CENTER FOR FACULTY EXCELLENCE PRESENTS

for your consideration...

SUGGESTIONS AND REFLECTIONS ON TEACHING AND LEARNING

April 2012

Evaluating collaborative coursework

Formal student collaboration is an increasingly prominent feature of higher education at UNC and at other institutions (Cheng & Warren, 2000). In some instances, this trend reflects an effort to cope with diminished instructional resources (e.g., reduced instructor-student ratios or limited equipment availability). However, instructors also assign group work to students in the belief that working collaboratively provides certain educational benefits. One such benefit is enhanced learning of course content. There is now a large body of evidence showing that, across disparate disciplines, students who work in cooperative learning groups achieve greater mastery of course subject matter than do individuals who work alone (Johnson, Johnson, & Stanne, 2000; Slavin, 1995). Importantly, it is not only the academically weaker students who benefit from working in groups. Academically stronger students, who you might think have less to gain from collaborating with less knowledgeable students, also achieve higher levels of retention, comprehension, and knowledge transfer and perform better on various other measures of learning when working in groups than when working alone (Johnson et al., 2000).

Why might this be so? As you undoubtedly know from your own experience as an instructor, a very good way to deepen and consolidate one's own understanding of new material is to explain the material to someone else. Accordingly, students working in groups may learn better in part because they have more opportunities to explain subject matter to their peers than do students working alone (Webb, 1989; 2009).

A second educational benefit of group work is the opportunity to develop greater proficiency at the process of collaboration itself. Because so many of the types of work that students undertake after graduation are collaborative in nature, many instructors feel that a student's program of study should help prepare him or her to function effectively as a member of a work team (Cheng & Warren, 2000). This preparation entails learning how to formulate goals for the group, assign roles, set deadlines, communicate effectively, resolve conflicts among members, and assess the work of one's peers. Indeed, in professionally oriented programs of study such as business administration, software development, or public policy, it is common to require students to participate in a capstone course that consists of a semester-long group project designed to emulate the type of collaborative work the students will undertake when they embark on their respective professional careers (Raban & Litchfield, 2006). Whether you assign your students to work in groups out of economic necessity, pedagogical principle, or both, you will need some way to derive individual student grades from work that students produce collaboratively. Specifically, you will need 1) a method to assess the quality of the work produced by the group as a whole, and 2) a method for assessing each individual student's relative contribution to the work of the group. Below we suggest methods for carrying out both types of assessment.

Assessing the quality of the work produced by the group

When assessing student work that can not simply be marked correct or incorrect and that allows for multiple dimensions of evaluation (e.g., a business plan, policy analysis, oral presentation, video, or software product) it is helpful to devise 1) a set of criteria that specify the component skills students must combine to successfully complete the assignment and the weight you assign to each component in the overall evaluation; and 2) an evaluative range (i.e., an ordinal rating scale) representing the different possible levels of achievement for each dimension. The set of criteria essentially answers the question, "What do I expect students to demonstrate in this assignment?" while the evaluative range answers the question, "How many meaningfully different levels of proficiency can I discriminate?" This type of assessment tool is commonly called a rubric. If you want your assessment to provide students the information they need to

improve their performance on future assignments (i.e., formative) as opposed to merely evaluating the degree to which the work produced met your expectations (i.e., summative), you should also produce detailed descriptions or operational definitions of each level of achievement for each dimension. These descriptions should include a clear characterization of the highest or expected level of achievement and, for lower levels of the evaluative range, examples of the ways student performance may deviate from the highest or expected level. An example of a grading rubric for a group project in an Information Systems course (Quesenberry, 2011) can be found at cfe.unc.edu/pdfs/Quesenberry.pdf, and a rubric used to evaluate a clinical pharmacy services group project (Skomo et al., 2008) can be found at cfe.unc.edu/pdfs/ProjectRubric.pdf. Detailed instructions on how to create a rubric can be found in the FYC on Grading Rubrics.

Share the grading rubric with your students before they begin work on the project, so that they will know what aspects of the project you will be assessing and what your expectations are for each aspect.

Deriving individual grades for collaborative work

While student collaboration on assigned projects can help students better learn course material and can help prepare them for post-graduate careers that involve collaborative work, it presents you, the instructor, with an assessment challenge. In any group of students, some members of the group may contribute more to the project than other members. In the extreme case, members may shirk their responsibilities entirely, effectively acting as free riders on the efforts of their peers. If you assign the same grade to all group members regardless of their level of effort, you actually create an incentive for this sort of exploitation. Even when all group members contribute roughly equal amounts of time and effort to the group project, the value or importance of some individuals' contributions may be greater than that of other individuals. To the extent that an individual's grade for an assignment ought to reflect the quality of the student's own work and not simply the quality of the work produced by his or her group, you will need to devise a method for deriving individual grades.

