

Peer & Self Evaluations as Means to Improve the Assessment of Project Based Learning

Maartje E.D. van den Bogaard, Gillian N. Saunders-Smiths
 Delft University of Technology,
 Faculty of Technology, Policy and Management, Faculty of Aerospace Engineering
 P.O. Box 5, 2600 AA Delft, the Netherlands

Abstract – Peer and Self evaluations are an excellent way to monitor and evaluate group skills in project based design work. Their use has become increasingly popular with increase in popularity of project based learning. Peer evaluations allow faculty to differentiate in individual grading of group work and prevent free-riding. This paper makes a distinction between three types of peer evaluation: ranking students, dividing assets between students and rating students on qualitative criteria. Based on these criteria it compares the system developed at Delft University of Technology with existing systems in Eindhoven (NL) and Sydney (Aus) based on their functionality and cultural dimensions. We will also discuss the hurdles faced by all parties in using this tool in grading and how we have overcome them. This results in a clear set of recommendations for lecturers who would like to use peer evaluation in their projects.

Index Terms - Peer evaluation, Self Evaluation, Assessment, Project Based Learning.

INTRODUCTION

Project based learning at engineering universities are increasingly popular to heighten the attractiveness of engineering degrees and to address other important professional skills such as teamwork, oral and written communications. Because of this popularity, a new type of problem is encountered. Lecturers have to evaluate every student on an individual basis in order to eventually merit them. However, in project based education, our students are a part of a group. This makes it harder for a lecturer to distinguish the exact contribution of a student to a group report or design. Individual learning processes and progress are difficult to assess as well.

To this end many universities have over the last 10 years started experimenting with different forms of monitoring and assessing the individual learning processes. One such way is the use of peer and self evaluation in this process. This paper aims to benchmark the existing system for peer and self evaluation as a monitoring and assessment tool at the Faculty of Aerospace Engineering at Delft University of Technology, which has existed for some 7 years now, against newly available similar systems used by other universities both in the Netherlands and abroad to see if the current tool is still fit-for-purpose.

In the late 90s the Faculty of Aerospace engineering at Delft University of Technology went through a major curriculum change. As part of that change traditional lab and design classes were replaced with group based design projects [1], [2]. With that the need started to arise to monitor the group process that took place outside the direct influence sphere of the lecturer and also the need arose to still be able to give an individual grade for a student's contribution to a group project.

In this paper we describe the purpose build Delft peer and self evaluation system, a system that is in use at Technical University Eindhoven, which is also based in the Netherlands. This system is paper-based. It is interesting because it includes only one item. The instrument has a high validity and is easy to administer. The third system we describe in this paper is an Australian online system that is specially fit for design exercises.

DEFINITION AND TYPES OF PEER ASSESSMENT FOR GROUP WORK

In the context of this paper peer evaluations are defined as assessment by peers in the same team (i.e. fellow team members) with regards to the work carried out in the team assignment. Self-evaluation is the evaluation of one's own performance on the team assignment. The assessment may or may not be based on a set of predefined criteria.

Peer and self evaluations generally can be divided in three types:

1. Evaluations based on ranking, i.e. students are asked to rank each team member with respect to each other resulting in a list with the "best" student at the top of the list and the "poorest" performing student at the bottom of the list. This is sometimes done per predefined learning objective.
2. Evaluation based on dividing a certain amount of assets. This type of peer evaluation is characterized by giving a group a set of assets to divide, e.g. a fictional amount of money which they can then choose to divide over the group members.
3. Evaluation based on quantitative descriptions of desired and undesired behaviour. This method is based on learning objectives. Students are asked to rate their peers and themselves on their performance in a predefined learning objectives. The results are a description of how others perceive a students meeting of the learning

objectives and how they themselves perceive their own performance. The difference with the first system is that students are not necessarily ranked amongst each other; it is more the contrast between one's own perceived performance versus the groups perception of that individual's performance.

Eight years ago, when the initial foundation for the current system in Delft was laid, an initial web search generated predominately type 1 and type 2 evaluation systems, most of which were paper-based and used in the United States of America. Type 3 evaluations appeared to be missing. It was therefore that the decision was made to design a type 3 system for use in project based education at the faculty of Aerospace Engineering at Delft University of Technology in 2000 which did meet the requirements of Faculty of Aerospace Engineering had. They therefore started to develop their own system: PeEv.

I. Motivation behind type 3 systems

In project based learning, the results are secondary to the learning process [3] combined with that team skills are often learning objectives in the projects and the degree course. When team skills are part of the learning objectives it is important that students have an opportunity to develop and learn these skills and that members of staff have the opportunity to monitor and coach this process. This effectively eliminates type 1 systems as they do not give any direction to improvement of student's performance. Type 2 does allow for student improvement and has as an advantage that you can express the progress quantitatively, but is more difficult to give feedback on multiple learning objectives.

CULTURAL DIMENSIONS

A second reason for developing our own Peer Evaluation system is the cultural dimension. At Delft several lecturers experimented with type 1 and type 2 evaluations. Type 1 and type 2 evaluations were mostly unsuccessful in Delft. Type 1 evaluations failed because students were unhappy at having to rank each other.

