



University of HUDDERSFIELD

University of Huddersfield Repository

Pulman, Mark

Assessing personal attributes in the group rehearsal

Original Citation

Pulman, Mark (2010) Assessing personal attributes in the group rehearsal. *Music Education Research*, 12 (4). pp. 395-414. ISSN 1461-3808

This version is available at <http://eprints.hud.ac.uk/9216/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

Assessing personal attributes in the group rehearsal

Mark Pulman

*Department of Music and Drama, University of Huddersfield,
West Yorkshire, United Kingdom*

Dr Mark Pulman
University Campus Barnsley
University of Huddersfield
Church Street
Barnsley
United Kingdom
S70 2AN
m.pulman@hud.ac.uk
01226 644254

This is a study of the marks that were awarded for students' personal attributes, when used as peer assessment criteria, in their band rehearsals. Successive cohorts of first year **undergraduate** students, from 2001 until 2009, were involved in the research comprising of 191 students and eighty-four bands. Data analysis focused on the strength of marking agreement and the variances between self- and peer-assessments. Personal attribute assessments that exhibited the greatest strength of marking¹ agreement arose from when criteria were formulated together by bands, especially those attributes to which the group, as a whole, aspired; to a lesser extent, personal weakness criteria formulated by bands for each member. High flyers and female students under-estimated themselves in their self-assessments, compared with those awarded by the band, especially when using criteria arising from their personal weaknesses; weaker students overestimated themselves. In considering such misjudgements, this study raises questions about band members' self-efficacy belief.

Keywords: popular music rehearsals; peer assessment; personal attributes; self-efficacy; group work

Introduction

Group projects can provide an opportunity to develop students' personal attributes including teamwork, communication, commitment, creative input, versatility, self-responsibility and leadership; skills that are highly regarded by employers seeking graduates (Harvey *et al.*, 1997). Group work, however, can raise problems for

¹ In this paper, the term 'marking' can be regarded as being the equivalence of 'grading'.

assessment such as, for example, deciding how to award marks for each group member. Assessing and measuring individuals' contribution to the assignment, including converting the personal qualities that each have demonstrated during the group activity into a numeric rating, is a formidable challenge (Johnston and Miles, 2004).

Peer assessment, as a technique with which to assess students' work, is no longer regarded as a practice that is either pioneering or unusual (Falchikov, 2005). Indeed, alongside its potential for providing valuable learning opportunities in which students may develop evaluative and judgemental capabilities, peer assessment can also, importantly, develop a deeper understanding of assessment criteria. **If the rationale for employing peer assessment arises from the opportunities that it provides for supporting learning**, then it makes sense to identify the kinds of activities or experiences that would bring this about.

Prins *et al.* (2005) recognise the potential for peer assessment to 'thrive on interaction' and for providing a natural setting through which students can formulate and negotiate assessment criteria. Habeshaw *et al.* (1993) and Brown and Knight (1994) are among those who are in favour of criteria that are determined by the students themselves since it is they, of course, who will be using such criteria to make their judgements. Indeed, Boud & Falchikov (2006) emphasise the wasting of learning opportunities that are available by omitting to involve students in establishing criteria. For example, the omission of this activity in the classic study of Goldfinch and Raeside (1990) was acknowledged when, on reflection, 'a discussion with the students could have helped to make them feel more involved and helped them to notice the skills as they were displayed in the group' (p.210). Further, student

generated peer assessment criteria may contribute to a more open and transparent assessment culture.

The *process* of generating peer assessment criteria can have implications for the practicalities: how *many* assessment criteria should be employed and should criteria be *category-weighted* (i.e. assessment comprising a number of discrete criteria, each of which having a specific weighting within the whole) or are *holistic* criteria more appropriate? Whichever criteria are chosen, it is important that students agree with them and, as Gibbs (1999) suggests, written copies are disseminated to those involved as soon as possible.

Some reported studies raise concerns about the quality of student-generated criteria and question, for example, whether they have sufficient clarity and meaning (Mindham, 1998; Pond and ul Haq, 1998; Doran *et al.*, 2000). However, Orsmond, Merry and Reiling (2000) found that when students themselves developed the assessment criteria they acquired a kind of special understanding which

may have developed a sense of “ownership” which related to both the *meaning* of the criteria and the *worth* of the criteria in terms of marks to be awarded. Students may feel that because they constructed the criteria they are arbiters of the quality of the criteria expression and/or of the subject matter the criteria expresses (p.320).

Rust *et al.* (2003) addresses the problems associated with how to absorb students into the assessment culture of their disciplines and recognises that written criteria might not be sufficient in communicating meanings and intentions to novice students. Criteria tend to be articulated in a written format only and this medium may not always be sufficient in conveying subtle musical meanings of a non-discursive character. One problem of initiating students into assessment culture, Rust suggests, is the ‘tacit’ nature of assessment and the difficulty of transferring such understanding to others. If socialisation, as Rust argues, is required to assist in a transfer of ‘tacit’

knowledge (p.161) then involving band members in setting their own criteria for group rehearsing might seem an ideal context for this to take place.

Norton (2004) questions the appropriateness of explicit criteria because of the danger of rigidity or inflexibility, which could limit students to focus on the purely visible. Indeed by adopting explicit and detailed criteria intentionally this, as Norton argues, may have a 'deleterious effect' on the peer assessment process. She suggests that we should come to replace the term 'assessment criteria' with 'learning criteria' in order that the focus is not on the purely visible but includes wider demonstrations of learning that are more centred on the individual. Might there be peer assessment contexts where, because of the nature of group-rehearsing activities extending over a sequence of rehearsals, Norton's concept of assessment based on learning criteria is more appropriate? If learning criteria is tailored to the individual, then what are the applications and implications for individual members of a band who are working together in their rehearsals and performances? Criteria that are formulated for each individual student and particular group of students are a key theme also, in Pulman (2008) as well as for this paper. Finally, we are reminded by Hanrahan and Isaacs (2001), Smith *et al.* (2002), Ballantyne *et al.* (2002), Pope (2005) that not only is the setting of clear peer assessment criteria important, but equally so is training students in their use.

In the higher education subject area of music, peer assessment has also been reported in a number of activities including solo performance (Hunter and Russ, 1996; Searby and Ewers, 1996; Daniel, 2004; Blom and Poole, 2004), group performance (Pulman, 2004), orchestration (Hunter and Russ, 2000), composition (Searby and Ewers, 1996), music history group seminar presentations (Hunter and Russ, 2000) and recording (Lebler, 2008).

