AC 2009-1603: AN ASSESSMENT STRATEGY FOR A CAPSTONE COURSE IN SOFTWARE AND COMPUTER ENGINEERING

Richard Stansbury, Embry-Riddle Aeronautical University

Massood Towhidnejad, Embry-Riddle Aeronautical University

AN ASSESSMENT STRATEGY FOR A CAPSTONE COURSE IN SOFTWARE AND COMPUTER ENGINEERING

Richard Stansbury and Massood Towhidnejad
Embry-Riddle Aeronautical University
Daytona Beach, FL 32114

{stansbur, towhid}@erau.edu

Abstract:

The assessment of individual student work on team projects is challenging. Ideally, every student would work toward the project goals with an equal level of effort resulting in all students in the course sharing the same final grade for the project. Unfortunately, this is not realistic. In this paper, a number of approaches to student assessment including peer and leader evaluation, ranking systems, etc. are discussed. Based on some of these ideas and the goal of providing students with frequent feedback, a new assessment approach for the computer and software engineering senior design course at Embry-Riddle Aeronautical University is presented, which utilizes rubrics to judge each student's personal contribution and frequent delivery of the assessment so that students may iteratively improve their performance. This approach is evaluated versus the previous, a more highly subjective, assessment approach for this course. The paper concludes with a discussion of how this process will be implemented for later years.

Introduction:

At Embry-Riddle Aeronautical University (ERAU), Daytona Beach campus, two courses, CEC 420/421: Computer System Design Lab and SE 450/451: Software Team Project, comprise the two halves of a single year-long capstone course provided by the Department of Computer and Software Engineering. There is typically a significantly higher number of software engineering students versus computer engineering students. The course also invites participants from other departments such as human factors in order to encourage a multidisciplinary teaming environment.

From past experience, it has been found that assessment of individual student performance on a large group project is difficult. Ideally, every student would work toward the project goals with an equal level of effort resulting in all students in the course sharing the same final grade for the project. Unfortunately, this is not realistic. To distinguish between the individual contribution of each student, typically, self evaluation, peer evaluation, or team lead evaluation are used. However, this approach is also flawed. For instance, one major problem is that students tend to normalize their peer evaluations, which results in team members evaluating their peers above average (unless there is a student that has not contributed at all). For self evaluations, students typically provide themselves with a higher evaluation than is likely deserved.

To remedy these problems, the instructor may introduce a subjective evaluation for each student based on the instructor's observations. Typically, there is not a formal representation of this subjective component to the grade. As a result, the grading criteria and expectations are never

explicitly defined for the students. This strategy has been used for several years for team oriented projects within our department. This paper proposes a new assessment strategy applied to the 2008-2009 academic year to assess student performance on the team project individually. This strategy is used for both determining a final grade and providing frequent feedback throughout the semester in order to motivate each student toward personal and team success.

At ERAU, each student's grade in the senior design capstone course is based on their performance on an individual research tasks and the overall project grade. For this new approach, each student's project grade is based on a grade given to the team for the set of final deliverables multiplied by the student's contribution factor. The contribution factor is based on their level of effort exhibited, willingness to support others, and overall attitude. A rubric defines the numeric contribution factor awarded given the expectations of the instructor. For students that exceed expectations, the contribution factor shall increase their project grade as a bonus. For students that fail to meet expectations, the contribution factor shall decrease their project grade as a penalty.

To keep the students motivated and aware of their assessed performance, monthly assessment reports are provided. Afterward, the instructor and student have the opportunity to discuss the student's current performance with respect to the rubric. Since a final grade will be only given at the end of the semester, these frequent updates shall help motivate students toward meeting or exceeding the instructor's expectations.

In this paper, a discussion of current issues with student assessment in capstone courses will be provided. Related assessment approaches shall be reported. The authors' new strategy for assessment shall be presented in greater detail. The qualitative and pending quantitative analysis of this new strategy versus past strategies shall be presented and/or discussed. Finally, the paper will conclude with a discussion of improvements that may be made and a brief discussion regarding how the techniques may be applied to other software and computer engineering capstone courses.

Related Work:

Assessment of individual student performance for team projects has interested educators for many years, and is a problem that is not exclusive to software team projects. Student assessment has also not exclusively focused upon grading. It has been applied toward determining if a course is delivering the educational outcomes promised to its students and as a mechanism to provide feedback to the students so that they may improve as discussed in McGourty et al.[1]. In this section, a survey of papers related to the assessment of individual students participating in team projects is presented.

