

Building Better Teamwork Assessments: A Process for Improving the Validity and Sensitivity of Self/Peer Ratings

Eric Van Duzer and Flora McMartin

Abstract: A process employing both quantitative and qualitative methods was developed to improve the validity and sensitivity of self/peer ratings in assessing teamwork skills. Preliminary results indicate a dramatic improvement in the sensitivity of scales in measuring differences between student skill levels. The data also indicate that the process improves the validity of the ratings in measuring what the developers intended.

Introduction

Traditional engineering education emphasized individualism, in contrast current practice increasingly involves team projects, cooperative learning and an emphasis on the synergy possible through group processes. Most faculty who interact regularly with their students have a general sense of a student's teamwork skills. However, in order for faculty to develop effective interventions, it is necessary for them to measure the underlying skills that contribute to successful teamwork. Building on a number of existing instruments (e.g., the Foundation Coalition's Team Evaluation Sheet) the Synthesis Coalition's assessment team developed a self/peer assessment instrument that incorporated the best practices in engineering and other disciplines. This instrument was initially pilot tested in the Fall of 1997, revised in the following spring and a follow-up pilot test was conducted in the summer of 1998. The revision process outlined in this paper significantly improved the sensitivity and validity of the teamwork instrument. It provides a model process for developing local self/peer assessment instruments.

Developing the Instrument

In order to determine which characteristics of self/peer assessment instruments would improve their validity and reliability, a review of the literature on self-assessment and peer evaluation was conducted. That literature revealed two primary types of bias, self enhancement and downward comparison, that can distort self-ratings on assessment instruments (Mabe & West, 1982; Groeger & Grande, 1996).

Self enhancement is the unreasonably optimistic self appraisal that may be triggered by threats to self-esteem, for example, by asking nursing students to rate their level of competence in treating their patients. *Downward comparison*, is a general tendency for positive self-bias and negative other bias when self-evaluation involves social comparisons, such as "Compared to other freshman". The literature indicates that the effects of these biases can be reduced and thus enhance response reliability and validity by:

- Using explicit language shared by respondents and testers in defining the traits and the criteria.

OVER

- Correlating self-assessments with the scores of multiple raters.
- Designing questions that are based on past performance and not on perceived ability or expected future performance.
- Making social comparisons with an explicit group of known individuals (e.g., the members of your team) when social comparisons are required.

To build on the literature review, a meeting of students, faculty, and industry participants was held to define a set of abilities related to specific Synthesis learning activities that could be measured by the teamwork instrument. The results of this meeting were disseminated back to the group and a brainstorm and negotiation process was used to refine the set of teamwork abilities to be included in the pilot instrument. By involving all of the constituents, we not only built upon an expert base of knowledge, but also developed a process for getting buy-in from the Coalition's partners and faculty. The final items for the instrument were culled from the contributions from these constituents and edited to meet the standards identified in the literature.

The structure of the instrument was patterned after sample instruments that were collected from a variety of sources including engineering education, communication, engineering practice, and teacher education (Johnson, 1997; Villson, 1997). In crafting the teamwork instrument, model items from the best of these sources were adapted to reflect the Synthesis skills defined by the brainstorm and negotiation process described above. The goal was to generate a concise, easily completed instrument that would inform faculty efforts to target deficiencies that impact on teamwork abilities.

The Initial Pilot Test

A freshman design course and an upper division design course at two different universities were selected for the pilot test. A total of 142 students completed the instrument. Once the data had been collected, there was an analysis of:

- The level of the variation among ratings.
- The correlation between pairs of items intended to measure the same abilities.
- The correlation between the distribution of 100 credit points among members of the team and the overall points scored in the self/peer rating section.
- The correlation between the content in written comments and the overall points given to a team on the instrument.

Once the initial trial was conducted and the results analyzed, the instrument (see Appendix A) was revised and retested with a group of 21 students enrolled in a summer school section of the lower division course involved in the initial pilot test. The same basic analysis was conducted and the results were compared to the pilot data. The only portion of the analysis that could not be compared were the written comments due to the small number of responses obtained in the subsequent pilot test.

Preliminary Results

The principle finding from the preliminary results was that there was unacceptably low variation among scores. Preliminary findings indicated that:

- On a scale of 1 to 4, over 90 percent of the items were rated at the highest level (4).
- The correlation between related items intended to measure the same constructs were difficult to interpret given the low variation of ratings, but were generally in the expected direction.
- Students were asked to distribute a hundred points among the members of the group as a measure of each individual's contribution to the group's efforts. The correlation between this distribution and the other items on the instrument indicated that the points were allocated almost entirely on the basis of the work completed, rather than general teamwork abilities, participation, and contribution.
- In the content analysis of written comments, there was a one-to-one correlation between the number of positive or negative comments and the overall points assigned to each team.