This present paper gives details of a peer assessment system employing criteria based on students' personal attributes arising from group rehearsals and analyses the marks that were awarded for these. It focuses on the marking data, over an eight-year period, arising from the *Performance Management* module involving first year undergraduate band rehearsals. **The tutor chiefly responsible for the instruction and assessment of this course is the author of this paper.** Prior approval for conducting research into this module, initially taught on the BA (Hons) Popular Music delivered at Barnsley College before subsequently transferring to The University of Huddersfield, was obtained from the relevant department at each institution.

Module outline and context

Providing opportunities for group work activities are particularly important for courses in popular **music**. Indeed, group work can be found in almost all areas of the popular music world; rehearsing, performing, recording, production, sound engineering and song-writing, for example, are collaborative activities that depend on professional musicians working together. For students of popular music there is, perhaps, much to be learned about how the appearance, or absence, of participants' interpersonal qualities, can make the difference between a productive rehearsal and a waste of everyone's time. The capacity to work effectively within a team is an attribute that is as prized in the music profession as it is in many other areas of human creativity.

If students' interpersonal skills and social qualities are an important part of the process that contributes to productive group rehearsing, then such attributes might, in a peer assessment system, be appropriated as rehearsal-related learning criteria (Norton, 2004). Employing such criteria as part of peer assessment could motivate

students towards bringing about an awareness of their personal attributes as they are displayed in the rehearsal. Identifying particular personal attributes that individuals would like to improve might, for example, provide a useful self-assessment activity; an activity in which students, in addition to rating themselves, could also involve comparing and reflecting on their self-ratings alongside the marks awarded by their peers.

It was decided to employ three kinds of personal attributes to be used as assessment criteria:

(1) attributes arising from what the band as a whole agreed were important to them; these ‘group-agreed attributes’ would reflect the intrinsic *in situ* and collective learning experience of their rehearsing;

(2) attributes specific to each individual within the band; these ‘individual attributes’ would, initially, be self chosen and arise from what students themselves considered they could improve upon in their rehearsing;

(3) ‘personal weakness’ attributes as above but this time being formulated by the band with whom they had previously rehearsed.

Using students’ personal attributes as assessment criteria was a technique that was gradually introduced and which developed out of the substantial experience of peer assessment had already been acquired. Up to three assignments were set each year: rehearsing for a ‘Christmas Party’ evening gig at a public venue assessed in December (abbreviated as CP); ‘Venues and Audiences’ (VA), rehearsing for a daytime gig usually in a school or café, assessed in March; and a Decade Tribute evening (DT), at a public venue, assessed in May. Bands were tutor-formed for the CP and VA gigs; bands for the DT gig were put together jointly by students and tutor.

A programme of training for peer assessment was provided for each first year cohort which included: an explanation of the rationale for peer assessment; activities involving the identification of personal attributes to be used as assessment criteria; safeguards for individuals, including transparency of the process, student agreement and, if required, tutor moderation of marks. The process through which students identified personal attributes, that might be appropriate to be used as peer assessment criteria for their rehearsing, commenced with a class discussion in which examples from previous years were introduced. A brainstorming activity usually then followed in which individuals were asked to identify or describe personal attributes, arising from their own experiences of rehearsing, that they considered should be displayed in their rehearsals. These were compiled on a white board in order to allow the class to discuss each attribute in terms of their possible meanings. Students, working individually, were then invited to consider five or six of these attributes (or any others that did not appear during brainstorming) that they regarded, arising from their own experience, as being the most important for them. The purpose of this was to help develop students' awareness of their personal attributes and to articulate those that they considered important for their rehearsing. Many of the personal attribute descriptors that students formulated, although exhibiting a range of various rehearsing qualities, tended to reoccur from year to year.

The three types of criteria, arising from the kinds of personal attributes previously described, consisted:

Group-agreed attributes

Bands were asked to devise three mutually agreeable *group attributes* that they believed were important to their rehearsing together, which could be used as peer

assessment criteria. It was emphasized that these were group-agreed attributes, arising from what they, as a band, considered important and that should be displayed in their rehearsals. Frequently formulated group-agreed attributes often used criteria descriptors comprising just one or two words, for example: *Communication; Team member; Commitment.*

Self-selected personal weaknesses

In order to develop a greater awareness of their *individual* involvement in the group rehearsals and in identifying areas for improvement, each student was asked to select three *attributes* that they considered as their ‘personal weaknesses’, which could be used as peer assessment criteria. Frequent self-selected personal weakness criteria included: *Confidence; Patience; Responsibility for own part.*

Band-determined personal weaknesses

Following students’ peer assessment experiences using self-selected personal weaknesses, a further training session included the question about who might speak with greater authority in identifying or noticing personal weaknesses: the individual him/herself, or the band with whom they have rehearsed? With two exceptions (2001-2002, before band-determined personal weaknesses were employed, and 2006-2007, where agreement was not forthcoming) students agreed to using personal weakness criteria that were determined by the band as a whole. It was important that care and sensitivity was taken in ensuring that individuals were assured that their personal weakness criteria referred only to the specific activity of rehearsing; In order to avoid situations in which individuals might interpret peers’ feedback as a personal slight (which was encountered on a few occasions early in the research) it became

imperative for band members to understand that such personal attribute criteria had no bearing on their personality or character whatsoever. Similarly, students were reminded of the purpose of the activity: learning about assessment criteria through identifying those appropriate for each individual and, importantly, being able to receive valuable feedback from the band about one's own personal attributes as displayed in the rehearsals. Bands considered each member in turn; individuals under discussion were offered the choice of either leaving at that point or remaining as a silent observer (most chose to do the former). It was important that written agreement was obtained for the band-determined personal weakness attributes that their band had formulated, including consent for these to be used as assessment criteria. Frequently chosen band determined personal weaknesses included: *Verbal input; Creative input; Tolerance of others' ideas.*

For the first peer-assessed assignment of the year, it was usual to devise six personal attribute criteria for each student, comprising three group-agreed and three self-selected personal weakness attributes. An additional purpose for inviting students to include group-agreed criteria characterised by a neutral potency, was to provide balance; employing personal weakness attributes alone might have resulted in an undue emphasis upon the negative and an unnecessary raising of students' anxieties about the process. The literature on peer assessment criteria (for example, Goldfinch and Raeside, 1990; Orsmond *et al.*, 2002) suggested that six was an appropriate number of criteria that would allow a range of qualities to be assessed while remaining manageable within the assessment process. It was decided that each personal attribute criterion would be weighted equally.

