – Support for Self Assessment; Group – Support for Group Assignments; Discussion - Support for a Discussion-oriented classroom.

Table 2: Overall Construct Inter-correlations

Construct inter-correlations and means

Two methods were used to assess the construct inter-correlations. Method 1 was more accurate as the correlations were calculated from the varying contribution of the items to the latent factor. Method 2 is a more commonly used, though less accurate approach, in which the correlations are calculated from the average of the scores for the contributing items. This method assumes each item is contributing to the latent factor in equal measure. The second method was included because the comparison of means by gender and level were based on the average scores. For Method 1, positive and significant correlations were obtained overall between the four constructs (Table 2) which were highly consistent in magnitude by gender, with variations of note by level (Table 3).

	Self				Gp.				Disc.			
	W	M	G	UG	W	M	G	UG	W	M	G	UG
Peer ¹	.56***	.54***	.56***	.46**	.56***	.57***	.52***	.67**	.33**	.37***	.39***	.26
Peer ²	.45***	.43**	.46***	.37***	.44***	.39**	.39***	.45***	.24**	.48***	.35***	.16
Self ¹					.24	.29**	.29*	.11	.25*	.24**	.08	.31
Self ²					.15	.24	.18*	.04	.18*	.09	.09	.24**
Gp ¹									.25*	.24*	.16	.34
Gp ²									.18*	.02	.10	.20

* p<.05; ** p<.01, *** p<.001. ¹ correlations calculated from the varying contribution of the items to the latent factor; ² correlations calculated from the average of the scores for the contributing items. W = Women; M = Men; G = Graduates; UG = First Year Undergraduates. Peer – Support for peer-assessment; Self – Support for Self Assessment; Gp. = Support for Group Assignments; Disc. = Support for a Discussion-oriented classroom.

Table 3: Construct inter-correlations by gender and level

The construct measuring support for peer-assessment was strongly correlated with the constructs measuring support for self-assessment and group assignment with a modest correlation with support for a discussion-oriented classroom. The other construct correlations were modest. By level, the first year undergraduates were less likely than the graduates to convey a relationship between support for peer-assessment and a discussion oriented classroom. Graduates on the other hand did not perceive a relationship between support for group-assignments and a discussion-oriented classroom. For Method 2, the averaged construct scores produced eclipsed inter-correlations to those obtained from Method 1. Since the patterns evident by gender and level are overall preserved, this provides reasonable confidence in the validity of the mean comparisons.

Construct	SRW	p
Group	.43	***
Self	.40	***
Discussion	.18	*
Level	-.09	ns
gender	.05	ns

* p<.05, *** p<.001; SRW = Standardised Regression Weight

Self – Support for Self Assessment; Group – Support for Group Assignments; Discussion - Support for a Discussion-oriented classroom

Table 4: Accounting for the variance in level of support for peer-assessment

The multiple regression analysis in AMOS identified support for group assignments as the strongest predictor of support for peer-assessment, followed closely by support for self-assessment. Support for a discussion-oriented classroom was also a modest but significant predictor. Both gender and level were not significant predictors when these factors were included (Table 4).

Construct	O (n=213).	G (n=130).	UG (n=83).	t- value	df	p value	Effect Size (Eta ²).
Peer	3.69 (.87)	3.79 (.91)	3.53 (.79)	2.149	211	*	.02
Self	3.74 (.81)	3.89 (.80)	3.50 (.80)	3.435	211	***	.05
Group	3.56 (.86)	3.72 (.85)	3.30 (.81)	3.567	211	***	.06
Discussion	4.75 (.68)	4.83 (.71)	4.64 (.62)	2.005	211	*	.02

* p<.05; *** p<.001. Standard deviations provided in parenthesis. O = Overall; G = Graduates; UG = First year Undergraduates. Peer – Support for Peer Assessment; Self – Support for Self Assessment; Group – Support for Group Assignments; Discussion - Support for a Discussion-oriented classroom.