Revising the Instrument

In an effort to improve the instrument both quantitative and qualitative results were analyzed. The strong correlation between written comments and the overall scores indicated that the instrument was capturing information about the team. However, the lack of variation in the scores on the pilot test indicated a lack of sensitivity and limited the ability to assess differences among students or types of skills. This led to a re-evaluation of the instrument as a whole.

Items with the lowest variation were removed or replaced. The items that received almost universal ratings of 4 on a scale of 1-4 were eliminated (e.g., items 1 and 6 in Part II of the instrument) because they showed little promise of providing useful information. Issues raised in the written comments as well as some professional judgement by the assessment staff influenced the selection of a set of replacement items that had not been included previously because of limited space.

Second, the scale which had been a standard “strongly agree, agree, disagree, strongly disagree” was replaced with “agree, tend to agree, tend to disagree, disagree.” This language was chosen to increase the sensitivity of the scale because it did not require respondents to choose between the dichotomous positions of agree and disagree.

Given the problem of “response set” (i.e., students gave the same rating to all items in a particular section) several items on the self/peer assessment were reversed. For instance, in the initial instrument students were asked to rate whether a team member “Helps divide the work fairly”. In the revised instrument, this was changed to “Failed to do their share of the work”. This was intended to get students to think more carefully about their answers.

A final change in the instrument was designed to increase the relationship between the distribution of credit points among the group's members (see section 4 of the instrument) and the overall scores in the teamwork ability ratings (Part II of the instrument). Rather than simply asking students to distribute one hundred points among the members of the team on the basis of their contributions, the prompt was changed to include an explicit reference to the various elements rated on the self /peer section. Additionally, in an effort to expand the usefulness of this instrument students were asked to also provide a job description for each member of the team (see Appendix B).

The Follow-up Pilot Test

To improve the clarity of the language used in the instrument and to insure that the respondents and authors had similar interpretations of the meaning of the responses, a set of one-on-one interviews with students were arranged. A total of eight upper and lower division students were invited to complete the form while talking out loud about their understanding of the question and the reason for the response. Any unclear or conflicting comments made by the students during these sessions were followed up on by the researchers once the students had completed the form.

Several students were asked to complete the original instrument. Building on their comments the instrument was revised and presented to subsequent student volunteers. During the process of testing the instrument by way of the verbal protocols, the instrument was revised two more times. The language was improved so that the interpretation of the students matched the interpretation of the authors who created the questions and analyzed the data. For instance, in the pilot instrument one item in the self /peer rating section was: "Often dominates the conversation". This was interpreted to be negative by the authors as a way to demonstrate a student's lack of ability to function smoothly in a team environment. However, the students involved in the interviews indicated that they gave high ratings to indicate a member's leadership during meetings. Therefore, the item was revised to read: "*Excessively* dominate the conversation". This simple wording change improved the clarity of the trait, and ultimately, the sensitivity of the instrument to measure what the authors intended.

Follow-up Pilot Test Results

As Table 1. illustrates, the process for revising the instrument resulted in considerably greater sensitivity in measuring differences among students. There is still an upward bias common to self/peer instruments, (i.e. over 80% of the scores reflected better than average ability), but the responses vary considerably more than in the pilot test.

Table 1: Comparison of Ratings for Initial and Revised Instrument

Rating	Pilot (N=142)	Revised (N=21)
1	.07 (%)	5
2	2	10

OVER

3	7	33
4	91	51

Rating Scale: 1=disagree 2=tend to disagree 3=tend to agree 4=agree

The data also seems to indicate that the effort to broaden the criteria upon which students allocate the 100 credit points was successful. In the pilot instrument there was not a significant correlation between the point distribution and the overall teamwork ratings in the self/peer section. The only significant correlation ($p = .05$) with the distribution of credit points was with the first item "Helps divide the work fairly". In the revised instrument the correlation between credit points and overall teamwork scores reached the .05 significance level.

Finally, the information gathered from the roles students assigned to their teammates, together with the distribution of credit points, provided useful data about the relative roles of team members and the value students place on particular sets of abilities and contributions (see Table 2). Students were asked to provide a brief (one or two word) description of the role each team member played. Examples (e.g., idea person, facilitator, etc.) were provided to stimulate student thinking. Instead of creating their own job descriptions, the majority of student responders used one of the exemplars. It is unknown to what degree this makes interpretation of the results problematic. In the current trials students are asked to both assign a role and provide a one-sentence rationale for it. Future results will help clarify these insights into the respondent's values.