In order to help illuminate 'tacit' meanings symbolised within each personal attribute criterion, students were also asked to state what they considered to be the

opposite. This resulted in a set of bi-polar descriptors in which one of the poles represented the desired attribute, while the other pole represented the opposite; for example, *Leadership skills - Always having to be led*. These were marked using a five-point directional scale (*one to five*), which was felt to offer peers sufficient differentiation and discrimination in awarding marks for their band members' personal attributes; *one* was described as *poor/negative*, *five* described as *excellent/positive*, and *three* described as *average/neutral*, with *four* and *two* being *above* or *below*. Peer marking was usually conducted after the final rehearsal prior to the performance² and, from 2003 onwards (see later explanation), was normally conducted in secret.

As previously explained, this peer assessment system was founded on the formulation, by the students, of three types of personal attributes assessment criteria in order to develop a greater awareness of themselves and of their other band members. Tutor practice, involving using students' personal attributes as peer assessment criteria, also improved; for example, identifying further opportunities for supporting learning and acquiring a better understanding of the operational mechanics of peer assessment. In this respect, the increasing experience of the tutor may have had an indeterminate causal effect on the peer assessments that *have arisen* from year to year. Our system of peer assessment has, with a few changes and refinements, now entered its ninth year of operation.

² The final grade for each student was obtained from combining the peer assessments, arising from the rehearsals, with the tutor assessment of the band performance. This was calculated from the totalled peer assessment marks, dividing each by their mean, and multiplying them by the 'band mark'. It is a process sometimes described as the 'zero-sum' method (Sharp, 2006) because any student who is peer assessed as providing zero contribution receives zero marks. It can be expressed thus:

$$\frac{\text{Totalled peer assessment mark of individual (rehearsals)}}{\text{Mean of totalled peer assessment marks of band (rehearsals)}} \times \text{Performance mark awarded to the band as a whole}$$

Research process and analysis procedures

First year cohorts of the Performance Management module, between 2001 and 2008, were involved in the research, comprising 191 students in total. Many had performing experience at their previous colleges as well as in bands unconnected with their academic studies. Approximately one third were mature students, many of whom had substantial professional and semi-professional gigging experience. There was a gender imbalance, not untypical for popular music courses, with males [n=140] outnumbering females [n=51].

The aim of this research paper is to analyse and interpret the quantitative data arising from the marks **that were** awarded by the students across the **following five measures:**

- (1) the extent of agreement within a band when assessing each band member across the three types of personal attribute criteria;
- (2) the variance between individuals' self-assessments and those made by band members;
- (3) the marks awarded to group-agreed attributes, self-selected personal weaknesses and group-agreed personal weaknesses;
- (4) the marks awarded the highest- and lowest-rated individual in each band;
- (5) the marks awarded to female and male band members.

Although there is a large body of research devoted to the agreement between student and teacher ratings (e.g. Falchikov and Goldfinch, 2000), there appears to be a lack of research into the consistency of ratings among student groups themselves. As Zhang *et al.* (2008) explain, traditional reliability indices, such as the correlation coefficient, the percentage of agreement, and the Kappa statistic, cannot easily be

applied in this situation. As each band member receives ratings only from other members of the band, calculating traditional reliability indices across all of the bands is problematic. Further, it is not convenient for these statistical techniques to accommodate the multiple marks awarded to each member by band members.

In view of the relatively large number of popular music students involved in this study [n=191], bands [n=84] and **individuals'** personal attributes that were peer assessed [n= 1420], it was decided that employing standard deviations [SD], to measure the dispersal of marks within bands, would provide an indication of marking agreement that was appropriate and stable. **This metric was employed in the analysis of (1) above, and calculated from the peer assessments awarded to each band member.**

Normalised absolute difference coefficients [ND] were used as a metric to calculate the difference between an individual's self-assessment mark and the mean of the band **and were employed in the analysis of (2).** The difference is simply rescaled [normalised] within the range of 0.00 [i.e. complete match] and 1.00, [i.e. gross disparity]. **The methods used to determine the remaining measures were calculated from obtaining the means of the marks awarded for group-agreed attributes, self-selected personal weaknesses and group-agreed personal weaknesses (3), highest- and lowest-rated individual in each band (4), and female and male band members (5).**

Results and Discussion

An extensive data corpus accumulated during the eight-year period. That said, the range of standard deviations (indicating strength of agreement) and normalised difference coefficients (the difference between an individual's self-assessment mark and the mean of the band) obtained from year to year (see Table 1) was not as wide as

might have been expected: standard deviations for group-agreed attributes were in the range of 0.31 to 0.68 only; for self-selected personal weaknesses they were 0.52 to 0.63; and 0.35 to 0.69 for band-determined personal weaknesses. Similarly, the ranges obtained for normalised difference coefficients across each attribute type were also relatively narrow: between 0.14 and 0.22 for group-agreed attributes; 0.10 to 0.22 for band-determined personal weaknesses and between 0.17 to 0.22 only, for self-selected personal weaknesses. Indeed, for the latter, as Table 1 indicates, five of the six years resulted in a ND that was confined to between just 0.17 and 0.18.

[Insert Table 1]

Overall, band members were in strongest agreement when marking group-agreed attributes (SD 0.47, n=651). The strength of agreement, although remaining impressive, was not quite as pronounced in the bands' marking of band-determined personal weaknesses (SD=0.55, n=391) and self-selected personal weaknesses (SD 0.56, n=378).

In order to compare the marks that were awarded by the bands for each attribute type the means of these, totalled by assignment, were calculated and expressed as a percentage. Table 2 reveals that students' group-agreed attributes were, on average, marked consistently higher [83%, n=665] than those awarded for band-determined personal weaknesses [80%, n=638] and, higher still, than self-selected personal weaknesses [77%, n=539]. This, of course, was not surprising; it would be expected that attributes, considered either by the band or selected by oneself as 'personal

weaknesses', would be marked lower than those suggestive of having a neutral potency (i.e. being neither strong nor weak), such as group-agreed attributes.

[Insert Table 2]

The cohort that was clearly the most generous of all in their marking, recording the highest mean marks for each of the three attribute types in the eight years, was that of 2007/8. Identifying the least generous proved to be more problematic, given the differing nature of each attribute type; 2003/4 and 2004/5 would, however, be among the contenders.