Because the group members themselves are typically the ones best able to judge the relative value of each member's contribution (Sharp, 2006), a commonly used method for deriving individual grades involves 1) obtaining peer ratings of each group member's relative contribution, 2) using the ratings to determine a weighting factor for each individual, and 3) assigning individuals their group grade adjusted according to their individual weighting factor (IWF). According to this method, those individuals judged by their peers to have contributed relatively more to the project will receive individual grades somewhat higher than the group grade while those judged by their peers to have contributed relatively less will receive individual grades somewhat lower than the group grade.

Obtaining peer ratings

The simplest way of obtaining peer ratings is to ask each group member at the end of the project to rate holistically the contribution of each other group member and then divide all the ratings each person received by the number of raters to obtain the student's average rating. For example, you can have the students assign each other direct numerical ratings, such as a number from 1 to 10, with 10 indicating an exceptionally strong contribution and 1 indicating little to no contribution, or a percentage score ranging from zero to 100, where the score represents the proportion of the total group effort a given student contributed. Brown (1995), however, found this approach unsatisfactory because of "an apparent psychological compulsion of students to give [each other] high numerical marks" (p. 3c2.17). He recommended instead using a set of verbal descriptors representing an ordered continuum of levels of contribution (e.g., excellent, very good, satisfactory, deficient, etc.). In Brown's scheme, the different levels are assumed to be evenly spaced along the continuum, and you convert each descriptive rating into a numerical rating (e.g., "excellent" = 100%, "very good" = 80%, "satisfactory" = 60%, etc.).

Although this method appears simple, in practice students often find it hard to decide on a single number or descriptor that summarizes, with adequate precision, the relative value of the many different types of contribution a peer makes to a group project over a period of several weeks or months (Raban & Litchfield, 2006; Lejk & Wyvill, 2001). This cognitive difficulty, rather than any "psychological compulsion," perhaps explains why students tend to give uniformly high holistic ratings to all but the most egregiously delinquent group members. To help students make more informed and discriminating judgments about their peers' contributions, you can ask them to evaluate their peers in each of several categories of performance instead of asking for a single holistic rating. Examples of performance categories often used in peer rating instruments include dependability (e.g., met deadlines); level of effort, quality of contribution (e.g., creativity, skill); and cooperativeness (Baker, 2008; Paswan & Gollakota, 2004). To obtain a single numerical value representing a rater's overall evaluation of a peer, you can simply take the average of the separate category ratings, or, if you feel some categories of contribution should count for more than others, a weighted average. Baker (2008) provided examples of empirically validated peer rating instruments that you can adapt to your own needs (see URL in references).

Another way to help students produce peer ratings that more accurately reflect group members' differing relative contributions is to ask them to provide ratings more than once over the course of the group project. Raban and Litchfield (2006) found that requiring weekly peer ratings greatly improved the spread of the final holistic peer ratings. Specifically, students in a computer science course that included a semester-

Table 1. Hypothetical peer ratings for a group of four students.

The rows indicate the ratings that each student received and the columns indicate the ratings that each student assigned.

Rater					
Barry	Charles	Diane	Average	Weighting factor	Final Grade
100	100	90	96.67	1.12	95.27
-	80	80	81.67	0.95	80.48
90	-	90	90.00	1.04	88.70
70	80	-	76.67	0.89	75.56
86.67	86.67	86.67	86.25	1.00	85.00

long system development project were required to rate the contribution of each of their fellow team members weekly, and to track how much time they themselves spent working on the group project during the week. The introduction of weekly peer ratings markedly reduced students' tendency (observed in prior semesters) to award everyone in the 10-member groups the same grade. Raban and Litchfield pointed out that the students were neither required nor even explicitly encouraged to use the weekly peer ratings in determining a final overall rating for each fellow team member. Rather, the authors suggested, "it seems that progressiveness and visibility of peer evaluation, feedback, and review empowered individuals to claim their 'rightful' share of the marks" (pp. 691-692).

How many times should students assess the contributions of fellow group members over the life of the project? In addition to providing peer ratings at the end of the project, the students in each group should evaluate the functioning of their group and the performance of each member at least one other time. This evaluation should occur early enough in the life of the project—preferably no later than mid-way through—that students can identify problems that may be hindering the progress of the group's work or individuals who are perceived not to be contributing their fair share to the group effort while there is still time to benefit from making changes.

Using the ratings to determine a weighting factor for each individual

Because the weighting factor is meant to represent how much a given student contributed to a group project relative to how much all the other students in the group contributed, the weighting factor must take into account both the ratings assigned to the individual student and the ratings assigned to all the other students in the group. For example, an average peer rating of, say, 80% signifies a below-average contribution if the average peer rating for the group as a whole is 90%, while it signifies an above-average contribution if the average peer rating for the group is 70%. Thus, a commonly used formula for obtaining an individual weighting factor is to divide the average peer rating for an individual by the average peer rating for the group as a whole (Brown, 1995; Sharp, 2006). This method yields a weighting factor greater than 1.0 for individuals receiving above-average peer ratings and a weighting factor less than 1.0 for individuals that receive below-average peer ratings. Table 1 shows how IWFs are derived from a hypothetical set of peer ratings according to this method.