In Kaufman et al.[2, 3], the authors present an assessment technique for evaluating individuals in cooperative learning teams based on a modified version of Autorating, as defined by Brown[4]. For this technique, students were asked to assess their peers using a pre-defined set of linguistic terms for a variety of criteria. A numeric value is assigned to each linguistic term. Each student's contribution factor is then calculated as their rating average divided by the team rating average. The authors' research sought to address the validity of Brown's technique by

performing correlation analysis with respect to a number of concerns such as the impact of peer evaluation on women and minority students. It also discusses the ability to diagnosis issues with teams such as identifying students that are not pulling their own weight, students that are acting as tutors to the team members, and those teams that are dysfunctional. The results showed that the technique was fairly successful, but did yield some results that warrant further investigation with regards to minority ratings being lower by a statistically significant level.

Wilkins and Lawhead[5] examine a variety of data gathering techniques to be applied for team member evaluation. The discussion focuses upon how peer evaluation forms may be developed to assess student performance including scales, rankings, matching, fill-in-the-blank questions, short answers, and journals. Scales allow the students to indicate a score from negative to positive where they feel their team or a peer falls with regard to specific criteria. Ranking may be used to rank the members of the team based on contribution. It was suggested that rather than use explicit ranking that exercises such as asking each team member to divide some amount of money amongst their team members based on their belief of what is "fair". Matching may be used such as asking the student to match three strengths to each team member and three weaknesses to each team member from a set of evaluation terms. Alternatively, given a set of job roles, the students could match teammates to each of the roles. Fill-in-the-blank questionnaires allow the students to provide their own feedback using the sentence structure to prompt them for their own opinion regarding their teammates. Lastly, short answer and journal writing/essays provide an opportunity for obtaining specific details, but are more difficult to prompt the student for evaluation content and require greater effort by the grader.

Hayes et al.[6] surveys current practices at a number of universities. For each, the assessment techniques vary from a small reported set of assessment schemes: assessing all students with the same grade; assessing students individually based on self-reports, peer evaluations, and self evaluations; quizzes and exams only; and cross-validating team grade with individual performance on exams, quizzes, and non-group assignments. Each technique is assessed versus a set of criteria believed essential for individual assessment: fairness, consistency, accurate reflection of educational objectives, provides feedback, encouraging to students, avoids grade inflation, and ease of grading. The authors conclude that none of the grading schemes meets all of these criteria. However, a combination of techniques may be fused to produce an evaluation process so long as it is explicitly defined to the students in the syllabus.

Clark et al.[7] presents a grading technique used for a capstone software engineering course. To assess individual performance, several products were examined including timesheets, self/peer evaluations (using a variety of the techniques presented in Wilkins and Lawhead[5]), individual contribution reports, and a quantity report (similar to the monetary ranking discussed in Wilkins and Lawhead[5] in which $100 is divided into shares to each team member). Given these assessment products, the authors use the various data to validate the assessment for consistency, and if consistent, the monetary amount assigned to each teammate is used as the percent contribution by the students. A student's individual mark may be computed given the team's performance and their contribution. The individual mark is then summed with the team mark to produce a final grade. The techniques were developed, refined, and evaluated over a period from 1998 to 2004.

In a 1984 publication, Sanders[8] addresses individual assessment for a senior-level directed project course in which students were divided into teams, and each team partnered with a real customer to develop a software application. Given the varied background of the students, the compositions of the teams, and the difference between projects, a team and project neutral evaluation scheme was required to assess each student's performance. Peer evaluations composed 10% of the grades. Checklists were used to evaluate student performance during group or customer meetings for 30% of the grade. The remaining 60% of the grade was derived from evaluating the status of the project at various stages throughout the semester. As an incentive for strong performance within the team setting, the instructor reserved the right to provide up to five bonus points and five penalty points.

As evident from the previous work, there is no single best case approach for evaluating individual student performance within a team project. Researchers have attempted to combine a number of different techniques in order to derive the fairest grade representing the student's contribution toward the success or failure of the project. Similar to the observational approaches discussed by Sanders[8], the remainder of this paper introduces another approach to including faculty subjective observation of individual personal contribution in order to derive an overall grade for each student.