Group-agreed attributes

Student agreement was at its strongest when assessing group-agreed attributes (SD 0.47, n=651); indeed the SD for 2001/2, of 0.31, was remarkably small in part through using a smaller marking scale of 1-4 only (rather than 1-5) for that year. Further, and as a consequence of certain band members awarding identical marks amongst themselves, it was also suspected that there were four marking cartels in operation during that year. Band members were also invited to consider marking collaboratively rather than in secret during 2001-2003; an invitation which might, unintentionally, have provided a *raison d'être* for, and perhaps tacit approval of, bands to award their members identical marks across each personal attribute criteria. Allowing bands the opportunity to decide through group discussion how they should mark each other's personal attributes, it was hoped, would be a valuable learning experience. In practice, however, these were not easy activities to facilitate; on occasions band members simply agreed to award marks uniformly (usually the top mark) to each other.

Allocating marks on this particular basis of mutual self-interest, although not surprising, unfortunately rendered each student's personal attribute criteria, upon which so much time was devoted to their formulation, irrelevant. Offering bands the opportunity to mark collaboratively, as an approved activity, ceased after 2003; discontinuation would not, in itself however, prevent the occurrence of further suspected pre-meditated agreements or marking cartels.

Irrespective of the occasional peer-marking manipulation, the strength of agreement among band members remains impressive. A high correlation exists also between individuals' self-assessments and those of their band (ND 0.18, n=358). This suggests that individuals' understanding of their achievements, as measured by group-agreed attribute criteria marks, differed relatively little from their bands' perceptions. What might individuals have learned from this? Many, on receiving their marking feedback would at least have been reassured, perhaps, that perceptions of themselves were largely confirmed by the assessment of their band.

The consistency with which group-agreed attributes attracted the highest marks overall, was clearly apparent from the analysis of individuals' self-assessments; the mean mark for the eight years being an astonishing 83% [n=481]. Although this self-assessment mean is identical with the bands' overall mean, it conceals large variances between self- and band -assessments during particular years, for example in 2007 and 2008.

Self-selected personal weaknesses

As previously noted, the mean overall mark awarded for self-selected personal weaknesses was lower than either group-agreed attributes or band-determined personal weaknesses. Band members remained in strong agreement with each other however, when assessing individuals' self-selected personal weaknesses (SD 0.56).

The difference between individuals' self-assessments of their personal weaknesses and those awarded by their band members was also relatively small (ND 0.18, n=294). Similarly, as observed from Figure 1, there was only a slight difference between the totalled eight-year means for self- and band- assessments for self-selected personal weaknesses (75% and 77% respectively).

[Insert Figure 1]

In summary, the assessment of individuals' self-selected personal weaknesses exhibited stability and consistency throughout the eight years, with relatively little overall difference between the marks that individuals awarded to themselves and those awarded to them by their bands.

Band-determined personal weaknesses

Band members strongly agreed with each another when assessing band-determined personal weaknesses (SD 0.55, n=391). Interestingly, the difference between individuals' self-assessments of their band-determined weaknesses and those awarded by their band members (ND 0.17, n=316) was slightly smaller than for either group-agreed attributes or self-selected personal weaknesses. Indeed, a ND of only 0.10 for band-determined personal weaknesses, by far the smallest occurring anywhere in the data, was recorded for the peer assessments in 2007-08. An examination of the means, from year to year, of students' self assessments of band-determined personal weaknesses (Figure 1) reveals, however, that individuals tended to under-estimate themselves when compared with the marks awarded by their band members. As in self-selected weaknesses, the marks awarded for individuals' band-determined personal weaknesses were also stable and consistent; if anything, individuals and

bands were slightly more ‘in-tune’ with each other about their band-determined personal weaknesses.

Band members with the lowest marks

Analysis was also conducted on individuals who received the lowest mark in their band. Bands containing two or more members who, jointly, received the lowest marks were excluded from the data analysis. An examination of Figure 2, which compares the means for the lowest rated students’ self-assessments with those awarded by the bands for each assignment, reveal these students (n=47) consistently, and sometimes grossly, over-estimated themselves in the marks they awarded for their personal attributes in the rehearsals. Their self-assessments exceeded those awarded by their bands by 15%, 12% and 8%, for group-agreed attributes, self-selected and band-determined personal weaknesses, respectively; this suggests that these weaker students had a somewhat distorted view of their achievement in the rehearsals.

[Insert Figure 2]

Although individuals were each assessing themselves on the basis of their personal attribute criteria rather than on contribution criterion, it may be that some of the weaker students deliberately inflated their grades, either in the mistaken belief that their self assessments would contribute to their final mark, or perhaps to moderate the overall profile of marks, as a protest about the poor marks that they anticipated would be awarded by their fellow band members. Perhaps the source of some of these variances arose from individuals’ differing perceptions of their personal weakness attributes as a consequence of the developing rehearsing contexts and experiences.

Some, for example, might have marked their self-assessments according to the meanings of the personal attribute criteria they believed they had originally agreed upon prior to rehearsing; bands, however, assessed these attributes in the retrospective light of their rehearsal experiences, in which *a priori* comprehension of the assessment criteria had changed.

Bands might have punished a weak student's poor contribution also, by interpreting their personal attribute criteria as being equivalent to the mark awarded for their contribution, as there were instances of low marks being uniformly awarded across each of the personal attributes of these weaker students. Such speculation arises from the author's observations of the students, tutorial involvement in their rehearsing and semi-structured interviews conducted with a number of individual students between 2002-2006 (Pulman, 2008). In order to examine these kinds of issues further, narrative self-reflections, open-ended anonymous surveys, or further semi-structured interviews would seem warranted.

Nonetheless, band members appeared to be in rather less agreement with each other when assessing weak students. Did they find assessing weaker group members problematic because some were not as willing to give low marks for their colleagues' personal attributes in comparison with their readiness to reward the best group member? Band members differed most, although still largely in overall agreement, when assessing self-selected personal weaknesses [SD 0.76], with group-agreed attributes [SD 0.70] and band-determined personal weaknesses [SD 0.69] exhibiting members' assessments that were somewhat more in harmony with each other. The greatest spread of marks, overall, occurred in the assessments of those who would become the lowest-rated individuals in each band.

