

First appearing in *Educational Developments* 4.1 2003

Achieving more by doing less: Integrating peer assessment and collaborative written work

James Hartley and Stephen Bostock
University of Keele

One way of improving the quality of student learning is to encourage students to work in small groups or pairs. Not only do students writing in groups or pairs halve marking loads at a stroke but research has also shown that working in groups has educational advantages. The literature in this field suggests that collaborative writing amongst academics can:

- be more efficient – because different aspects of the task can be shared out;
- be of better quality – because different individuals can contribute different expertise; and
- lead to better-written papers – because each individual contributor can assist in the editing of the paper, seeing it from different perspectives.

Similarly, research with students has also reported positive findings for paired and collaborative writing. Quantitative studies suggest that writing in pairs or groups is beneficial for the students in that:

- pairs or groups do as well or better than individuals on various tests of quality (Louth et al, 1993; O'Donnell et al, 1985; Tang, 1998), and
- many students express positive feelings about the experience (Jenkins, 1998; Louth et al, 1993).

Qualitative studies in this area have also shown benefits for collaborative over individual student writing (see Topping, 1996).

However, nothing is as simple as it sounds. Studies with students suggest, first, that it is important to explain to the students what the ground rules are and, second, to give them practice in working with others. Thirdly, the studies indicate that working together should first be encouraged when the assessment is not important for student progression. Indeed, assessment is the crucial issue here - especially if there is a need to differentiate between the different contributions of members of the group to the final product.

Lejk, Wyvill & Farrow (1996) summarise three different ways of assessing individual contributions to group work as follows:

1. All marks for the contributions are *shared* equally, and are implicit in the final mark.
2. An individual's mark for his/her contribution is *added* to (or taken away from) from the group mark first awarded.
3. The group mark first awarded is *multiplied* by each individual contributor's mark.

The first method gives all members of the group the same mark. This procedure, whilst simple and effective, has an obvious difficulty in that it is unfair to different students if not all members of the group have contributed equally. With this method, therefore, some form of appeal system is typically set up for students who feel disadvantaged by this approach. Both the second and third methods involve asking the students to assess in some way their relative contributions to the joint piece of work, and to use these self- and peer-assessments in an additive or multiplicative way to allocate marks to individuals. Both methods involve exploring different ways of getting students to assess their own contribution and that of the others to the work.

Difficulties lie with the students not normally being familiar with such procedures and, perhaps at first, not being willing to use them, but guidance can be given (e.g. see Orsmond, Merry & Reiling, 2001). Students can either be asked to

assess each other's contributions on criteria, such as those in Panel 1, or be asked for an over-all, holistic assessment. Lejk and Wyvill (2001) argued that the holistic assessment was better for summative peer assessment of group contributions and this is used in the examples that follow.

Panel 1. Suggested criteria for assessing a student's contribution to an essay (adapted from Habeshaw, Gibbs & Habeshaw, 1993).

Student..... has contributed to the groups' work in the following ways:

	Good Contribution	Satisfactory Contribution	Little Contribution
Leadership and direction	+5	+3	+1
Ideas and suggestions	+5	+3	+1
Finding resources	+5	+3	+1
Structuring the essay	+5	+3	+1
Word-processing the essay	+5	+3	+1
Clarifying the argument	+5	+3	+1
Clarifying the wording	+5	+3	+1
Clarifying the conclusions	+5	+3	+1
Checking the references	+5	+3	+1
TOTAL			

There is also an issue of whether or not these self- and peer-assessments are private or public. Lejk & Wyvill (2001a), for example, found that private assessments led to a higher spread of marks for individuals than did public ones. But they also found that the general tendency for the better students to undersell

themselves, and for the weaker ones to do the reverse, was more marked in the private assessments. Lejk & Wyvill recommended that students make private peer-assessments (and avoid self-assessments) in awarding marks, and that this procedure is discussed first with the students concerned.

One way of awarding marks to a piece of work (based on the second, additive method described above) is as follows. Suppose - for example - that the final mark awarded by a tutor to a group report written by four students (A, B, C and D) comprises the sum of a mark out of 80% for the group report and one out of 20% for each individual's contribution. Thus, for each student, we need to adjust the mark initially awarded by the tutor to the group in the light of (i) the 80% weighting and (ii) the students' assessments of their contributions.

Suppose the initial mark for the report awarded by the tutor is 65%. To give this an 80% weighting we need to multiply it by 80 and divide the result by 100 - leading to a weighted mark of 52%.

To assess the student contributions, we might ask each one, for example, to rate their colleagues privately on a holistic 5-point scale, as follows:

- 20 Very good contribution to the group
- 15 Good contribution to the group
- 10 Satisfactory contribution to the group
- 5 Poor contribution to the group
- 0 Very poor contribution to the group

Panel 2 summarises the calculations needed to arrive at their final marks. Initially, for each student the average mark is taken. In this example the median is used as the marks are ordinal (although similar calculations can be done with means). The median for all the data is 12.5 whereas the expected value for accurate, fair assessments is 10 (the mid point of the scale). Each median is adjusted (standardized) by multiplying it by 10/12.5. In the case of the median of

15 for A, this is multiplied by 10 and divided by 12.5 to give 12.0. This is added to the shared report mark of 52% to give A a final mark of 64%. Without such standardization students may, for good or bad reasons, all award each other high marks for their contributions when what is needed is a valid method of assessing relative contributions to the report: it is a 'zero-sum game'. (In the example in Panel 2 the students tend to be generous to each other, on average).

Panel 2. An example of the calculations required using the additive method.

<i>Students' assessments of</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D:</i>
A's assessment	--	20	10	5
B's assessment	20	--	15	0
C's assessment	15	15	--	5
D's assessment	15	10	10	--
Median	15	15	10	5
Standardized median (20%)	12.0	12.0	8.0	4.0
Tutor's mark (80% weighting)	52.0	52.0	52.0	52.0
Final mark (total)	64.0	64.0	60.0	56.0
	Median of all data			12.5

Such procedures, which may at first seem complicated, can of course be automated (Freeman & McKenzie, 2002).

No doubt different rating scales and different weightings will be needed for different kinds of writing, depending on their importance. In the example given above, the students rated each other holistically on a simple scale. Other (less reliable) systems involve students rating themselves and others on several

different subscales (as in Figure 1), and deriving a total score (see Lejk & Wyvill, 2001a).

Panel 3 shows a similarly worked example for method 3 where the students' assessments are used to multiply the tutor's mark. Standardized median peer assessments are calculated as before. Then each standardized score is divided by 10 (the mid point of the scale) to give a multiplier in the range 0-2. Student A has a standardized score of 12 and thus a multiplier of 1.2. This leads to a final mark of $1.2 \times 65 = 78.0$. A merely satisfactory student, accurately assessed by fair peers, would have a standardized score of 10 and a multiplier of 1.0 that would not affect the initial mark of 65. (Again similar calculations can be done with means rather than medians, and this is similar to dividing the total pot of 260 marks - four times 65 - between the four students according to the proportion of the total peer assessments gained by each student).

As can be seen in the example, multiplicative adjustments can produce much larger variations between the students than can additive adjustments for a given set of peer assessments, and even lead to marks exceeding 100. To avoid this we could put narrower limits on the range of marks. For example the peer assessments on a scale of 0-20 could be used to create multipliers in the range of 0.9 to 1.1. So, tutors need to experiment beforehand with different scales and different weightings in order to decide what is appropriate for their particular concerns. While it may seem arbitrary, the additive method has the benefit of simplicity (for both tutor and students) and it avoids any possibility of marks exceeding 100. However, like method 1, it may be generous to students contributing less than others. The multiplicative method will not protect them but may produce some correspondingly high marks for better students.

Panel 3. An example of the calculations required using the multiplicative method.

<i>Students' assessments of</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D:</i>
A's assessment	--	20	10	5
B's assessment	20	--	15	0
C's assessment	15	15	--	5
D's assessment	15	10	10	--
Median	15	15	10	5
Standardized median	12.0	12.0	8.0	4.0
Multiplier	1.2	1.2	0.8	0.4
Tutor's mark	65.0	65.0	65.0	65.0
Final mark	78.0	78.0	52.0	26.0

In this article we have suggested that students will gain more from writing together than from writing individually and, because less marking is involved, tutors will have less to do in order to achieve this. Furthermore, the quality of the feedback on the assessments given by tutors might also improve – given that there is more time to comment in more detail on each piece of written work. We have not, however, considered here the assessment of team skills – the group process – only the contributions to the group product. That is a further, significant challenge in assessment.

References

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

- Habeshaw, S, Gibbs, G. & Habeshaw, T. (1993) **53 Interesting Ways to Assess Your Students**. Bristol: Technical and Educational Services Ltd.
- Jenkins, L. A. (1998) Writing laboratory reports: Individuals versus groups. Unpublished undergraduate dissertation, Department of Psychology, Keele University, UK.
- Lejk, M. & Wyvill, M. (2001a) 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(6)**, 551-561.
- Lejk, M. & Wyvill, M. (2001b) Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education*, **26(1)**, 61-72.
- Lejk, M., Wyvill, M. & Farrow, S. (1996) A survey of methods of deriving individual grades from group assessments. *Assessment and Evaluation in Higher Education*, **21(3)**, 267-280.
- Louth, R., McAllister, C. & McAllister, H. A. (1993) The effects of collaborative writing techniques on freshmen writing and attitudes. *Journal of Experimental Education*, **61(3)**, 163-167.
- O'Donnell, A. M., Dansereau, D. F., Rocklin, T., Lambiotte, J. G., Hythecker, V. I. & Larson, C. O. (1985) Co-operative writing: Direct effects and transfer. *Written Communication*, **2(3)**, 307-315.
- Orsmond, P., Merry, S. & Reiling, K. (2002) The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, **27(4)**, 309-323.
- Tang, C. (1998) Effects of collaborative learning on the quality of assignments. In B. Dart and G. Boulton-Lewis (Eds) **Teaching and Learning in Higher Education**. Melbourne: Australian Council for Educational Research (102-123).

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, **32**, 321-345.

Contacts

The authors can be contacted at:

j.hartley@psy.keele.ac.uk

s.j.bostock@keele.ac.uk