Table 2: Role Identification and Credit Point Allocation on Revised Instrument

Role	# of Points Allocated to Role	# of Time Role was Listed
Technician	28.5	14
Programmer	27.8	10
Organizer	25	15
Problem solver	25	5
Idea person	24.07	23
Negotiator	21.9	8
Yes man	20	5

There were too few written comments in the second set of data to make any conclusions concerning the correlation between positive and negative comments and student peer ratings. The small number of respondents in the second set of data ($N = 21$) make any conclusions tentative. However, both the follow-up test data and the student responses in the verbal protocols do demonstrate that there was a dramatic change in the levels of variation among the self/peer scores (see Table 1). This process of revising the instrument was effective in making it more sensitive and thus provided better information about the relative teamwork abilities of the students.

Finally, in an effort to determine the accuracy of the ratings in capturing teamwork skills and therefore validate the instrument, interviews and focus groups were conducted with team members after they had completed the instrument during their engineering course. There was

strong agreement between the interview data and the instruments. The one exception was a lack of correlation between a member's work effort as discussed in the interview and as rated on the forms. Here there was a tendency among the team members to be more generous in their ratings of one another on the instrument, than during the interviews. Future versions of the teamwork instrument will need to explore ways of improving the accuracy of this element.

Conclusion

The early results with the revised instrument suggest that it provides an accurate measure of the student's participation in teams. Studies underway should help improve the instrument and provide additional evidence of its utility in assessing team skills. The process of combining constituent input, statistical analysis, content analysis and verbal protocols provides a useful model for developing local self/peer assessment instruments. As a result of this process, this tool can help faculty identify problems early and provide remedies that will help teams to be successful. Because the instrument measures a variety of teamwork abilities it provides information at a relatively fine scale, allowing faculty to tailor their teaching to weak areas and improve their effectiveness in promoting the skills. For faculty who are inclined to grade on the basis of teamwork contributions, the revised instrument also seems to provide an accurate reflection of the team's impression of student contributions through the overall credit distribution. This provides useful feedback for either mid-term corrections, or summative grading based on team participation and contributions.

Acknowledgements

The authors would like to acknowledge the support of the Synthesis Coalition funded by the National Science Foundation Engineering Education Coalitions Program (Award No. EEC 9625456).

References

- Cole, D.A.; Maxwell, S.E.; Martin, J.M. (1997). "Reflected self-appraisal: Strength and structure of the relation of teacher, peer, and parent ratings to children's self-perceived competencies." *Journal of Educational Psychology*. 89(1), 55-70.
- Groeger, J.A. & Grande, G.E. (1996). "Self-preserving assessments of skill?" *British Journal of Psychology*. 31(4), 61-79.
- Johnson, V. (March 1997). "The Design Of Survey Questionnaires to Identify the Critical Education and Support Processes Most in Need of Improvement". Presentation at the Rose Hulman Best Assessment Practices in Engineering Education: A Working Symposium. Terra Haute, Indiana.
- Katzenbach, J.R. (1993). *The Wisdom of Teams: Creating the High-performance Organization*. New York: Harper Business.

Krathwohl, D.R. (1993). *Methods of Educational and Social Science Research*. New York: Longman.

Mabe, P. A. and West, S.G. (1982): "Validity of self-evaluation of ability: A review and meta-analysis." *Journal of Applied Psychology*. 67(3), 280-296.

Steel, R.P.; Mento, A.J.; Davis, C.L.; & Wilson, B.R. (1989). "Psychometric properties of a measure of sense of competence. *Educational and Psychological Measurement*. 49(2), 433-446.

Villson, V.L. (March 1997). "Questionnaire and Survey Development: Principles of Good Practice". Presentation at the Rose Hulman Best Assessment Practices in Engineering Education: A Working Symposium. Terra Haute, Indiana.

APPENDIX A
Self/Peer Team Assessment #2
 Team name: _____

Part I: Team Effectiveness ratings:

Please circle the rating that best describes your response for each of the three questions below:

Do all members of the group share in the team's responsibilities?

- | | | | |
|---------------------|-------------------------|------------------------------------|------------------------------------------------|
| 1 | 2 | 3 | 4 |
| nobody did anything | one person did the work | a few group members
do the work | the work is
widely shared by all
members |

Did the team stay on-track during meetings?