Weak students and free riders

There is a distinction that may be made, between those who, in spite of their attempts at contributing to rehearsals, remain weak, and others, who by demonstrating their lack of involvement, effort or commitment, are deserving of being considered 'free riders'. Lateness and absence from timetabled rehearsals, accompanied by unconvincing excuses, complaints from the group about a particular member's attendance, attitude and lack of contribution have a bearing on a tutor's understanding of the situation and how best to identify the support that may be required for the group and free rider alike. On this basis, and from the total of 191 students involved in this study, there were 15 'free-riders' belonging to the various bands across the years. Table 3 identifies and tracks the journey of these free riders during the module year [their identities have been anonymised]. Of the four individuals who were considered free riders in the CP rehearsing assignment of 2002, for example, there were three remaining in the DT rehearsals. Similarly, the three students [JT, BQ and GJ] appearing in the Christmas Party 2001, by the time of their assessments resulting from the V&A rehearsals, in which they demonstrated their commitment, coupled with improvements to their personal attribute marks had, in the opinion of the author, exonerated themselves. It was unfortunate, however, that three other individuals (DN, DD, SM) emerged as fresh free riders during the Venues & Audiences rehearsals. However, of the fifteen students who were initially regarded as being free riders, this number, happily, was reduced to nine at the conclusion of their final assignment.

[Insert Table 3]

Band members with the highest marks

Similar analysis was performed on the highest rated individual [n=47] of each band; bands having two or more members who had jointly received the highest marks were excluded from the calculations. Figure 3, which compares the means of the highest rated students' self-assessments with those awarded by the bands across all assignments, reveals that individuals consistently under-estimated themselves in the marks they awarded for their personal attributes. Variance, in terms of their under-estimation for group-agreed attributes was 5%, increasing to 13% and 9% for self-selected and band-determined personal weaknesses respectively. These modest self-ratings may be attributed to a number of possible explanations: an intentional or unintentional setting of high rehearsing standards for themselves arising, perhaps in part, from their previous experiences; the natural anxiety, occurring in year one students, associated with joining a band consisting of peers with whom they were unfamiliar; errors of judgement in self-efficacy belief (Bandura, 1977,1986), especially involving personal *weakness* attributes (either self-selected or band-determined) intended to provide opportunities for individuals to demonstrate that s/he possessed the capability of developing or improving them. Implications for self-efficacy belief arising from this study are discussed later in this paper.

[Insert Figure 3]

Students were in considerable agreement, as measured by mean standard deviations, when assessing would-be higher-rated students, compared with those who were destined to become the lowest rated individual of the band. Indeed, as Table 1 shows, this relationship was repeated in no less than seventeen out of the eighteen

occasions when personal attributes were used as assessment criteria. Of the three types of criteria that were employed, the strongest agreement occurred when assessing group-agreed attributes [SD 0.37] followed closely by self-selected personal weaknesses [SD 0.42] and band-determined personal weaknesses [SD 0.49].

Shooting stars and shining stars

Although the analysis so far has included all of the highest marked members of each band, a distinction may be made between those who achieved that lofty position on one occasion only, and those who impressed further by repeating this achievement in other peer assessed rehearsals. The former can be likened to shooting stars, because they shine brightly in one assignment only and then fade; the latter, shining brightly from one assignment to another, shining stars. From the total number of students involved over the eight years [n=191], shining stars, being the highest rated student across multiple assignments, numbered just nine. From the author's experience in facilitating the activities whereby bands formulated the 'personal weaknesses' criteria of their individual members, shining stars were very rarely subjects of animosity, arising out of jealousy. It was, rather, the opposite: shining stars were regarded as natural bandleaders, indicated also by the frequency with which bands selected personal attributes such as *leadership skills, confidence and organisational abilities* to which their members might aspire.

Gender

Similar analysis was conducted for all female [n=51] and male [n=140] students and a summary of this appears in Figure 4. Female students consistently under-estimated

themselves in the self-assessments of their personal attributes compared with those awarded by their bands. The average under-estimation of their group-agreed attributes [n=120] was 2%, increasing further to 5% and 6% for self-selected [n=83] and band-determined [n=74] personal weaknesses respectively. Female students underestimated themselves most of all, therefore, when self-assessing their personal weaknesses. A similar tendency towards under-estimation when self-assessing their personal attributes, although less pronounced, was also found among male students. The average under-estimation for male students when self-assessing their group-agreed attributes [n=361], self-selected personal weaknesses [n=164] and band-determined personal weaknesses [n=182] was 2%, 1% and 3% respectively. In their self-assessments therefore, male students were in close proximity, overall, to the assessments made by their bands. In terms of the overall mean marks for rehearsing awarded by bands, it is interesting to note that females outperformed males across both group-agreed attributes and self-assessed personal weaknesses; for band-determined personal weaknesses, however, the marks were the identical. These data suggest that female students were somewhat more successful than male students, as measured by the marks awarded for their personal attributes in rehearsing by their bands.

[Insert Figure 4]

Discussion

This is a discussion about the findings of an eight-year study arising from using peer assessment based on students' group-agreed, self-selected personal weaknesses and band-determined personal weaknesses criteria for rehearsing.

Marking personal attributes

Students agreed with each other most of all when assessing their band members' group-agreed attributes. One possible explanation for this relates both to the nature of, and causal effects resulting from, formulating group-agreed rehearsal criteria. These criteria typically reflect group working and general team skills; qualities in which it would be highly improbable for members not to have displayed on several occasions. Further, the situational setting within which a band, as an entity, decided their mutually agreeable attributes might have been significant. If so, then perhaps strongly held group agreements, arising from a face-to-face context about what they considered important for them, were demonstrated at the point of assessment by strong agreements in marking. The assessments of individuals' self-selected or band-determined personal weaknesses also exhibited a degree of stability and consistency; personal weaknesses that were formulated by the band, rather than being self-selected, achieving slightly stronger levels of marking agreements. Band members were not always in harmony with each other when assessing weaker students, however, as revealed by the wider spread of marks that appeared on occasions.

Peer assessment can also raise questions about the impartiality and honesty of those involved in awarding marks. Examinations of the variance and ND data were useful indicators: if variances had been large or the intra-rating agreements (ND) among the bands had been weak, then these findings would cast doubt on the honesty of the marks that were awarded. The variance and ND data suggests that this was not the case.