Table 5: Comparison of construct means by level

Construct	W (n=166)	M (n=47)	t- value	df	p value	Effect Size (Eta ²)
Peer	3.64 (.84)	3.86 (.97)	-1.551	211	ns	
Self	3.72 (.79)	3.79 (.91)	-.514	211	ns	
Group	3.50 (.86)	3.75 (.83)	-1.718	211	ns	
Discussion	4.70 (.68)	4.95 (.65)	-2.276	211	.02	.02

* p<.05; ns = not significant. Standard deviations provided in parenthesis. W = Women; M = Men

Table 6: Comparison of construct means by gender

With respect to research question *What is the level of support among pre-service primary teachers regarding self-, peer-assessment and group assignments?* pre-service teachers' mean scores were 3.7 (sd 0.8), 3.7 (sd 0.9), and 3.6 (sd 0.9) respectively indicative of comparable but cautious student support for participative assessment strategies (Table 5). With respect to research question *Is support for these practices influenced by level and gender?* graduates compared to first year undergraduates were consistently more supportive of alternative assessment practices (Tables 5). For the undergraduates the comparison of means and associated inter-correlations suggests that level of support for peer-assessment is far more independent of their support for a discussion-oriented classroom than the graduates. Women and men were consistent in their cautious support of participative assessment practices and men were significantly more supportive of a discussion-oriented classroom compared to women (Table 6).

Discussion

The items used in this study aggregated to reliable constructs for measuring the level of student support for peer- and self-assessment among first year undergraduate and graduate pre-service teachers. The inter-relatedness of these four measures has also been established.

If these assessment practices are beneficial for learning, as the literature contends, then it follows that we would want education students to also use them in their school teaching. These assessment practices are on the rise in higher education, but the research is telling us that the more students are exposed to these practices their level of support tends to decline (Rushton, Ramsey & Rada, 1993). Hence blithely implementing them without consideration to and discussion of the opinions of students is not a recipe for good teaching practice. It is proposed that these scales be used by education lecturers to determine systematically base-line support among a student cohort prior to the implementation of a peer or self-assessment strategy and that the results be used to initiate a dialogue with students about their concerns.

Further scale development is planned to include additional items that are more focused on the perceived pedagogical benefits of both peer and self-assessment. The scale for measuring support for group assignments should be read with caution as it contains only three items. However, it does provide a basis for further scale development as does the validated scale for measuring support for a discussion-oriented classroom.

Self-report measures conducted about sensitive issues such as the present study suffer from responses biased toward social desirability (SD) (Hofstee, Berge, & Hendriks, 1998). This can interfere with the extraction of factors using an EFA. Hofstee and colleagues (1998) suggest that in at least some cases SD can account for up to 10% of the variance. The present study did not contain enough positively and negatively worded item pairs to calculate a SD deviance score; therefore no correction could be conducted. A follow up study including enough pairs of items to calculate a SD is suggested and would be a relatively straightforward addition to the current questionnaire items.

A further shortcoming with the present study was the lower than expected response rate (49%), also making interpretation of the data more problematic. Many students are spending less time on campus and more time in outside paid employment to help cover the cost of their studies and this is taking a toll on student engagement in all sorts of areas (McInnis, 2001) and must also impact the quality of relationships formed between students. Anecdotal comments from students in the graduate diploma course were consistent with McInnis' (2001) findings. The implication is that students are simply too busy to take the time to respond to voluntary questionnaires, conducted outside scheduled class time. This

has a direct bearing on the level of trust that can be built up between students and therefore the confidence they would subsequently have in each other's judgements for peer-assessment and potentially a significant threat to the validity of peer assessment as a pedagogical tool.

A further issue to consider when interpreting these results is the question of motivation. There appear to be differences in the motivation of people who leave careers in industry to teach compared to those who take a more direct path to teaching (Watt, Richardson, & Tysvaer, 2007). This may be a factor contributing to the differences obtained between the graduate and undergraduate support for participative assessment. It may be that older students who have experienced life in the workplace have developed a trust in and valuing of others' and their own views that may be lacking in the more inexperienced younger undergraduates. An equally plausible explanation is related to the earlier findings of Baxter-Magolda (1992) and Perry (1970) who noted a shift associated with the migration of knowledge authority from external to internal among undergraduates through the college years. The positive correlation between support for a discussion-oriented classroom and participative assessment may represent a belief in the wisdom of the group (Surowiecki, 2005).

In the context of an increasing use of peer-assessment strategies in higher education, these results are informative in having identified that a large cohort of Australian pre-service primary teachers indicated only modest support for the practice that was independent of gender. While level of support from students is not necessarily indicative of a lack of validity of scores generated through peer or self-assessment, a lack of perceived validity among students does not provide the best climate for increasing its use. Issues such as the 'illusion of unanimity', 'self censorship' and 'pressure for conformity' identified by Janis (1982) in other contexts are currently subtext in classrooms where peer-assessment is used. Open-discussion of these issues is likely to be productive and may serve to enhance the validity of peer-assessment. By collecting data using the scales provided may provide a fruitful beginning for initiating discussions with students.

Laughlin and Simpson (2004, p. 135) propose that we need to breakdown the mystique associated with assessment, stating that, "many students remain unsure of the assessment procedure and its aims [and that the process for students is largely about] guessing what the teacher wants rather than an educational experience". It is not surprising that Falchikov and Goldfinch (2000) advocate student involvement in setting criteria, and Hanrahan and Isaacs (2001) advocate considerable time is given to educating students about the process before peer assessment is conducted. The quelling of anxieties among students who are looking for 'expert' assistance through assessment may result.

All the participants in this study were planning a career as primary (elementary) school teachers and assessment will be a significant aspect of their

work. Therefore, in this study our aim was to better understand what pre-service teachers think about participative assessment practices. Further research identifying the subjective elements associated with peer- and self-assessment practices in classrooms, particularly those issues identified by Janis (1982) may show a change in perceptions of the costs and benefits of participative assessment. In our view students would be more likely to experience the benefits of peer assessment once they were satisfied that the potential for negative outcomes were adequately controlled for through discussion with the people who were to carry them out. Discussing the 'shadow' side of assessment practices may be as useful as student involvement in criteria setting. By doing this they can deepen their understanding of the complexity of the explicit and tacit, specific and general elements, and subjective, objective and intersubjective processes involved in producing valid assessments. If the aim of any assessment practice is to provide feedback that is accurate, helpful and appreciated by the receivers (Brown, Bull, & Pendlebury, 1997), there is much to be done to both understand student experiences of participative assessment and enhance the validity and effectiveness of these assessment practices that are increasingly used in higher education.

References

- Ballantyne, R., Hughes, K. & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research project, *Assessment and Evaluation in Higher Education*, 27(5), 427-441.
- Baxter Magolda, M. (1992). *Knowing and reasoning in college: Gender –related patterns in students' intellectual development* (San Francisco, Jossey-Bass).
- Brew, C. (2006). Peer-assessment: Economic convenience or pedagogical preference? *International Journal of Knowledge, Culture and Change*, 6(8), 145-158.
- Brew, C., Riley, P., Walta, C. (2009). Participative assessment practices: A comparison of pre-service primary teachers and teaching staff views, *Assessment and Evaluation in Higher Education*, 34(6), 641-657.
- Brindley, C. & Scoffield, S. (1998). Peer assessment in undergraduate programs, *Teaching in Higher Education*, 3(1), 79-89.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*, (London, Routledge).
- Byrnes, B. (2001). *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming* (New Jersey, Lawrence Erlbaum).
- Carter, L., Howlett, D., & Daruwalla, P. (2004). Evaluation of a peer assessment instrument on group work for marketing and management students: A