Using the individual weighting factors to determine individual grades

To determine an individual's grade on a group project, simply multiply the group grade by the individual's weighting factor. For example, if you assign a grade of 85% to the work produced by the group as a whole, and the peer rating method described above yields IWFs of 1.12, 0.95, 1.04, and 0.89 for group members Alice, Barry, Charles, and Diane, respectively (Table 1), then the students' individual grades for the project are 0.95 (1.12*.85), .80 (.95*.85), .88 (1.04*.85), and .76 (.89*.85), respectively (Table 1). Thus, while in this example you judged the work produced by the group as a whole to merit a grade of 'B,' by taking into account the group members' own assessments of each others' relative contribution to the project, you end up assigning individual grades that range from 'C' to 'A.' As in this contrived example, the use of peer ratings to adjust individual grades for group work typically increases the spread of final grades.

Further considerations

In contrast to the weighting factors presented in Table 1, the authors who have used peer ratings to adjust individual grades for work produced collaboratively report that the actual adjustments tend to be modest (Brown, 1995; Cheng & Warren, 2000), usually resulting in adjustments of only a few percentage points (i.e., from a B to a B+). This may in part reflect students' disinclination to give very low ratings to a fellow group member, which has the effect of compressing the ratings variance within a small range. But it probably also reflects the fact that, when students know that their grade on a group project will depend in part on how their peers rate their contribution, they feel motivated to put in a good effort in order to earn favorable ratings. Nonetheless, it is possible with the method described above to end up with a situation where two students in a group receive markedly different individual grades

for the same piece of work. To avoid this outcome, you can impose a limit (e.g., 10 percentage points) on the amount by which a student's final grade can deviate from the grade assigned to the group as a whole. Sharp (2006) provided a method for scaling the raw peer ratings to keep the IWFs within the predetermined range.

A very different concern is that the peer ratings assigned to a student may reflect not only the student's actual relative contribution but also the (perhaps implicit) gender, racial, and other biases of the person doing the rating. For example, suppose that in a group comprising both males and females where everyone contributed equally, the males in the group gave systematically lower ratings to the female students than to male peers. This would result in the females unfairly receiving lower grades. Kaufman, Felder and Fuller (1999) addressed this concern in a study of the ratings that students in two chemical engineering courses assigned to the members of their cooperative learning teams, and found no evidence of either gender or racial bias.

There are many good reasons to assign students to work together in groups, but in order to do so you need a way both to assess the quality of the work produced by the group's collective effort and a way to derive individual grades that reflect each student's relative contribution to the work. In this essay we have suggested using a multidimensional grading rubric to assess the work of the group as a whole and then adjusting the group grade up or down for each individual according to a weighting factor based on peer ratings. Staff at the Center for Faculty Excellence are available to assist you in designing grading rubrics and peer assessment forms and to help you create a spreadsheet you can use to obtain individual student weighting factors and adjusted grades.

Bibliography

Baker, D. F. (2008). Peer assessment in small groups: A comparison of methods. Journal of Management Education, 32(2), 183-209. Available from http://jme.sagepub.com/content/32/2/183.

Brown, R. W. (1995). Autorating: Getting individual marks from team marks and enhancing teamwork. 1995 Frontiers in Education Conference Proceedings. Pittsburgh, PA.

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to group project. Teaching in Higher Education, 5(2), 243-255.

Davis, B. G. (2009). Tools for teaching (2nd ed.). San Francisco, CA: John Wiley & Sons.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Cooperative learning methods: A meta-analysis. Unpublished manuscript.

Kaufman, D. B., Felder, R. M., & Fuller, H. (1999). Peer ratings in cooperative learning teams. Proceedings of the 1999 Annual ASEE Meeting, Session 1430. Available at http://wwwtemp.asee.org/conferences/annual/pdf/upload/1999-PIC-IV-Best-Paper.pdf

Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. Assessment & Evaluation in Higher Education, 26(1), 19-39.

Paswan, A. K., & Gollakota, K. (2004). Dimensions of peer evaluation, overall satisfaction, and overall evaluation: An investigation in a group task environment. Journal of Education for Business, 79(4), 225-231.

Raban, R., & Litchfield, A. (2007). Supporting peer assessment of individual contributions in groupwork. Proceedings of the 23rd Annual Ascilite Conference: Who's Learning? Whose Technology?, 23(1) 34-47.

Sharp, S. (2006). Deriving individual student marks from a tutor's assessment of group work. Assessment & Evaluation in Higher Education, 31(3), 329-343.

Slavin, R. E. (1995). Cooperative learning: Theory, research, and practice. Boston, MA: Allyn & Bacon.
Webb, N. M. (1989). Peer interaction and learning in small groups. International Journal of Educational Research, 13(1), 21-39.

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. British Journal of Educational Psychology, 79(1), 1-28.