Formulating personal attribute criteria, especially from a symbolic interactionist perspective, might be considered as a means whereby the *self* or *identity* is presented

and constructed. Self-selecting personal attributes in the context of band rehearsing, presents individuals with an activity that involves the construction of meanings about themselves, that are symbolised through peer assessment criteria. This process, especially for self selected personal weakness criteria, involves and indeed encourages individuals to become more aware of their own rehearsal image. Deciding on a descriptor to represent this (for example, *versatility*) is an activity that might, for example, originate in class discussion or from the complex socio-linguistic meanings that may be encountered and comprehended by the individual; a highly cognitive activity in each case. Indeed, of particular interest for this study were the thinking processes that might have occurred between the activity of formulating a personal attribute criterion and, through an individual's self-awareness, deciding on a descriptor that best represented those meanings symbolised by that human quality.

In speculating on such meanings, could there have been a causal effect that might have accounted for the variance of the marks awarded (between an individual and their band) for personal weakness attributes, as a consequence of changing rehearsing contexts? If so, then some band members, perhaps, might have self-assessed their personal attribute criteria on a basis of what they believed was agreed prior to rehearsing, while the rest of their band were assessing those qualities according to meanings that were being shaped by the experiences of rehearsing with him/her.

Free riders and weak students

Most students appeared to be pulling their weight; one factor that might, perhaps, have caused the low proportion of free riders was the knowledge that each would be peer assessed on their rehearsing. There was some evidence of premeditated marking agreements among bands, however, arising from the awarding of identical marks

(usually 5 out of 5) to each member; free riders were usually excluded from, and also punished by, the assessments of those marking cartels. Most noticeable of weak students and free riders was the regularity with which they over-estimated themselves in their self-assessments. Especially for those students, but also for everyone involved, there were valuable opportunities to learn about themselves through reflecting on the over-estimation of their rehearsing attributes. Individuals' might learn much about themselves not only from the marks awarded by their bands for their personal weakness attributes but also from a comparison of these with their self-assessments. Reflecting on the variance between the self-assessments and the band marks might help individuals to develop a greater awareness and self-knowledge of, for example, their personal weaknesses in the rehearsal. This could also be facilitated through written comments directed towards focused improvements and goal setting. Such reflection has implications also for self-efficacy appraisal. The under- and over-estimations exhibited by shooting stars and free riders respectively, of their self-assessments compared with those of their bands, suggests misjudgements about self-efficacy.

Self-efficacy

A guiding principle of self-efficacy belief, Bandura (1977, 1986) suggests, is that unless people believe that their actions can lead to improvements, they have little incentive to act on it. In this study, self-selected personal attributes such as personal weaknesses may be illustrative of an individual's self-efficacy belief. For example, self-selecting *confidence* as a *personal weakness* attribute would demonstrate an individual's belief that s/he possesses the capability of developing or improving it. Conversely, that same individual might not self-select certain other personal

weaknesses that they consider to possess, because of doubting their efficacy to mount a similar effort to improve those attributes. Errors of judgement in self-efficacy, arising from the under- and over-estimations of self assessments exhibited by shooting/shining stars and weak students/free riders respectively might for example, as Bandura advises, require to be checked periodically in order to ascertain the effect of further rehearsal experiences. This suggests that where a large variance occurs between, for example, a self assessed and band assessed personal weakness attribute, a reselection of that attribute in a subsequent peer assessed rehearsal assignment is necessary in order to improve the self-efficacy judgement of the individual. In fact, such reselections regularly occurred between assignments when, for example, an individual's self-selected personal weakness attribute that s/he formulated in one assignment also appeared in the next as band-determined personal weakness. Bandura also identifies a number of sources from which individuals might form their self-efficacy beliefs that may have relevance for band rehearsing. They are described as: mastery experience (meaning their previous rehearsing); modelling (observation of others during rehearsing); social persuasions (feedback and judgements received from other band members); and somatic and emotional states (personal anxiety, mood or ego, for example, during rehearsing). Such self-efficacy principles may also illuminate situational factors (working in band rehearsals) and cognitive processes (reflection on self- and peer-assessed personal attributes) whereby individuals develop self-awareness and motivation, through their peer marking, to improve their *personal weaknesses* in this particular context.

Conclusion

In the introduction to this article, it was suggested that providing activities and experiences that support learning should be at the centre of a peer assessment system.

Encouraging individuals to learn more about themselves and others in the rehearsal, through using personal attributes as peer assessment criteria, forms a key pedagogical underpinning of this performance module. Establishing individuals' own personal attribute criteria, self assessing these, formulating band-determined personal weaknesses, marking those of other band members, receiving marking feedback from the band, reflecting on these marks and comparing these with their self assessments are activities which, in the opinion of the author, can provide valuable learning opportunities for every student. In this study, the personal attribute assessments that exhibited the greatest strength of marking agreement arose from when criteria were formulated together by bands, especially those attributes to which the group, as a whole, aspired; to a lesser extent, personal weakness criteria formulated by bands for each member. High flyers, characterised as either shooting or shining stars, together with female students under-estimated themselves in their self-assessments, compared with those awarded by the band, especially when using criteria arising from their personal weaknesses; weaker students and free riders overestimated themselves. Although no one single finding arising from the analysis might have a greater weight than any other, these data, in which band members marked the attributes of their peers in rehearsal, provide an interesting insight into the rehearsing life-worlds of these students.

Of course there is so much that cannot be explained by an analysis of peer assessment marks alone: as the mathematician Alfred Einstein was reputed to have remarked, not everything that can be counted counts, and not everything that counts can be counted³.

³ McKee, M. 2004 Editorial Commentary *British Medical Journal* 328, No.7432: 153.

Table 1. Standard deviations (SD) and Normalised difference coefficients (ND) for group-agreed attributes, self-selected and band-determined personal weaknesses for each cohort year.

Cohort year	Bands & members	Group-agreed attributes (n=)		Self-selected personal weaknesses (n=)		Band-determined personal weaknesses (n=)	
		SD	ND	SD	ND	SD	ND
2001/2	18/85	0.31(212)	0.14 (141)	No assessment (n/a)		No assessment	
2002/3	14/67	0.58 (176)	0.17 (153)	0.52(51)	0.17 (81)	No assessment	
2003/4	13/68	0.53(75)	N/a	0.60(71)	N/a	0.69(129)	0.18 (87)
2004/5	8/37	No assessment		0.55(37)	0.17 (39)	0.64(72)	0.17 (63)
2005/6	9/39	0.68(53)	0.23 (39)	0.60(42)	0.18 (28)	0.47 (78)	0.22 (60)
2006/7	5/31	No assessment		0.63(57)	0.17 (38)	No assessment	
2007/8	9/49	0.43(75)	0.17 (72)	0.51(63)	0.17 (57)	0.35(66)	0.10 (66)
2008/9	8//38	0.56(60)	0.22 (54)	0.52(57)	0.22 (51)	0.49(46)	0.17 (40)
Year mean		0.47(651)	0.18(359)	0.56(378)	0.18(294)	0.55(391)	0.17(316)