Approach:

For the senior capstone course, a common customer-oriented project assigned to all students. In past years, the project has included unmanned systems development. For 2008-2009, the students are contributing to the development of an intelligent hybrid control system for the Saturn Vue as part of the General Motors and US Department of Energy funded EcoCar Challenge[9]. The students are divided into teams with respect to the major functionalities to be delivered. The first semester focuses primarily upon the first half of the software lifecycle (requirements and design), but there is some opportunity for development toward the end of the fall semester. The spring semester is composed entirely of system development utilizing the Crystal Clear[10] agile process. Under this process, the students must deliver a functional and testable product every two weeks for demonstration and usability.

There are two major components that contribute to the overall assessment of the student's performance, Team Project Grade, and individual Research Grade. Student's grade is calculated with the following formula.

Student Grade = (0.75 * Team Project Grade) * Personal Contribution Factor + 0.25 * Research Grade

Where Personal Contribution Factor = (Effort + Support + Attitude) / 3

The *team project grade* is based on an assessment of deliverables and milestones throughout the semester. Table 1 presents the point distribution for the *team project grade*.

The *personal contribution factor (PCF)* acts as a multiplier upon the team project grade, which may result in either an increase or decrease in points toward the student's final grade. The PCF is computed as the mean of three subfactors: level of effort, support of others, and attitude. For each, the student is evaluated using the following linguistic descriptors and their respective

numeric weight: Poor (0.5), Moderate (0.75), Good (1.0), or Excellent (1.25). Therefore, if they perform above and beyond the call of duty, they are rewarded. However, if their contribution is moderate or lower, they will not receive full credit for the team project grade. For example, if a student is evaluated as having a effort=Good, support=Moderate, attitude=Good, their *PCF=(1 + 0.75 +1)/3=0.9167*.

Table 1: Team grade point distribution.

Team Grade	75
Team organization	10
Chemistry	5
Dynamic	5
Project Plan	10
Completeness	5
Accuracy	3
On time delivery	2
Requirement	12
Acquisition	3
Completeness	4
Quality	5
Design	12
Completeness	6
Quality	6
Quality	13
Test/quality plan	3
Process document	2
Inspection	2
Test case	2
Test execution	2
Traceability metrix	2
Social/economical/ethic	3
Completeness	2
Quality	1
Operation/Demo (HW)	8
Overall quality	3
Overall completeness	5
Presentation	7
Quality	7

Evaluation of the personal contribution is subjective. However, it is influenced by a number of factors and artifacts produced by the students including personal time logs and self/peer evaluations. In addition to these factors, the instructor and teaching assistants' observations are also considered. Given this information, the following rubrics are applied to derive each of the individual contribution factors.

Level of effort's rubric as shown in **Error! Reference source not found.** explicitly tries to capture and reward those that put more time actually working in or out of class than their peers. It also penalizes those students that are simply not pulling their weight, or being an outright distraction to others.

Support of others is explicitly defined to capture team member's ability to be "team players". Its rubric is presented in Table 3. By supporting others, the students demonstrate the understanding that this is a group project and that they cannot be obliged to only look out for themselves and still expect to succeed professionally.

Attitude is the most subjective of the personal contribution factors and requires the greatest effort of the instructor to assess. Its rubric is presented in Table 4. Peer evaluations may also act as

indicators of team dysfunction, which may be the result of personality clashes. This contribution also tries to reward those students that help motivate others toward success.

Table 2: Rubric for assessment of student level of effort.

Level of Effort	Possible Characteristics
Poor (0.5)	Student is riding off of the work from others. When in class, the student is seen frequently doing tasks other than working and causing a distraction to others.
Moderate (0.75)	Student is willing to work during class time and some other times. Student may sometimes lack focus while others are working. However, student may be periodically absent. Student focuses on doing only what it takes to get his job done.
Good (1.0)	Student is considered a good workhorse that is always willing to put in the hours if necessary. Upon completion of a task, the student will begin work on the next without being told to do so.
Excellent (1.25)	Student works above and beyond the call of duty. The student is an integral part of the success for his team and the entire class. Student is self motivated and willing to find work if nothing appears on the current iteration's "to do" list.

Table 3: Rubric for assessing student support of others.

Support of Others	Possible Characteristics
Poor (0.5)	Student has an "I" attitude toward the project. Student focuses solely on his contribution to the project with little concern toward the project as a whole. Student may leave class early, skip meetings if his/her portion of the project is done.
Moderate (0.75)	Student will work to assist members on own team in order to contribute to its overall success. Student rarely volunteers or contributes to tasks that will benefit the overall class.
Good (1.0)	Student works toward supporting own team and contributes to one or more task that supports the class.
Excellent (1.25)	Student works toward supporting team and entire class. Student is seen as taking an overall leadership role toward completing several support tasks

Ideally, students would be self aware of their personal contribution based on the state of the project, interaction with their peers, and interaction with the instructors. However, often students inflate their presumed level of contribution. In order to keep students level headed and to avoid surprises at the end of the semester, students receive performance evaluation reports with their personal contribution scores periodically. This feedback should motivate students by allowing them to know exactly where they must improve.

The *Individual Research Score* is the last component of each student's grade. Table 5 presents the points distribution. These points are assigned based upon the quality of research reports or presentations provided by the students.

Table 4: Rubric for assessing student attitude.

Attitude	Possible Characteristics
Poor (0.5)	Student demonstrates little interest in succeeding in the final project. Student is frequently late or absent to classes or meetings. Student may frequently blame other team members or other teams for difficulties in parts of the project. Student frequently lowers team or group morale. Student may be the frequent cause of personality clashes.
Moderate (0.75)	Student demonstrates some interest in his/her portion of the project. Student is willing to work toward solving problems based on the motivation of a grade. Student may occasionally lower team or class morale. Student may periodically contribute to personality clashes, but generally they are resolved without incident
Good (1.0)	Student demonstrates an interest in the project and a willingness to work. Student is amicable with classmates. Student rarely is involved with personality clashes. Student helps maintain the overall team or class morale.
Excellent (1.25)	Student demonstrates a positive attitude toward the project and works to keep teammates and peers motivated. Student is willing to put in long hours without complaint. Student will work to help boost the morale of team or class.

Table 5: Individual research contribution point distribution

Individual Research	25
Willing to learn	5
Quality of work	8
Creativity	4
Difficulty	8

Evaluation:

One qualitative assessment of this new approach is to determine the acceptability and perceived motivation of the new assessment procedure from the 2008-2009 senior design students. These students have experience with the previous assessment strategy from a number of courses earlier in the undergraduate curriculum.

Students from previous semesters of senior design have never been formally surveyed. However, there has been plenty of feedback. Most often, a student has queried as to why *they*

received a lower grade even though *their team* succeeded in achieving all tasks goals. Generally, a discussion about how individual grading was assessed based on personal contribution often resulted in further argument.

The new assessment procedure was implemented partly to provide transparency regarding the assessment of personal contribution to the students and provide a clear cut equation for deriving the final grade given that personal contribution. After having gone through two full cycles of the assessment process and the Fall semester's final grading, a questionnaire was presented to the students. To ensure that the student would not worry about giving an honest opinion an online survey tool was used that presented the surveyor with only the aggregate tally for each question.

The survey was comprised of the following statements. For each, the student was asked to indicate if their feelings toward the statement are: strongly agree, agree, neutral, disagree, or strongly disagree.

1. Compared with previous team projects (e.g. SE 300), I feel that my personal contribution to the project is better assessed under this new assessment approach.
2. The rubrics allow me to better understanding how I was evaluated for each of the personal contribution factors (effort, support, and attitude).
3. I feel that I am being fairly assessed under the new rubrics for personal contribution.
4. Receiving periodic feedback regarding my personal contribution score helps motivate me by letting me know where I may improve my performance.
5. The new approach to personal contribution assessment should be used in future semesters of CEC 420 and SE 450.
6. Other courses should adopt this technique for assessing personal contributions to our software and/or hardware team projects.

There is an additional comment section that is available to the respondent to provide further feedback.

During the Fall 2008 semester, the assessment process for personal contribution was utilized once the implementation of the software system had begun. Two cycles of assessment were conducted. The first was strictly used to provide student with feedback after one month of work during the team-based implementation phase. The second was used for assigning the student's grades. In the results section, this paper shall present the minimum, maximum, mean, and standard deviation of personal contribution factors. Some insights and concerns are also presented.

Quantitative assessment of this approach is planned. Upon completion of the project, an analysis of project artifacts from the past three years of senior design will be conducted. During this analysis, a number of software engineering metrics (man-hours, lines-of-code, defects logged, etc.) will be used to provide a comparison regarding the quality of the product and the productivity of the students. These results will be presented during the conference presentation.

