On the Evaluation of Student Team Software Development Projects

Anya Tafliovich
Dept. of Computer and Mathematical Sciences
University of Toronto Scarborough
anya@cs.utoronto.ca

Andrew Petersen
Dept. of Mathematical and Computational Sciences
University of Toronto Mississauga
andrew.petersen@utoronto.ca

Jennifer Campbell
Dept. of Computer Science
University of Toronto
campbell@cs.utoronto.ca

ABSTRACT

Project experience and teamwork have been identified as two of the most important deficiencies of recent graduates, so experiences with team projects are a critical component of computer science and software engineering education. However, evaluating these projects is difficult, as it requires a balance between rewarding the team's effort (and the development of skills that enable the team to work effectively) and recognizing individual contributions. We report on an investigation of the student perspective on evaluation of software development team projects. We find that (a) computer science and software engineering educators hold several misconceptions about students' preferences in teamwork evaluation methods, (b) that students' preferences with regard to evaluation change dramatically as they advance through the program and even in the course of a single term, and (c) that the change in preference occurs early in the term, before they complete any work as a team, with most students shifting their preference towards putting more weight on the team's effort.

Categories and Subject Descriptors

K.3.2 [**Computers and Education**]: Computer and Information Science Education—*Computer Science Education*

General Terms

Human Factors

Keywords

student teamwork; evaluation; students' perspective; motivating students; undergraduate software development project

1. INTRODUCTION

The ability to work effectively in teams is consistently rated highly by employers of Computer Science and Software Engineering graduates. Project experience and teamwork have been identified among the most important knowledge deficiencies of recent graduates, i.e. the most important skills that the prospective employers found were lacking in recent graduates of Computer Science and Software Engineering programs [28, 29]. Unfortunately, the development of teamwork skills is an area poorly understood by the Computer Science and Software Engineering Education community. Educators agree on the importance of having team projects in undergraduate CS and SE courses. However, there is no agreement on best methods of forming and managing student teams and on best policies of evaluating student team projects [26].

Educators agree that an appropriate method of evaluation of team software development projects is essential to ensuring a positive student experience. A method of evaluation that is perceived as unfair can become a significant demotivational factor for students, discouraging team cooperation and communication and hindering the development of a stimulating learning environment.

Anecdotal evidence suggests that educators hold some shared beliefs about the student perspective on evaluating teamwork. For example, the authors encountered the following opinions repeatedly in conversation with colleagues: "Stronger students prefer to be evaluated individually rather than as a team", and even "Most CS students prefer to be evaluated individually rather than as a team". Studying the existing research in the field of Computer Science Education (see Section 2) revealed that very little is known about the students' perspective on teamwork evaluation methods.

In this work, we report on the results of a research study that investigates the student perspective on the evaluation of software development team projects in an undergraduate Computer Science or Software Engineering course setting. We answer the following questions:

1. *What team project evaluation policy do second, third, and fourth year students prefer?*
2. *Which of the factors under consideration, if any, are predictors of a student's initial preference of evaluation policy?*
3. *How do students' preferences change at the end of the term?*
4. *Which of the factors under consideration, if any, are predictors of the change in a student's preference?*
5. *When do the students change their preferences?*

We consider multiple factors as potential predictors of the student's preferences: gender, being in a CS program, English as first language, GPA, expected grade in the course, student's individual performance in the course, the student's

SIGCSE'15, March 4–7, 2015, Kansas City, MO, USA.

ACM 978-1-4503-2966-8/15/03 ...$15.00.
http://dx.doi.org/10.1145/2676723.2677223.

	Team portion (% of the total grade)	Individual portion (% of the total grade)
Option 1	100%	0%
Option 2	80%	20%
Option 3	50%	50%
Option 4	20%	80%
Option 5	0%	100%

Table 1: Survey Question

peer evaluation scores in a recently completed project phase, the student's evaluation of his or her peers in a recently completed project phase, and the grade in this project phase.

The remainder of the paper is organized as follows. Section 2 introduces related studies and work on undergraduate student teamwork. Section 3.1 describes the courses under consideration and the research methodology. Section 4 reports on the results of the study. Section 5 draws conclusions and suggests directions of future research.