Table 2. Group-agreed attributes, self-assessed personal weaknesses and band-determined personal weaknesses: mean marks awarded by bands, expressed as a percentage

Group-agreed attributes			Self-selected personal weaknesses			Band-determined personal weaknesses		
Cohort Year	Rehearsal	Mean mark	Cohort Year	Rehearsal	Mean mark	Cohort Year	Rehearsal	Mean mark
2001/2	CP	85%	2002/3	CP	72%	2003/4	VA	77%
	VA	84%		CP	74%		DT	73%
2002/2	CP	81%	2004/5	CP	74%	2004/5	VA	71%
	DT	82%	2005/6	CP	78%		DT	77%
2003/4	CP	80%	2006/7	CP	72%	2005/6	VA	75%
2005/6	CP	79%	2007/8	CP	87%		DT	83%
2007/8	CP	89%	2008/9	CP	84%	2007/8	DT	92%
2008/9	CP	85%	Year mean		77%	2008/9	DT	90%
Year mean		83%				Year mean		80%

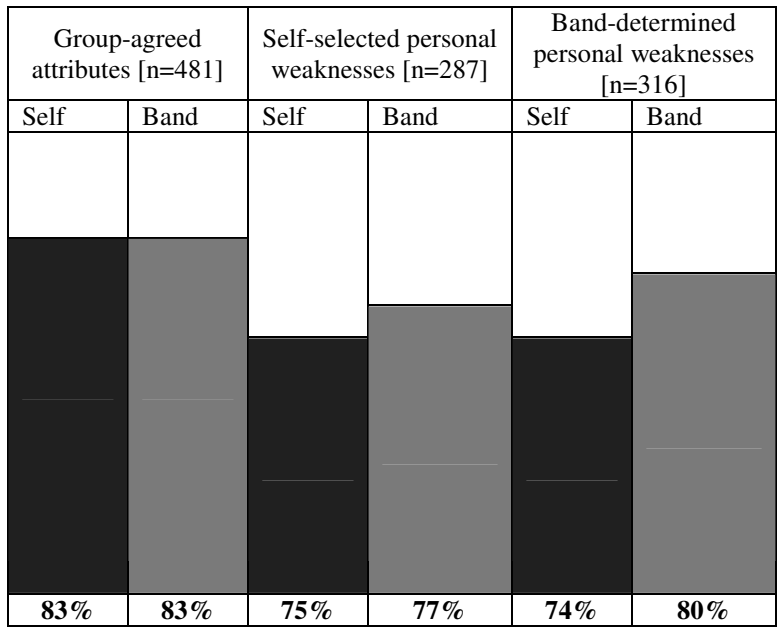


Figure 1. Comparison of self- and band assessments: mean marks, expressed as a percentage

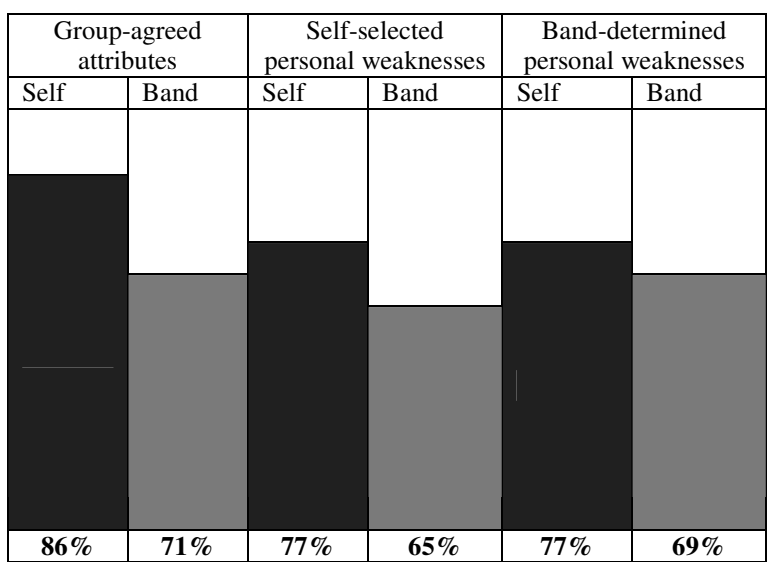


Figure 2. Self-assessments: mean marks for the lowest rated students [n=47], expressed as a percentage, compared with band assessments.

Table 3. Overview of free riders

Year	Rehearsal assignments			Free riders	
	CP	VA	DT	Total	Remaining
2001/2	JT/BQ/GJ	Peer assessed as 'weak' students rather than 'free riders' (due to improved attendance, attitude & contribution, etc)	No assessment	3	0
		DN/DD/SM free riders	N/a	3	3
2002/3	TG/SD/BQ/DG	No assessment	SD/BQ/TG	4	3
2003/4	LK	Peer assessed as 'weak'; as above	N/a	1	0
		SM	Peer assessed as 'weak'	1	0
2004/5	NE	NE	N/a	1	1
2005/6	DP/RS	DP	DP	2	1
2006/7	HK	No assessment	N/a	1	1
2007/8	PS/RM	No assessment	PS	2	1
2008/9	KP	No assessment	N/a	1	0
Total				15	9
Comparison of mean SD and ND totals		Group-agreed attributes	Self-selected personal weaknesses	Band-determined personal weaknesses	
Free riders mean SD		0.67 [n=57]	0.70 [n=70]	0.74 [n=74]	
Total student mean SD		0.47 [n=651]	0.57 [n=378]	0.55 [n=391]	
Free riders mean ND		0.69 [n=42]	0.79 [n=79]	0.74 [n=74]	
Total student mean ND		0.18 [n=359]	0.18 [n=294]	0.17 [n=316]	