Results:

Of 15 students, only 13 responded to the questionnaire. The results of the survey are shown in Figure 1. From these results, only one student indicated disagreement for any of the statements, which was statement three regarding the fairness of their assessment given the rubric. This result is somewhat understandable as students are likely to defend their performance when assessed lower than expected. For all other categories, the majority of students responded favorably with at least 70% reporting agrees or strongly agrees for all questions.

Some additional student feedback was received. Three felt that the categories were too broad and would prefer that the categories and rubrics respectively be divided up into further subcategories. Another student requested specific checklists in addition to the rubric to ensure that they touch all of the expected activities. Both of these comments will be considered for future implementations. In addition, number of students provided positive comments regarding the added transparency associated with this new approach.

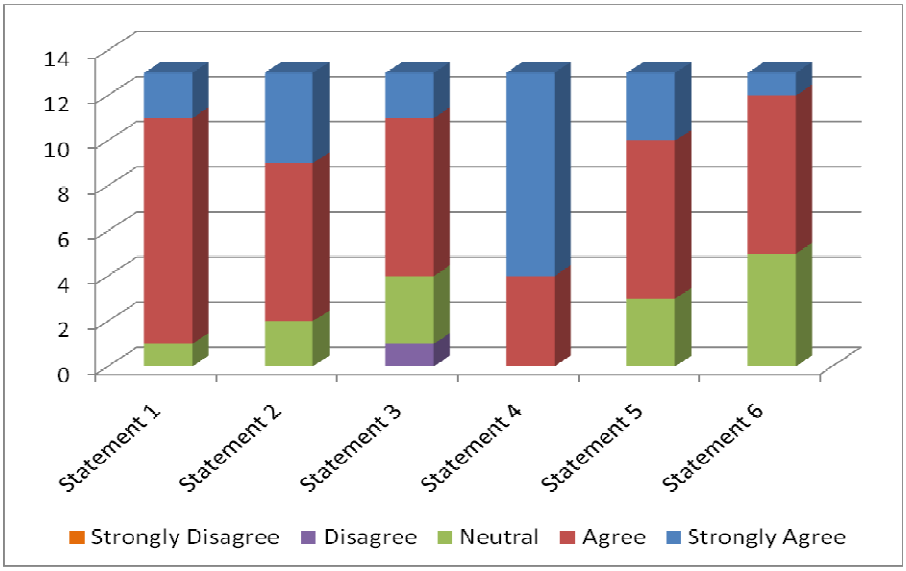

Figure 1: Responses to Survey Questions.

From the first assessment period, one concern regarding the assessment approach was identified. Given the current personal contribution weights assigned to poor, moderate, good, and excellent respectively, there is a strong bias toward negatively assessing the student. It has always been our expectation that students should strive to consistently meet a good performance level. Several students during the first evaluation period fell into moderate for two or more contribution factors. As a result, it was realized that they would likely be too negatively weighted. Therefore, two modifications to the approach shall be conducted during the 2009-2010 school year. First, as suggested by several of the students, each contribution shall be broken into subcategories such that their contribution factor may be distributed across additional factors such that negative assessments under one subcategory will not have as significant of a bias. Second, the weight factors may be shifted upward for both poor and moderate.