- multi-institutional study. ANZMAC 2004, Conference Proceedings, Victoria University, New Zealand.
- Cheng, W. & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise, *Studies in Higher Education*, 22(2), 233-240.
- Chin, P., & Overton, T. (2003). *Assessing group work LTNS physical services*. Available online at: <http://www.iml.uts.edu.au/Learnteach/assess/guides/Oll.htm> (accessed 7 June 2010).
- Conroy, D. E., Motl, R. W., & Hall, E. G. (2000). Progress toward construct validation of the Self-Presentation in Exercise Questionnaire (SPEQ), *Journal of Sport and Exercise Psychology*, 22, 21-38.
- Conway, R., Kember, D., Sivan, A. & Wu, M. (1993). Peer assessment of an individual's contribution to a group project, *Assessment and Evaluation in Higher Education*, 18(1), 45-56.
- Davies, P. (2000). Computerized peer assessment, *Innovations in Education and Training International*, 37(4), 346-355.
- Dochy, F., Segers, M., & Sluijsman, D. (1999). The use of self-, peer and co-assessment in higher education: A review, *Studies in Higher Education*, 24(3), 331-350.
- Dochy, F. J. R. C., & McDowell, L. (1997). Assessment as a tool for learning, *Studies in Educational Evaluation*, 23, 415-432.
- Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59(4), 395-430.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education. A meta-analysis comparing peer and teacher marks, *Review of Educational Research*, 70(3), 287-322.
- Goldfinch, J. (1994). Further developments in peer assessment of group projects, *Assessment and Evaluation in Higher Education*, 19(1), 29-35.
- Groundwater-Smith, S., Cusworth, R., & Dobbins, R. (1998). *Teaching, Challenges and Dilemmas* (NSW, Harcourt Brace & Company Australia Pty Ltd).
- Habeshaw, S. (1989). Teaching practicals, in: S. Habeshaw (Ed). *Preparing to teach* (London, Routledge Paul).
- Hanrahan, S. J. & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views, *Higher Education Research & Development*, 20(1), 53-68.
- Heron, J. (1981). Self and peer assessment, in: T Boydell & M Pedler (Eds.). *Management development, concepts and practices* (Farnborough, Gower).

- Holmes-Smith, P., Coote, L., & Cunningham, E. (2004). *Structural Equation Modelling: From the fundamentals to advanced topics* (Melbourne, School Research, Evaluation, and Measurement Services).
- Hofstee, W., Berge, J. T., & Hendriks, A. (1998). How to score questionnaires. *Personality and Individual Differences*, 25(5), 897-909.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit, in: R. H. Hoyle (Ed). *Structural equation modelling: Concepts, issues and applications* (Thousand Oaks, CA, Sage).
- Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The ringelmann effect: Studies of group size and group performance, *American Journal of Sociology*, 97, 371-384.
- Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (Boston, Houghton Mifflin).
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment* (New York, Free Press).
- Kane, J. S. & Lawler, E. E. (1978). Methods of peer assessment, *Psychological Bulletin*, 85(3), 555-586.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration, *Journal of Personality and Social Psychology*, 65, 681-706.
- Kline, R. B. (2005). *Principles and practice of structural equation modelling* (2nd ed). (New York, Guilford).
- Kwan, K. & Leung, R. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise, *Assessment & Evaluation in Higher Education*, 21(3), 205-214.
- Langan, A. M., Wheeler, C. P., Shaw, E. M., Haines, B. J., Cullen, W. R., Boyle, J. C., Penney, D., Oldekop, J. A., Ashcroft, C., Lockey, L. & Preziosi, R. (2005). Peer assessment of oral presentations: effect of student gender, university affiliation and participation in the development of assessment criteria, *Assessment and Evaluation in Higher Education*, 30(1), 21-35.
- Latane, B. (1981). The psychology of social impact, *American Psychologist*, 36, 343-356.
- Larisey, M. M. (1994). Student self assessment: A tool for learning. *Adult Learning*, July/Aug, 9-10.
- Laughlin, P. & Simpson, N. (2004). Peer assessment in first year university: How the students feel, *Studies in Educational Evaluation*, 30, 135-149.
- Lejk, M. & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments, *Assessment and Evaluation in Higher Education*, 26(1), 551-561.