2. RELATED WORK

Many groups have looked at team composition and whether attributes of the team members negatively impact the team's overall performance [2, 4, 7, 12, 20, 27, 31]. Faraj et al. found that a team's performance is affected by their communication and coordination styles [10], and Karn and Cowling found that personality issues that lead to disruptions were the most significant factor impacting quality of work [21]. Other groups have found productivity to be linked to trust and to cultural and disciplinary issues between teammates [9] or to social sensitivity [33, 3]. Andreescu et al. reports that successful project teams needed a combination of software development application domain and project management knowledge [1]. A related area of study focuses on tools that support good team formation [6, 23].

Other studies focus on the impact of rewards on teams. Researchers in management cite the need to reward personal achievement to maximize productivity [19]. However, Hoffman et al. argue that a need for cooperation between teammates necessitates more equal sharing of rewards [17]. Similarly, in the classroom, students must balance the need to cooperate with teammates with the urge to demonstrate personal ability (for a better mark). Hazzan's 2003 study asked students to distribute a financial award within a team and observed their behaviour and found that the students tended to divide the majority of the funds equally, reserving a small portion to reward personal achievement [14]. Several groups have worked to teach teamwork skills in the context of software engineering teams and to raise awareness of the tension between collaboration and reward [15, 18, 30].

Self and peer assessments are critical elements in assessment for student team software projects, but some have expressed concerns about the effectiveness of peer assessment [5, 11, 24]. They found evidence of teams colluding to make sure that marks were shared equally – or to punish a single group member [22, 13]. Herbert found that students are both too generous *and* too hard on themselves, suggesting that students' perceptions of effort are not objective [16]. In contrast, Dunsmore et al. notes that peer critiques of contribution were consistent and correlated with perceptions of how much work was completed by team members [8]. Smith and Smarkusky propose a competency based framework for assessment to assist students in making more objective judgements [32].

3. ENVIRONMENT

The study was performed at a large, North American research-intensive university. It included students from four sections of a second-year course, a single-section third year course, and a single-section forth-year course. Each course included a significant team-based software project within a twelve week term.

3.1 Courses

The second-year course (we will call it SE2) introduced students to the programming language Java (the first year is taught in Python), basic Android development, object-oriented analysis and design, simple design patterns, basic version control, and introduction to software modelling. This is the first course in the program that involves significant teamwork, though the students may opt to work with a partner in first year courses. The course project spanned seven weeks (roughly the second half of the term) and consisted of three distinct stages. The end product of the project was a simple Android application. The first stage was the design stage, done in groups of two. The second and third stages involved revisiting the design and developing the application; these were done in groups of four. The teams followed a prescribed, simplified, Scrum-like software development process, closely monitored by the course teaching assistants. Between four sections taught on two campuses, the peak enrolment in the course was 575 students.

The third-year course (SE3) is an introduction to software engineering. In this course the students learn about software processes, with heavy emphasis on Agile software development techniques. The course also covers aspects of software design and software architecture. Since the second year course has not always included a team component in past years, the third year course might also be the first time that a student is involved in a significant team project. The course project spanned 10 weeks of the term and consisted of four distinct stages. The students worked in teams of four or five and developed either a sophisticated Android or web application. The teams followed an extreme programming (XP) software development process under close supervision of the teaching assistants and course instructor. The peak enrolment in the course was 84 students.

The fourth-year course (SE4) is an advanced software engineering course. The students worked on a large (200K+ lines of code) open source project, in teams of six or seven. In the course of the project, the students performed a reverse engineering and analysis of the code base, completed at least two bug fixes, and investigated, designed, and implemented a new feature. The project spanned 10 weeks of the course and consisted of four stages. The teams chose their

Course	Gender		CS major		English First Lang.		GPA				Expected grade			
	Male	Female	Yes	No	Yes	No	< 1.5	1.5-2.5	2.5-3.5	> 3.5	A	B	C	Don't Care
SE2	256	82	291	48	159	176	3	45	71	21	229	95	3	9
SE3	41	4	43	2	24	21	1	16	22	6	27	14	3	1
SE4	38	3	41	0	14	26	0	14	20	6	19	12	3	7

Table 2: Descriptive Statistics

own software development process. The peak enrolment in the course was 63 students.

3.2 Evaluation Preferences

At the beginning of the term, during lecture, the following information was collected: (a) whether the student is majoring in CS, (b) gender, (c) GPA, (d) whether English is the student's first language, and (e) expected grade in the course. At the same time, students were asked to provide informed consent. This information was provided to the investigators at the end of the term. The investigators also received access to all consenting students' term and final grades, as well as peer evaluation data collected during the term. The students were also asked the following question:

> Assume that you are a member of a student team working on a course project that involves developing software. Your project grade will consist of two portions: Team and Individual. In the Team portion of the grade, each team member *receives the same grade*. In the Individual portion of the grade, each team member's grade is allocated *based on his/her individual performance* on the project. The table below outlines five grading options:
> *[See Table 1 for the table in the survey.]*
> Which grading option would you prefer?

In SE2, the students answered the above question three times: (1) before the project teams were formed, (2) after the teams were formed but before they completed a project stage, and (3) at the end of the term. In SE3 and SE4, the students answered the questions twice: (1) at the beginning of the term before teams were formed and (2) at the end of the term.

Each time, the students were informed that their survey responses were made anonymous for the purposes of the study, and that their responses had no impact on the actual method of project evaluation performed in the course.

3.3 Team Formation and Peer Evaluation

In SE2, the students formed teams of two for the first stage of the project. The course instructors then formed teams of four for the second and third stages of the project by merging the pairs from the first stage.

In SE3 and SE4, we allowed students to request a single partner for the team. The course instructors then formed teams and, wherever possible, put the requested partners on the same team.

We used CATME team formation and peer evaluation software to form the teams in all three courses [25]. The team formation criteria included personal schedules (to maximize time available for group meetings), gender (females are not outnumbered in a team) [1], GPA (to group students with dissimilar GPAs), and commitment level (to group students with similar self-reported commitment).

In all three courses we also used CATME to perform peer evaluation, following completion of each stage of the project. In SE2, the students ranked themselves and their teammates, on a scale of 1 to 5, on (1) contributing to the team's work, and (2) interacting with teammates. In SE3 and SE4, there were additional evaluation criteria: (3) expecting quality and (4) having related knowledge, skills, and abilities. The students were also invited to provide comments to the instructor. Based on these criteria, the CATME software calculates an *adjustment factor* for each student. In SE2, the individual grade for a project stage was calculated as $0.8 \times$ *team grade* $+ 0.2 \times$ *adjustment factor* $\times$ *team grade*, with the adjustment factor capped at 1.05. In SE3 and SE4, the formula was *adjustment factor* $\times$ *team grade*, where the adjustment factor was capped between 0.8 and 1.05.

[1]This condition could not be met in SE4, which only had three female students enrolled.

4. RESULTS

A total of 430 participants provided consent for the study: 342 from SE2 (59.5%), 46 from SE3 (54.8%), and 42 from SE4 (66.7%). Table 2 summarizes the descriptive statistics of the participants' responses to the survey questions (omitting blanks).

4.1 Team Evaluation Preferences: Before

Figure 1 shows the students' team evaluation preferences as reported before the student project teams were formed.

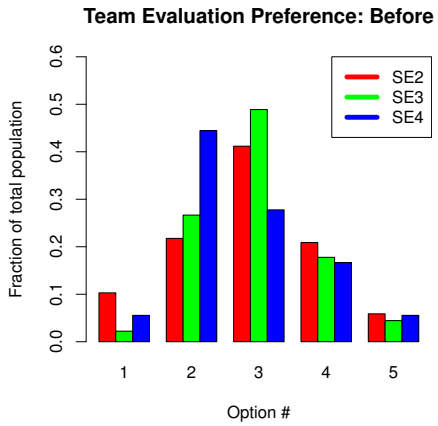
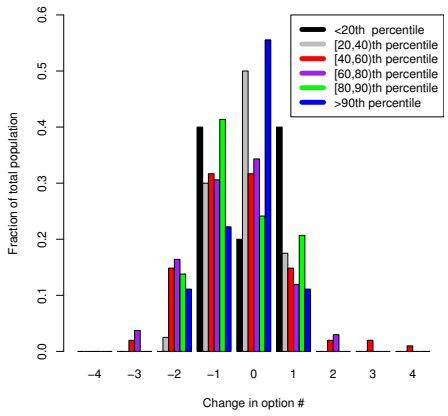


Figure 1: Team Evaluation Preference: Before

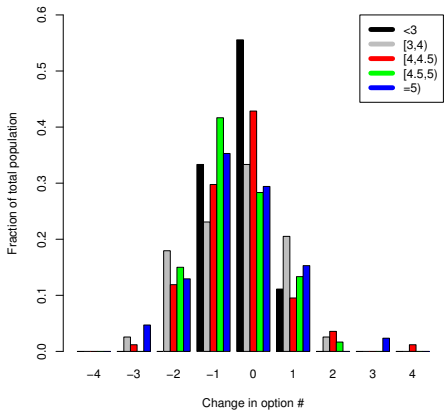
Recall that Option #1 is the 0-100 split (0% individual, 100% team) and Option #5 is the 100-0 split (100% individual, 0% team). In the second year course the students' answers form a bell curve around the 50-50 option, but by the fourth year course, the students favour the 20-80 split.

We then asked the question, *"Which of the factors under consideration, if any, are predictors of a student's initial preference of evaluation policy?"*

We used the statistical software package R to run linear regression tests to see whether the following factors are predictors of the student's initial preferred team evaluation



(a) By Individual Grade



(b) By Team Satisfaction Score

Figure 2: No Effect on Preferences: Examples

policy: gender, being in a CS program, English as first language, GPA, and expected grade in the course. We found that none of these factors were predictors.[2]

4.2 Team Evaluation Preferences: After

Figure 3 shows the students' team evaluation preferences as reported at the end of the term.

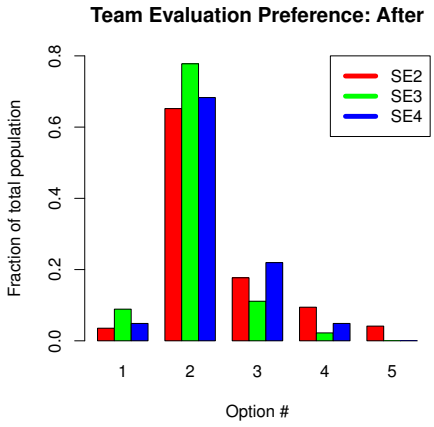


Figure 3: Team Evaluation Preference: After

The students' answers are dramatically different after the course, with the vast majority of students in all three courses preferring the 20-80 split.

Linear regression tests show that, as with the initial preference, gender, being in a CS program, English as first language, GPA, and expected grade in the course are not predictors of the student's preference as recorded at the end of the term.

In addition, we used the same approach to investigate whether the student's individual performance in the course, the student's peer evaluation scores in a recently completed project phase, the student's evaluation of his or her peers in a recently completed project phase, or the grade in this phase are predictors of the team evaluation option. Surprisingly, we found that, in general, these factors were not predictors of the student's preference. In SE4, there was a slight positive correlation between a preference toward higher weight for the individual component and (1) a student's individual grade ($R^2 = 0.19$, $p = 0.007$), (2) the grade in the last project phase ($R^2 = 0.10$, $p = 0.038$), and (3) team satisfaction in a recently completed project phase ($R^2 = 0.17$, $p = 0.005$). However, the smaller sample size in SE4, lack of evidence in the other two courses, and the low values of R^2 suggest that we cannot draw any conclusions from these results.

4.3 Team Evaluation Preferences: Change

The next question we address is, *"How do the students change their preference at the end of the term?"* Figure 4 shows the students' change in team evaluation preferences at the end of the term.

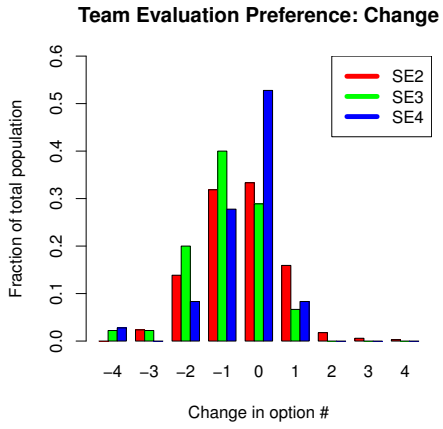
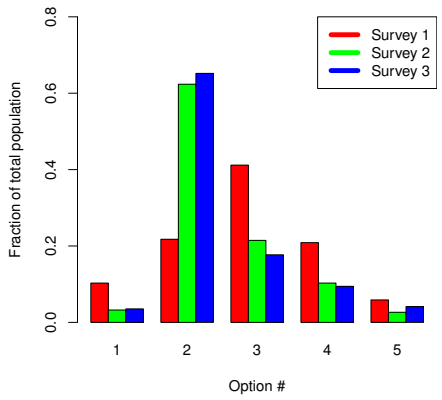


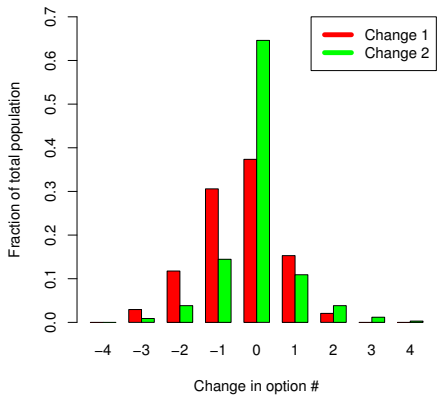
Figure 4: Team Evaluation Preference: Change

We observe that the students in all three courses tend to change their team evaluation preference by increasing the team portion. This effect is more significant in the sec-

[2]The full results of the linear regression analysis are available at `http://www.utsc.utoronto.ca/~atafliovich/`.



(a) Preferences in SE2: the three surveys



(b) Preferences in SE2: the two changes

Figure 5: Timing of the Change in Preference

ond and third year courses than in the fourth year course (where students were in at least their second significant team project and where there was already a strong preference for a 20-80 split at the beginning of the term).

We then asked the question, *"Which of the factors under consideration, if any, are predictors of the change in students' preferences of the team evaluation policy?"*

Again, linear regression tests show that gender, being in a CS program, English as first language, GPA, expected grade in the course, individual performance in the course, the student's peer evaluation scores in a recently completed project phase, the student's evaluation of his or her peers in a recently completed project phase, or the grade in this project phase are not predictors of the student's change in team evaluation preference at the end of the term.

Anecdotal evidence suggests that educators are frequently surprised that the student's level of satisfaction with his or her teammates' performance is *not* positively correlated with opting for a higher team portion, and neither is it the case that stronger students (students with higher individual grades) prefer an evaluation option with a higher individual portion. Figures 2a and 2b show example distributions of students' change in preferences based on the individual term grade and on the students' ratings of his or her teammates in a recently completed project phase. The distributions for each bin are normally distributed around the mean of the original distribution, confirming that these factors do not affect the student's preference toward evaluation.

Another interesting question is, *"When do the students change their preferences?"* Recall that in SE2 the first answer was recorded before the project teams were formed, the second answer was recorded after the teams were formed but before the students completed a project stage, and the third answer was recorded at the end of the term. Figure 5a provides a histogram of the students' answers in the first, second, and third survey in SE2. We observe that the biggest change occurs between the first and the second surveys, with only a minor further change at the last survey. This tells us that the students' preferences change and settle very soon after they know their teammates, before they had a chance to perform a significant piece of work together or get any feedback on their teamwork. Figure 5b, which shows the distributions of the changes in team evaluation preference that occur between the first and second and the second and third surveys in SE2, confirms this observation.

5. CONCLUSIONS

Our data analysis demonstrates that the student's preference for the method of evaluation in a software development team project depend remarkably on two factors: the student's year of study and the point in the term when the preference is expressed. First, there is a significant difference of opinions between students in their second, third, and fourth year. We found that, when asked in the beginning of the term, in the second year course the students' answers form a bell curve around the 50-50 option (50% for individual contribution and 50% for team grade), but by the fourth year course, the students favour the 20-80 split (20% for individual contribution and 80% for team grade). Second, in all three courses, the students tend to change their preference during the term, to favour a team option more strongly, and this effect is more significant in the second and third year courses than in the fourth year course. Third, we found that the change in the students' preferences occurs very early in the course of the project: before completion of a team project phase, and before receiving any feedback or grade from the TA or instructor.

Additionally, we examined a number of factors that we conjectured may predict the student's initial team evaluation preference or the change in preference during the term. Surprisingly, and contrary to what anecdotal evidence suggests is a commonly shared belief among educators, none of the following factors are predictors: gender, whether the student is majoring in CS, GPA, whether English is the student's first language, expected grade in the course, individual performance in the course, the student's peer evaluation scores in a recently completed project phase, the student's evaluation of his or her peers in a recently completed project phase, or the grade in this project phase.

Our results suggest that the experience of being a member of a development team changes the students' perception of what constitutes fair evaluation of teamwork. Surprisingly, the change is not affected by the team's performance,

nor by the individual team members' contributions. Moreover, even a brief exposure to teamwork is enough to cause a significant change in evaluation preference, with students opting to increase the team's portion of the grade.

An important question that our study does not answer is *why* the students change their team evaluation preference. The two possible explanations we consider are decidedly different. It is imaginable that getting to know the team members increases the student's trust in the team, and creates an atmosphere of support and team spirit. On the other hand, one can imagine that students realize the difficulty in assessing the individual contribution, and thus do not trust its results, whether the assessment is performed via peer evaluations or via the TA's or instructor's observations.

As part of future research in this direction, we plan to conduct student interviews with the aim of better understanding the students' reasoning behind the shift in team evaluation preferences. Gaining a better understanding of the students' perspective on peer and supervisor's evaluation will also contribute to this line of research. Since our current results indicate that educators hold several misconceptions about students' preferences in teamwork evaluation, we identify a lack of available research data that would help explain these preferences, and the need for such data.

6. REFERENCES

[1] A. Andreescu, I. Intorsureanu, A. Uta, and R. Mihalca. Team work in software development student projects. In *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, pages 80:V.18–80:1, 2008.

[2] M. Barrick, G. Stewart, M. Neubert, and M. Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83(3):377–391, 1998.

[3] L. Bender, G. Walia, K. Kambhampaty, K. E. Nygard, and T. E. Nygard. Social sensitivity and classroom team projects: An empirical investigation. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, pages 403–408, 2012.

[4] G. Beranek, W. Zuser, and T. Grechenig. Functional group roles in software engineering teams. *SIGSOFT Softw. Eng. Notes*, 30(4):1–7, 2005.

[5] N. Clark, P. Davies, and R. Skeers. Self and peer assessment in software engineering projects. In *Proceedings of the 7th Australasian Conference on Computing Education - Volume 42*, pages 91–100, 2005.

[6] M. Craig, D. Horton, and F. Pitt. Forming reasonably optimal groups: (frog). In *Proceedings of the 16th ACM International Conference on Supporting Group Work*, pages 141–150, 2010.

[7] Y. Dubinsky and O. Hazzan. Using a role scheme to derive software project metrics. *Journal of Systems Architecture*, 52(11):693–699, 2006.

[8] H. E. Dunsmore, D. P. Moffet, and S. T. Ward. Software engineering team project group member evaluations: Some empirical results. *SIGCSE Bull.*, 21(2):40–45, 1989.

[9] K. Egea, S.-K. Kim, T. Andrews, and K. Behrens. Approaches used by cross-cultural and cross-discipline students in teamwork for a first-year course in web design. In *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103*, pages 87–96, 2010.

[10] S. Faraj and L. Sproull. Coordinating expertise in software development teams. *Management Science*, 46(12):1554–1568, 2000.

[11] A. Gates, N. Delgado, and O. Mondragon. A structured approach for managing a practical software engineering course. In *30th Annual Frontiers in Education Conference*, 2000.

[12] N. Gorla and Y. W. Lam. Who should work with whom?: Building effective software project teams. *Commun. ACM*, 47(6):79–82, 2004.

[13] J. H. Hayes, T. C. Lethbridge, and D. Port. Evaluating individual contribution toward group software engineering projects. In *Proceedings of the 25th International Conference on Software Engineering*, pages 622–627, 2003.

[14] O. Hazzan. Computer science students' conception of the relationship between reward (grade) and cooperation. *SIGCSE Bull.*, 35(3):178–182, 2003.

[15] O. Hazzan and Y. Dubinsky. Students' cooperation in teamwork: binding the individual and the team interests. In *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, pages 151–152, 2010.

[16] N. Herbert. Quantitative peer assessment: Can students be objective? In *Proceedings of the Ninth Australasian Conference on Computing Education - Volume 66*, pages 63–71, 2007.

[17] J. R. Hoffman and S. G. Rogelberg. A guide to team incentive system. *Team Performance Management: An International Journal*, 4(1):23–32, 1998.

[18] J. M. Hogan and R. Thomas. Developing the software engineering team. In *Proceedings of the 7th Australasian Conference on Computing Education - Volume 42*, pages 203–210, 2005.

[19] D. Ian Brown. TeamâĂŘbased reward plans. *Team Performance Management: An International Journal*, 1(1):23–31, 1995.

[20] M. Ikonen and J. Kurhila. Discovering high-impact success factors in capstone software projects. In *Proceedings of the 10th ACM Conference on SIG-information Technology Education*, pages 235–244, 2009.

[21] J. Karn and T. Cowling. A follow up study of the effect of personality on the performance of software engineering teams. In *Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering*, pages 232–241, 2006.

[22] D. Kaufmann, R. Felder, and H. Fuller. Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education*, 89:133–140, 2000.

[23] R. Layton, M. Loughry, M. W. Ohland, and G. D. Ricco. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education*, 2, 2010.

[24] R. Layton and M. Ohland. Peer ratings revisited: Focus on teamwork, not ability. In *Proceedings of American Society for Engineering Education Annual Conference*, 2001.

[25] R. A. Layton, M. L. Loughry, M. W. Ohland, and G. D. Ricco. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education*, 2(1):1–28, 2010.

[26] R. Lingard and E. Berd. Teaching teamwork skills in software engineering based on an understanding of factors affecting group performance. *Frontiers in Education*, 3, 2002.

[27] V. Pieterse, D. G. Kourie, and I. P. Sonnekus. Software engineering team diversity and performance. In *Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, pages 180–186, 2006.

[28] A. Radermacher and G. Walia. Gaps between industry expectations and the abilities of graduates. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, SIGCSE '13, pages 525–530, 2013.

[29] A. Radermacher, G. Walia, and D. Knudson. Investigating the skill gap between graduating students and industry expectations. In *Companion Proceedings of the 36th International Conference on Software Engineering*, ICSE Companion 2014, pages 291–300, 2014.

[30] A. Radermacher, G. Walia, and D. Knudson. Investigating the skill gap between graduating students and industry expectations. In *Companion Proceedings of the 36th International Conference on Software Engineering*, pages 291–300, 2014.

[31] S.-T. Shen, S. Prior, A. White, and M. Karamanoglu. Using personality type differences to form engineering design teams. *Engineering Education*, 2:54–66, 2007.

[32] H. H. Smith, III and D. L. Smarkusky. Competency matrices for peer assessment of individuals in team projects. In *Proceedings of the 6th Conference on Information Technology Education*, pages 155–162, 2005.

[33] A. Woolley, C. Chabris, A. Pentland, N. Hashmi, and M. Thomas. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–688, 2010.