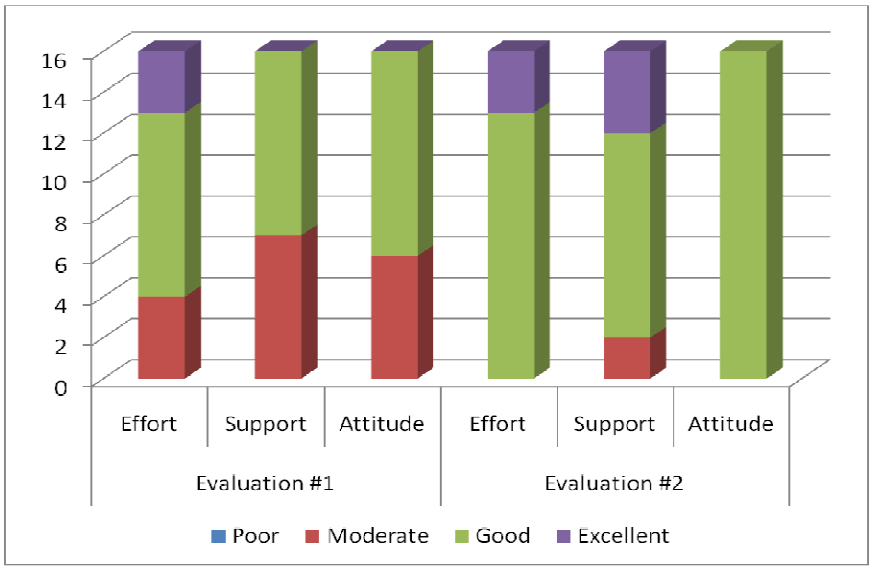

Figure 2: Student assessment totals from first and second assessment periods.

Conclusions:

The new assessment approach as discussed within this paper has shown some initial success under implementation for the 2008-2009 academic year. Although this is relatively early in the process, initial feedback from our questionnaire has been resounding positive feedback toward the mechanism and its implementation. Further analysis will be conducted at the end of the Spring 2009 semester and shall be presented at the ASEE 2009 conference.

From our survey, some suggestions have been made that shall be considered for future implementations of this assessment process. The three personal contribution categories will be divided into subcategories such that evaluation points may be applied to the students performance, and to better inform the students regarding where their performance may require improvement. The numeric weights may also require some adjustment in order to reduce the risk of over penalizing students. Checklists or other tools may also be included with the rubrics in order to better guide the students when applicable. Additional work is required toward implementing or utilizing tools to help automate the data collection process for each student assessment cycle as there is a significant workload upon the instructor in collecting and processing the various pieces of data.

Our initial evaluation of this assessment process shows potential for applicability toward other team-based class projects within and outside of the computer and software engineering disciplines. In order to implement this process, it would be important that the instructor select a variety of analysis tools in order to adequately assess the student performance toward the rubrics. The current rubrics have been written to be applied toward any team project so long as adequate data is collected. However, if course specific checklists are to be added for student guidance, the instructor would be required to tailor those checklists for their own course.

Bibliography:

[1]Jack McGourty, Peter Dominick, Mary Besterfield-Sacre, Larry Shuman, and Harvey Wolfe. "Improving Student Learning Through the Use of Multiple Assessment and Feedback." In the Proceedings of the 30th ASEE/IEEE Frontiers in Education (FIE) Conference, Kansas City, MO, Session F1A, Vol. 1, pp. 21—26 October 18-21, 2000.

[2]Debora B. Kaufman, Richard M. Felder, and Hugh Fuller. "Peer Ratings in Cooperative Learning Teams." In the Proceedings of the 1999 Annual ASEEE Meeting, June 1999.

[3]Deborah B. Kaufman and Richard M. Felder. "Accounting for Individual Effort in Cooperative Learning Teams." Journal of Engineering Education, 89(2), 133—140, 2000.

[4] R.W. Brown. Autorating: Getting individual marks from team marks and enhancing teamwork. In the Proceedings of the 1995 Frontiers in Education Conference. Pittsburgh, Online at: http://fie-conference.org/fie95/3c2/3c24/3c24.htm, November 1995.

[5]Dawn E. Wilkins and Pamela B. Lawhead. "Evaluating Individuals in Team Projects." In the Proceedings of SIGCSE 2000, Austin, TX, pp. 172—175, March 2000.

[6]Jane Huffman Hayes, Timothy C. Lethbridge, and Daniel Port. "Evaluating Individual Contribution Toward Group Software Engineering Projects." In the Proceedings of 25th International Conference on Software Engineering (ICSE'03), pp. 622—627, 2003.

[7]Nicole Clark, Pamela Davies, and Rebecca Skeers. "Self and Peer Assessment in Software Engineering Projects." In the Proceedings of the 7th Australasian conference on Computing education, Newcastle, NSW, Australia, Vol. 42, pp. 91—100, 2005.

[8]Dean Sanders. "Managing and Evaluating Students in a Directed Project Course." Proceedings of the 15th SIGCSE. ACM SIGCSE Bulletin, Vol. 16, Issue 1, pp. 15—25, 1984.

[9]EcoCar Challenge, Online at ecocarchallenge.org, 2009

[10]Alistair Cockburn, "Crystal Clear: A Human Powered Methodology for Small Teams", Addison Wesley Publishing, 2004.