- Li, L. K. Y. (2001). Some refinements on peer assessment of group projects, *Assessment and Evaluation in Higher Education*, 26(1), 551-561.
- McConnell, D. (2002). The experience of collaborative assessment in e-learning, *Studies in Continuing Education*, 24(1), 73-92.
- McInnis, C. (2001). *Signs of disengagement? The changing undergraduate experience in Australian universities. Inaugural professorial lecture*: Centre for the Study of Higher Education, University of Melbourne.
- Newstead, S. & Dennis, I. (1994). Examiners examined: The reliability of exam marking in psychology, *The Psychologist*, 5, 216-219.
- Nulty, D. D. (2010). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education*.
- Orsmond, P., Merry, S. (1996). The importance of marking criteria in the use of peer assessment, *Assessment and Evaluation in Higher Education*, 21(3), 239-250.
- Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS* (3rd ed.). (Crows Nest, Australia, Allen & Unwin).
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme* (New York, Holt, Rinehart & Winston).
- Pope, N. (2001). An examination of the use of peer rating for formative assessment in the context of the theory of consumption values, *Assessment & Evaluation in Higher Education*, 26(3), 235-246.
- Pope, N. (2005). The impact of stress in self- and peer assessment, *Assessment and Evaluation in Higher Education*, 30(1), 51-63.
- Reynolds, M., & Trehan, K. (2000). Assessment: A critical perspective, *Studies in Higher Education*, 25(3), 267-278.
- Rowtree, D. (1987). *Assessing students: How shall we know them?* (London, Kogan, Page).
- Rushton, C., Ramsey, P. & Rada, R. (1993). Peer assessment in a collaborative hypermedia environment: A case study, *Journal of Computer-Based Instruction*, 20(3), 75.
- Searby, M. & Ewers, I. (1997). An evaluation of the use of peer assessment in higher education: A case study in the School of Music, Kingston University, *Assessment and Evaluation in Higher Education*, 22(4), 371-383.
- Sharp, S. (2006). Deriving individual student marks from a tutor's assessment of group work, *Assessment & Evaluation in Higher Education*, 31(3), 329-343.
- Sluijsmans, D. M. A., Brand-Gruwel, S. van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions, *Assessment and Evaluation in Higher Education*, 27(5), 443-454.

- Stefani, L. A. J. (1994). Peer, self, and tutor assessment: Relative reliabilities, *Studies in Higher Education*, 19(1), 69-75.
- Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor Books.
- Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed). (New York, Harper Collins).
- Thompson, B. & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines, *Educational and Psychological Measurement*, 56, 197-208.
- Topping, K. J. (1998). Peer assessment between students in college and university, *Review of Educational Research*, 68(3), 249-276.
- Topping, K. J., Smith, E. F., Swanson, I. & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students, *Assessment and Evaluation in Higher Education*, 25(2), 146-169.
- Vandiver, B. J., & Worrell, F. C. (2002). The reliability and validity of scores on the almost perfect scale revised with academically talented middle school students. *The Journal of Secondary Gifted Education*, 13, 108-119.
- Watt, H. M. G., Richardson, P. W., & Tysvaer, N. M. (2007). Profiles of beginning teachers' professional engagement and career development aspirations (Ch 10, pp. 155-176). In A. Berry, A. Clemans, & A. Kostogriz (Eds.), *Dimensions of professional learning: Professionalism, practice and identity*. (Rotterdam, The Netherlands, Sense Publishers).
- Williams, E. (1992). Student attitudes towards approaches to learning and assessment, *Assessment & Evaluation in Higher Education*, 17(1), 45-58.
- Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in preservice mathematics education, *Journal of Mathematics Teacher Education*, 4, 95-113.