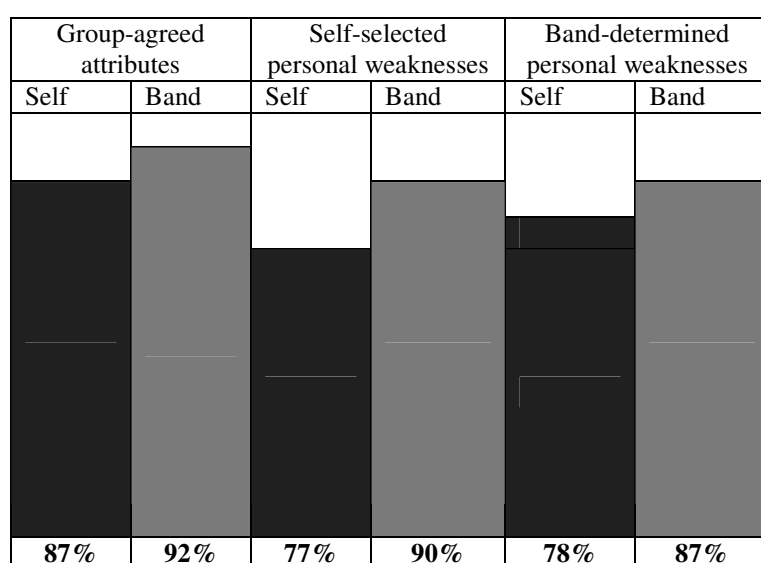


Figure 3. Self-assessments: mean marks for the highest rated students [n=47], expressed as a percentage.

Group-agreed attributes				Self-selected personal weaknesses				Band-determined personal weaknesses			
Self		Band		Self		Band		Self		Band	
♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂
86%	81%	88%	83%	74%	76%	79%	77%	73%	76%	79%	79%

Figure 4. Comparison of female and male self- and band-assessments: mean marks, expressed as a percentage.

Notes on contributor

Dr Mark Pulman is a senior lecturer in popular music at the University of Huddersfield. He pursued careers both as a freelance professional musician and in music promotion where he has substantial experience, working for many years in publishing and record companies, including Warner/Chappell Music, Music Sales Ltd and Novello. Prior to this, he was head of music at two secondary schools. He obtained a master’s degree in music education from the University of Nottingham and his PhD was in the field of peer learning in popular music.

References

Ballantyne, R., K. Hughes, and A. Mylonas. 2002. Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education* 27, no. 4: 427-441.

Bandura, A. 1977. *Social Learning Theory*. New York: General Learning Press.

Bandura, A. 1986. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Blom, D., and K. Poole. 2004. Peer assessment of tertiary music performance: opportunities for understanding performance assessment and performing through experience and self-reflection. *British Journal of Music Education* 21, no. 1: 111-

- Boud, D., and N. Falchikov. 2006. Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education* 31, no. 4: 399-413.
- Brown, S., and P. Knight, 1994. *Assessing Learners in Higher Education*. London: Kogan Page.
- Daniel, R. 2004. Peer assessment in musical performance: the development, trial and evaluation of a methodology for the Australian tertiary environment. *British Journal of Music Education* 21, no. 1: 89-110.
- Doran, S., C. Durston, A. Fletcher, and J. Longmore. 2000. Assessing students in seminars: evaluation of current practice. In *The Practice of University History Teaching*, ed. Booth, A., and P. Hyland. Manchester: Manchester University Press.
- Falchikov, N. 2005. *Improving assessment through student involvement: practical solutions for aiding learning in higher and further education*. London: Routledge.
- Falchikov, N., and J. Goldfinch. 2000. Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational ...*
- Gibbs, G. 1999. Using Assessment Strategically to change the way students learn. In Brown, S. and A. Glasner. *Assessment Matters in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.
- Goldfinch, J. and R. Raeside. 1990. Development of a Peer Assessment Technique for Obtaining Individual Marks on a Group Project. *Assessment & Evaluation in Higher Education* 15, no. 3: 210-231.
- Habeshaw, S., G. Gibbs, and T. Habershaw. 1993. *Interesting Ways to Assess Your Students*. Melksham: Cromwell Press.
- Hanrahan, S., and G. Isaacs. 2001. Assessing Self and Peer-assessment: the Students Views. *Higher Education Research and Development* 20, no. 1: 53-70.
- Harvey, L., S. Moon, V. Geall, and R. Bower. 1997. Graduates' Work: Organisational Change and Students Attributes. University of Central England UPDATE Newsletter (April 1997).
- Hunter, D. and M. Russ. 1996. Peer assessment in performance studies. *British Journal of Music Education* 13, no. 1: 67-78.
- Hunter, D. and M. Russ. 2000. *Peer Learning in Music*. Coleraine: University of Ulster.
- Johnston, L. and L. Miles. 2004. Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education* 29, no. 6: 751-768.
- Lebler, D. 2008. Popular music pedagogy: peer-learning in practice. *Music Education Research* 10, no. 2: 193 – 213.
- Mindham, C. 1998. Peer assessment: Report of a project involving group presentations and assessment by peers. In *Peer Assessment in Practice*, ed. S. Brown. Birmingham: SEDA Publications.
- Norton, L. 2004. Using assessment criteria as learning criteria. *Assessment & Evaluation in Higher Education* 29, no. 6: 687-702.
- Orsmond, P., S. Merry, and K. Reiling. 2000. The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education* 25, no. 1: 23-38.
- Orsmond, P., S. Merry, and K. Reiling. 2002. The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education* 27, no. 4: 309-323.
- Pond, K., and R. Ul-Haq. 1998. Learning to access students using peer review. *Studies in Educational Research* 23, no. 4: 331-348.
- Pope, N. 2005. The impact of stress in self- and peer-assessment. *Assessment &*

- Evaluation in Higher Education* 30, no. 1: 51-63.
- Prins, F., D. Sluijsmans, P. Kirschner, and J. Strijbos. 2005. Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education* 30, no.4: 417-444.
- Pulman, M. 2004. Peer Assessment in Popular Music. In *How Am I Doing? Valuing and Rewarding Learning in Musical Performance in Higher Education* ed. D. Hunter. Coleraine: University of Ulster.
- Pulman, M. 2008. Knowing Yourself Through Others: Peer Assessment in Popular Music. Unpublished PhD Thesis. Sheffield Hallam University.
- Rust, C., M. Price, and B. O'Donovan. 2003. Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education* 28, no. 2: 147-164.
- Searby, M., and T. Ewers. 1996. Peer assessing composition in higher education. *British Journal of Music Education* 13, no. 2: 155-163.
- Smith, H., A. Cooper, and L. Lancaster. 2002. Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development. *Innovations in Education and Teaching International* 39, no.1: 71-87.
- Zhang, B., L. Johnston, and G. Kilic. 2008. Assessing the reliability of self- and peer rating in student group work. *Assessment & Evaluation in Higher Education* 33, no. 3: 329-340.