- | | | | |
|--------------------------|------------------------------------------------|--------------------------------------------------------------|-------------------------------------------------------------------------------|
| 1 | 2 | 3 | 4 |
| never did get
on-task | often off-task,
with little
accomplished | wandered at first
but became focused
once we warmed up | stayed on-task
and accomplished our
agenda with little
wasted effort |

Did your group accomplish its purpose?

- | | | | |
|---------------------------------|----------------------------------------|--------------------------------------|----------------------|
| 1 | 2 | 3 | 4 |
| did not accomplish
our goals | accomplished some but not
all goals | satisfactorily accomplished
goals | group exceeded goals |

PART II: Self and Peer Assessment:

Using the following scale (1 = Strongly disagree, 4 = Strongly agree) please rate yourself and each other member of your team on how well the following phrases describe the team's work:

- | | | | |
|--------------------------|--------------------------|-----------------------|-----------------------|
| Strongly disagree | Somewhat disagree | Somewhat agree | Strongly agree |
| 1 | 2 | 3 | 4 |

The Team Member:

1. Values diverse opinions
2. Helps divide the work fairly
3. Helps guide the group to rationale decisions
4. Participates in team discussion without dominating the conversation
5. Delivers assignments on a timely basis
6. Works with team to productively resolve conflicts

	self	name	name

OVER

Section 3: Write a brief description of the problems you encountered in working with this group and how they were resolved.

Section 4: Please distribute 100 credit points to your team members. Each member (including yourself) should get the points appropriate to their contribution to the teams efforts. The total points should add up to 100.

	# of Points
Name:	
(Self)	
Total	100

OVER

APPENDIX B
Self/Peer Team Assessment

Team Name: _____ **Date** _____

1. Please circle the rating that best describes your team for each of the three items below:

a. Did all members of the group share in the team’s responsibilities?

- | | | | |
|---------------------------------------|------------------------------------------|----------------------------------------------------|-----------------------------------------------|
| some members
did no work
at all | a few members
did most of
the work | the work was
generally shared by all
members | everyone did an
equal share of
the work |
|---------------------------------------|------------------------------------------|----------------------------------------------------|-----------------------------------------------|

b. Which of the following best describes the level of conflict at group meetings:

- | | | | |
|-----------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------|
| no conflict, everyone
seemed to agree on what to
do | there were disagreements,
but they were easily
resolved | disagreements were
resolved with considerable
difficulty | open warfare: still
unresolved |
|-----------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------|

c. How productive was the group overall?

- | | | | |
|-------------------------------------------------------------------|--------------------------------------------------------------------|----------------------------------------------------------------|----------------------------------------------------------------------|
| accomplished some but not
all of the project’s
requirements | met the project
requirements but could
have done much better | efficiently
accomplished goals
that we set for ourselves | went way beyond what we
had to do exceeding even
our own goals |
|-------------------------------------------------------------------|--------------------------------------------------------------------|----------------------------------------------------------------|----------------------------------------------------------------------|

2. Please rate yourself and each team member on how well the following phrase describe your team’s work:

Disagree	Tend to disagree	Tend to agree	Agree
1	2	3	4

Team Members’ Names: | self: | _____ | _____ | _____ | _____

a.	Failed to do an equal share of the work				
b.	Kept an open mind/ was willing to consider other’s ideas				
c.	Was fully engaged in discussions during meetings				
d.	Took a leadership role in some aspects of the project				
c.	Helped group overcome differences to reach effective solutions				
e.	Often tried to excessively dominate group discussions				
f.	Contributed useful ideas that help the group succeed				
g.	Encouraged group to complete the project on a timely basis				
h.	Delivered work when promised/needed				
i.	Had difficulty negotiating issues with members of the group				
j.	Communicated ideas clearly/effectively				

OVER

3. Please review the items on the front side of this form and then write a brief description of any problems or conflicts you encountered in working with this group and how they were resolved.

4. Based on your opinion of each team member's performance (including yourself):

- Please distribute 100 points among the members for their overall contribution to the team's efforts (including work, communication, problem solving, etc.). The total points should add up to 100.
- Assign a job title such as organizer, yes-man, negotiator, idea person, technician, pessimist, obstructor, etc. that best describes the role each member assumed in the group.
- Please provide one reason why you assigned that role.

Name:	# of points	Job Title	Reason for assigning role
(Self)			
Total	100		

5. Please circle the name of the member who you think provided the most leadership in this group.

6. Describe what you learned about team projects and working with other team members during this assignment:

