TEAMS IN SOFTWARE ENGINEERING EDUCATION

Judith E Sims-Knight[1], Richard L Upchurch[2], T. A. Powers[3], Sara Haden[4], Raluca Topciu[5]

Abstract—*The ability to work as an effective member of a development team is a primary goal of engineering education and one of the ABET student learning outcomes. As such, teaming has received increased attention in both the classroom and the literature over the past several years. Instructors of software engineering courses typically organize students into teams, but expect, erroneously, that students learn the skills they need and learn to avoid dysfunctional patterns simply by working in teams. This paper describes the development of tools that can incorporate an assessment-based continuous improvement process on team skills into engineering classes. The primary focus is on the development of 1) a self-report assessment tool that would provide pointers toward improvement and 2) a test of students' knowledge of best teaming practices. The paper also describes a first pass at embedding these assessment tools into a continuous improvement process.*

INTRODUCTION

To meet the ABET student learning outcome "an ability to function on multidisciplinary teams" engineering faculty across the country are trying different strategies to incorporate such assessments into their courses. Most approaches involve either self assessments or peer assessments or both [1-3]. Many reports have addressed the feasibility of incorporating team training and assessment, although relatively few have provided evidence of validity or effectiveness. Thompson [3] demonstrated that peer evaluations of team skills improved over the course of a semester. Dominic, Reilly and McGourty [1] demonstrated experimentally that peer evaluations, whether shared with the target team member or not, resulted in better team performance, as measured by ratings of videotaped team activity, compared to a control group who had no peer evaluations.

These studies suggest that peer and self assessment is a viable technique for course-based assessment of teaming. Nonetheless, it is less time-consuming and potentially less divisive to assess the team as a whole. Thus, we have focused on developing a tool by which each student assesses the team as a whole. The purpose of our second tool is to assess students' knowledge of how effective teams function. Our ultimate goal is to produce assessments that can be used in courses as part of a continuous improvement model. To do this assessments should not only be reliable, valid, and easy to administer, but they should provide pointers toward improvement for both students and faculty. The two instruments used in this study are designed to provide such pointers.

The first is a self-report inventory of individual students' assessments of their team's processes, The Team Process Check (TPC), which is an adaptation of an instrument used at Arizona State University [4-5]. As described in Powers et al. [5], this adaptation provided two distinct scores, one for agency issues (team process and decision making) and one for affiliation issues (communication and conflict resolution). Thus, if teams had, for example, high scores on affiliation and low scores on agency, it would indicate that they needed to work on agency issues. At the end of the TPC students were asked directly to indicate areas they felt the team needed work on.

The second assessment was a Team Knowledge Test (TKT). When people do not understand much about a skill, self assessments are likely to exhibit overconfidence, particularly among men [6-8]. To counteract overconfidence and to provide a better knowledge base for their judgments, students were assigned readings in a team training website. This material covered the basics of teaming--team roles, effective meetings, communication styles, and conflict resolution. In addition the instructor did two teaming exercises. The TKT tested their understanding of this material.

In Power et al. [5] the TPC and the TKT were used in a freshman engineering course that was part of an integrated cluster of courses. In this study they were used in a software engineering (SE) course, which is a senior-level course taken by both Computer and Information Science and Computer Engineering students. The course required a semester-long team project. Teams of 46 students worked together to specify and build a product during the semester. The instructor chose project leaders who met with the instructor to receive a draft problem statement. The managers reviewed the resumes of those enrolled in the course and chose individuals they wished to have on their team. Managers then took turns choosing team members. They were asked to defend their choices based on project criteria and the contents of the resume.

Each project had a real customer with whom the teams met to determine the requirements. Each person on the team

[1] Judith E. Sims-Knight, Psychology Department, U. of Massachusetts Dartmouth, N. Dartmouth, MA 02747, jsimsknight@umassd.edu
[2] Richard L. Upchurch, CIS Department, U. of Massachusetts Dartmouth, N. Dartmouth, MA 02747, rupchurch@umassd.edu
[3] T. A. Powers, Psychology Department, U. of Massachusetts Dartmouth, N. Dartmouth, MA 02747, tpowers@umassd.edu
[4] Sara Haden, Psychology Department, U. of Massachusetts Dartmouth, N. Dartmouth, MA 02747, g_shaden@umassd.edu
[5] Raluca Topciu, Psychology Department, U. of Massachusetts Dartmouth, N. Dartmouth, MA 02747, g_rtopciu@umassd.edu

0-7803-7444-4/02/$17.00 © 2002 IEEE November 6 - 9, 2002, Boston, MA
32nd ASEE/IEEE Frontiers in Education Conference

was assigned a process role, (quality assurance, configuration management, project planning, project tracking) for the duration of the project. The process roles were distributed among team members through negotiations among team members and the project manager as were the development activities, e. g., design and coding.

Improving team process was one of the objectives of the course.

THE STUDY

Participants.

The 43 students of the software engineering course participated. There were 10 teams.

Measures

The Teamwork Knowledge Test was developed by the researchers and was loosely based on the format used by Stevens and Campion [13] for their Teamwork KSA Test [in 4 is a description of the KSA test with engineering freshmen]. The current test, however, was designed for use with an undergraduate college population rather that an industrial or corporate population. Its 21 items are designed to sample students' understanding of the four domains covered on the team training website--team process, decision making, communication, and conflict resolution. The questions are multiple choice questions with one right answer. A question from each domain appears in Appendix A. A copy of the full test is available from the authors.

Students took this test on the web at the beginning and end of the course.

The Team Process Checks (TPC; see Appendix B) were 20 statements about one's team. Participants were to rate how true, on a 1-5 scale, each item was of their team. The measure was constructed to cover two broad dimensions of team functioning, namely team agency and team affiliation. The team agency dimension attempts to assess areas such as team process and team decision making. The affiliation dimension is intended to assess interpersonal functioning, particularly communication and conflict resolution. The students completed the TPC individually, as a web-enabled survey, in the middle and end of the semester.

At the end of the semester the instructor filled out faculty ratings (see Appendix C). These were 8 statements about the team that the instructor rated on a 5-point scale ranging from *never* to *always* to evaluate the teams on these dimensions. The instructor had met with 6 teams every week and with 4 teams every other week (the latter were being jointly supervised by a part-time instructor). He based his ratings on his experiences with the teams throughout the semester.

Grades on their projects were a combination of process activities and products. Total possible score was 400 points, divided into 200 points for documents, 100 points for reviews, and 100 points for participation.

Procedure

The time-line of the procedure was as follows:

1. TKT pre-test (administered via the world wide web during first month of classes)
2. Assignment of team training material posted on a website (assigned at the end of the first month of classes)
3. TPC time 1 (administered via the web at the end of the second month of classes)
4. TPC time 2 (administered via the web in the last two weeks of classes)
5. TKT post-test (administered via the web in the last two weeks of classes)
6. Faculty Ratings (provided by the instructor at the end of the semester)

RESULTS

The means and standard deviations of all measures are presented in Table 1.

Table 1

Means, Standard Deviations, and Cronbach's alpha of Assessment Measures

Assessment	Time	Mean	Stand. Deviation	Cronbach alpha
TKT	pre	15.64	3.89	.78
	post	15.29	2.67	.76
TPC Agency	time 1	3.73	.58	.81
	time 2	3.59	.53	.82
TPC Affiliation	time 1	3.96	.61	.72
	time 2	3.77	.55	.74
Faculty Ratings of Team		3.50	.80	
Project Grade		275.44	67.54	

TKT

Students answered 74% of the TKT questions correctly at the beginning of the semester (see Table 1). They did not, however, improve their performance over the course of the semester, $t(44) = .78$, *NS*.

Several assessments of the quality of the TKT were undertaken. Both internal consistency and test-retest reliability was assessed. Cronbach's alphas for both occasions were high (see Table 1), as was the consistency across administrations, $r(34) = .75$, $p < .01$. Item analyses of all items were also conducted. The average discrimination index (DI) for the preTKT was .53 and for the postTKT, .45. Only 1 DI of the 21 items on the preTKT was less than .40. The discriminability of the postTKT was less impressive--7 items had DIs lower than .40. Nonetheless, overall the TKT appears to be a reliable test that generally discriminates between knowledgeable and not-so-knowledgeable students. These results are consistent with those found in our earlier study of freshmen in an integrated engineering program [5].

TPC

Students' self ratings of their teams were moderately high at both administrations (see Table 1). The ratings significantly

decreased from the middle to the end of the course, $t(43) = 2.05$, $p < .05$ for agency, and $t(43) = 2.69$, $p = .01$ for affiliation.

Powers et al. [5] found two distinct factors in the TPC, agency (6 items) and affiliation (7 items) (see Appendix B). The two-factor structure was verified with these data by confirmatory factor analysis, although the fit was impressive only at time 2. The structure was verified, with $\chi^2 = 98.04$, $p < .01$ and *RMSEA* = .11 for time 1 and $\chi^2 = 62.10$, *NS* and *RMSEA* = .00 for time 2. The resulting scales evinced adequate reliability (see Table 1).

As Table 2 reveals, the two subscales are substantially intercorrelated, both within and across times. Although these correlations are substantial, it is worth remembering that when two variables have a correlation of .50, they only share 25% of the variance (i. e., 75% of what influences the score is left unexplained).

Table 2
Intercorrelations Between TPC Subscales For Two Administrations[a]

		Agency		Affiliation	
		Time 1	Time 2	Time 1	Time 2
Agency	Time 1	1.000	.71**	.53**	.46**
	Time 2		1.000	.62**	.68**
Affiliation	Time 1			1.00	.70**
	Time 2				1.000

a. n = 38
** p< .01.

The two scales of the TPC evinced discriminative validity. At both administrations, students rated their teams affiliation skills higher than their agency skills, $t(36) = 2.4$, $p < .05$ and $t(37) = 2.57$, $p < .05$ for time 1 and time 2, respectively.

Intercorrelations Among Measures

The most important question to answer about an assessment tool is whether it is valid, that is, does it measure what it is supposed to measure. In this case, the question is whether the TPC and TKT really assess how well students function in their teams. This is a difficult proposition, because no single validated "test" of teaming skill is available nor are differentiated measures of affiliation and agency. Thus the best one can do is to compare various measures that are available. In this study we have three other measures that may be related to the TPC—pre and post TKT, faculty ratings of teaming, and project grade.

Table 3 shows the correlations among these measures. Note that the TKT assessments are strongly correlated with all TPC assessments except one. This suggests they are both tapping, in part, the same underlying construct, presumably teaming skills. Faculty ratings were not so highly correlated with the two teaming assessments. Only the affiliation subscale of the TPC correlated significantly. Project grades correlated with both administrations of the TKT and with the affiliation subscale of the TKT at time 2. Project grades, however, are a weak validation measure, because they assessed both process and product. Project grades might be more highly correlated with TKT because they are both typical of the academic skills that lead to high grades.

Table 3
Correlations among Teaming Assessments

		Pre-TKT	Post TKT	Faculty Ratings	Project Grade
Agency	Time 1	.21	.46**	.26	.06
	Time 2	.49**	.54**	.17	.30+
Affiliation	Time 1	.36**	.50**	.39*	.20
	Time 2	.50**	.49**	.30+	.34*
TKT	Pre			.30+	.44**
	Post			.28	.41*

+ p < .10 * p < .05 ** p < .01

To see whether the TPC and TKT were measuring different aspects of teaming, we computed a hierarchical multiple regression. The variables were blocked by time of administration (first the pre TKT, then the 2 TPC subscales at time 1, then the TPC subscales at time 2, and finally both the faculty ratings and Post TPC). Block 1, the pre TKT scores, yielded a significant $R = .44$, $F(1, 29) = 7.06$, $p < .05$. None of the other blocks added significant additional variance. This demonstrates that the TPC, TKT, and faculty ratings are all tapping the same aspects of teaming, at least so far as their relation to project grades is concerned.

DISCUSSION

Developing assessments of team process that have demonstrated reliability and validity and are still useful in classrooms is a difficult job. This study has tested two distinct measures. The TKT assesses students' declarative knowledge of what ought to be done in teams and the TPC assesses students' opinion of their team's success in applying that knowledge in two domains--affiliation and agency.

The TPCs generate two separable scales, each of which has respectable reliability. These two scales represent the two major dimensions of teaming [9]. Not surprisingly, the two scales share substantial variance, approximately 25% of the total variance. These same two subscales were found by Powers et al. [5] with freshman engineering students.

The decrease in students' ratings of their teams using the TPC has been found in other research. In our earlier study students rated their teams higher on the first TPC than on the other two administrations (although the time 1 in [5] was earlier in the semester). Knight et al. [2] also found that team self assessments went down over the course of the semester in their control group and that the teams coached by counseling students simply managed to stay the same. This pattern has been explained as produced by a response shift construct [10-11]. In situations in which people are relatively ignorant of the domain at first testing, their ratings may be based on different understanding at earlier administrations than they are at later administrations.

Of the two TPC subscales, affiliation was the stronger. It

correlated with both faculty ratings and project grades, albeit at one time only. This is inconsistent with the findings of Powers et al. and the field in general, which finds agency the more powerful variable [12]. Whether these differences are due to the difference in settings--senior vs. freshmen, one large semester-long project vs. a short more circumscribed one, computer science/engineering students vs. varied engineers (electrical, mechanical, computer)--can not be resolved at this point.

Students did not improve on the TKT. This suggests that students require more than homework assignments and one or two team exercises to gain the knowledge tested on the TKT. Perhaps it would have been better to grade the post TKT (i. e., to give differential amounts of credit for differential performance). This was done with the freshmen in Powers et al., [5] and they improved from pre to post test.

The TKT predicted project grades. Because the SE students did not improve in the post TKT, it is NOT likely that this correlation with project grades has to do with what they learned about team practices. At best, the pre TKT tapped preexisting knowledge of teaming, which helped students to succeed in this team-based course. At worst, students did well on the pre TKT and the course because of their skills and motivation to do well.

Self- and peer-assessments of teaming have been shown to be effective in promoting improved team process as assessed by professional observations of videotaped team meetings [1]. Demonstration of validity of such measures in the classroom will depend on a network of weak measures--grades that include evaluation of team process, faculty ratings, and student improvement. All of the criterion measures in this study can be improved. The team process grade included 100 points for participation, but team process involves more than simply participation. The instructions for the faculty ratings implied that judgments should be made on the whole semester, but that prevents accurate ratings for those who change. Perhaps this could be solved by having faculty make ratings twice during the semester and specifying that the second rating be done for behaviors since the first rating. Finally, the instructional intervention needs to be improved, as will be discussed next.

IMPROVING THE PROCESS

Outcomes-based assessment models expect faculty to determine if programs are meeting their goals by collecting data from students. Teaming is an important expectation of engineering education. The exploration reported here suggests that faculty who wish to help students improve in teaming can not do so passively. In this course a substantial structure for team improvement was built, including self-assessments of both knowledge base and team process plus team training reading and exercises. This structure was not sufficient to promote improvement even in basic declarative knowledge.

The project did provide pointers for improvement. First, giving the TKT as a test after the assignment of the team training material would ensure that the students took the reading assignment seriously and that they had the knowledge early in the semester. Second, having the students meet to discuss their TPC results and to write specific plans on how their teams might improve would provide the basis for students to improve using the TPC as the assessment tool. Fahr et al. [13] have demonstrated that peer ratings are effective in such a developmental context. Third, the instructor should do class exercises that address the areas the students have designated as needing improvement.

REFERENCES

[1] Dominick, P. Reilly, R. R., & McGourty, J. (1997). The effects of peer feedback on team member behavior. *Group & Organization Management*, 22(4), 1997, 508-520.

[2] Knight, D. Poppen, W., Seat, J. E., Parsons, J., Klukken, G., Glore, A. Coaching engineering design teams. *Proceedings of the American Society for Engineering Education Annual Conference and Exposition*, 1999.

[3] Thompson, Robert S. A Repeated Measures Design for Assessment of Critical Team Skills in Multidisciplinary Teams. *Proceedings of the American Society for Engineering Education Annual Conference and Exposition*, 2000.

[4] Powers, T. A., Sims-Knight, J., Haden, S. C. and Topciu, R. A. "Assessing team functioning with college students." Presented at the American Psychological Association Convention. San Francisco, California, 2001.

[5] Powers, T. A., Sims-Knight, J., Haden, S. C. and Topciu, R. A. "Assessing Team Functioning in Engineering Education." To be presented at the American Society for Engineering Education Annual Conference and Exposition, Montreal, 2002.

[6] Bjork, R. A. Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), Attention and performance XVII; Cognitive regulation of performance: Interaction of theory and application. Attention and performance. Cambridge, MA: MIT Press, 1999, pp. 435-459.

[7] Harvey, N., Overconfidence in self-assessment of motor skill performance. In J-P Caverni & J-M Fabre (Eds.), *Cognitive biases. Advances in psychology, 68.* Amsterdam, Netherlands: North-Holland,1990, pp. 499-509.

[8] Yarab, Paul E., Sensibaugh, Christine Cregan, Allgeier, Elizabeth Rice. Over-confidence and under-performance: Men's perceived accuracy and actual performance in a course. Psychological Reports, 81(1), 1997, 76-78

[9] Bettenhausen, K.L. Five years of group research: What we have learned and what needs to be addressed. *Journal of Management*, 17, 1991, 345-381.

[10] Hoogstraten, J. Influence of objective measures on self-reports in a retrospective pretest-posttest design. *Journal of Experimental Education*, 53(4), 1985, 207-210.

[11] Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D., & Gerber, S. L. Internal invalidity in pretest-posttest self-report evaluations and a reevaluation of retrospective pretests. *Applied Psychological Measurement*, 3, 1979, 1-23.

[12] Carless, S. A. & DePaola, C. The measurement of cohesion in work teams. *Small Group Research*, 31, 2000, 71-88.

[13] Farh, J-L., Cannella, A. A., Jr., & Bedeian, A. G. Peer ratings: The impact of purpose on rating quality and user acceptance. *Group & Organization Studies*, 16(4), 1991, 367-386..

Appendix A

Sample Items from the Team Knowledge Test

Example of a Team Process Question

Your team leader comes to your scheduled meeting without an agenda. What should you do?

1. *Make your first agenda item developing an agenda as a team.*
2. Let the meeting proceed without an agenda.
3. Tell the team leader to write out an agenda right now and take the rest of the team for coffee until s/he is done.
4. Suggest the meeting be postponed until the team leader gets his act together.

Example of a Decision-Making Question

You know consensus has been reached when

1. five to ten minutes has passed with no objections being raised.
2. a vote reveals that the majority of team members are in favor.
3. *every team member feels that the decision is workable and defensible, even if not what s/he would have chosen on his or her own.*
4. a vote reveals unanimity.

Example of a Communication Question

Effective discussions of team business often get bogged down by people who are argumentative or dominating or ramblers. No matter what their problem, to get the meeting moving forward you need to:

1. *let them know that you have understood and appreciated their point.*
2. make sure the meeting leader is assertive enough to insist that such members be quiet.
3. let them talk. Eventually they will run out of steam and you will still be on good terms.
4. Argue back until they realize that they are wrong.

Example of a Conflict Resolution Question

If a member of your team is hostile or critical it is generally useful to

1. criticize him/her to let that person know how it feels.
2. *try to find some area of agreement or acknowledge some truth in what he/she is saying to diffuse the attack.*
3. threaten to "fire" the individual from the team if he/she does not stop the behavior
4. try to ignore the behavior and push on.

Appendix B

The Team Process Checks Subscales

Students answered each item on a five point scale: Never, Rarely, Sometimes, Frequently, Always

Affiliation

My team may agree on a solution but not every member "buys into" that solution.
When arguments break out, my team members are able to step back, calm down, and work out our differences.
My team members criticize ideas, not each other.
My team ignores conflicts among team members
My team encourages differing opinions to be expressed.
When conflict arises in the team, it is likely to be a battle or, at best, a waste of time.
As a team we find it difficult to accept criticism openly and non-defensively.

Agency

We have a difficult time staying focused and on track.
We are careful to assign tasks to each of the team members when appropriate.
My team tends to start working without an explicit plan.
Some people seem to do the bulk of my team's work.
My team is able to generate potential solutions and evaluate them in an effective and systematic fashion.
18. My team can assess itself and develop strategies to work more effectively. (Agency)