We believe that the reason for this lies in the difference in culture between the United States of America and the Netherlands. Hofstede in his standard work on cultural differences [4] defines 5 cultural dimensions which can be used to explain the difference in working methods of people in general. These five dimensions are listed in Table 1.

The Netherlands and the US have some things in common, as is reflected in similar scores on some dimensions. Power distance is quite low in both countries. This is indicative of a greater equality between societal levels. This orientation reinforces a cooperative interaction across power levels and creates a more stable cultural environment. Both countries score high on individualism, indicating more individualistic attitudes and relatively loose bonds with others. People tend to be more self-reliant and look out for themselves and their close family members. For the Netherlands Hofstede specifies this dimension a little more closely: privacy is

considered the cultural norm and attempts at personal ingratiation may meet with rebuff. Individual pride and respect are highly held values and degrading a person is not well received, accepted, or appreciated.

TABLE 1
HOFSTEDE'S CULTURAL DIMENSIONS [5]

Cultural Dimension	United States of America	The Netherlands	World Average
Power Distance	40	38	55
Individualism vs. Collectivism	91	80	43
Uncertainty Avoidance	46	53	64
Masculinity vs. Femininity	62	14	50
Long vs. Short Term Orientation	29	44	45

The countries score differently on the other dimensions. The USA's low ranking in the Uncertainty Avoidance dimension represents a society that has fewer rules and does not attempt to control all outcomes and results. It also has a greater level of tolerance for a variety of ideas, thoughts, and beliefs. The Netherlands score in the midrange on this dimension, indicating a cultural tenancy to minimize or reduce the level of uncertainty within the population by enacting rules, laws, policies, and regulations to cover most any and all situations or circumstances.

The most distinct difference between the United States and the Netherlands is the masculine-feminine. A low score on this value may be indicative of a low level of differentiation and discrimination between genders and people in general. This low ranking may also be displayed as a more openly nurturing society. The differences in scores on uncertainty avoidance and masculinity effectively make a type 1 system seem too harsh and uncaring for Dutch students. A type 2 system will often yield an equal distribution of assets. A type 3 system is more suitable as it allows students to care for each other and to show appreciation. A type 3 system is probably less threatening to students and it allows for the personal growth of the student.

DELFT: PEV[6]

When looking for a Type 3 system one of the authors came across an anonymous table listing 5 criteria which the author deemed important learning objectives. This table was extremely well-suited for implementation into a type 3 system. For each of these criteria 5 descriptions of team member behaviour were given (See table 2) ranging from undesirable to desirable behaviour. It was this table that inspired the author to turn it into a Peer and Self Evaluation System, initially on paper. Teaching-assistants working for the author in project education at the time [5] turned the system into an online system, which they have developed since into what PeEv is now.

TABLE 2
PEEV ASSESSMENT TABLE

	Job Performance	Attitude	Leadership/Initiative	Management of Resources	Communication
Excellent	Consistently does more than required. Work is of exceptional quality.	Positive and professional attitude which favourably influences other company members.	Takes initiative to seek out work, concerned with getting the job done. Very involved in the technical project.	Uses time effectively in and out of group and works to get others to do the same. All tasks completed on or ahead of schedule.	Oral and written skills excellent. Very effective within the group and to reviewers.
Good	Sometimes does more than required. Work is of high quality. A producer.	Positive attitude toward project and the team.	Readily accepts tasks, sometimes seeks more work. Gets involved in the project.	Uses time effectively in and of group. Completes all tasks on time.	Usually effective.
Satisfactory	Performs all assigned tasks. Quality of work is acceptable.	Neutral attitude.	Gets involved enough to complete tasks. Does his/her share.	Wastes some time in group, but works hard when a deadline is near. Most tasks completed on time.	Generally gets the point across. Tries to improve in weak areas.
Marginal	Performs all assigned tasks. Work must be redone or repaired to meet standards.	Negative attitude toward project and/or project.	Tends to watch others work. Gets involved only when necessary. Volunteers to help when it will look good.	Wastes most of group time. Seldom seen doing productive work. Some tasks completed late.	Skills ineffective. Makes an effort to improve.
Poor	Performs some assigned tasks. Work must be redone by others to meet standards.	Negative attitude which adversely affects other company members or project.	Lets others do the work; does the minimum he/she thinks is needed to get by.	Does little useful work in group or out; wastes his/her time and others. Work is constantly late.	Skills ineffective. Makes little or no effort to improve.

I. Working principle

PeEv works on the principle that each student in a project group evaluates the other members of their project group and themselves by indicating how each member of the group fits in the matrix. PeEv then calculates the average of the group's assessment and then returns that value as well as the students own score. The feedback given to the student is graphically or in text using the phraseology of table 2. Students never see the actual numerical score. The final score is only computed after all students have evaluated or after the session is closed. This is to avoid students working towards a desired score. Tutors and staff can see at any time who has or has not filled in the evaluation. Only staff can change see overall result of all groups.

Next to that project group tutors can also score the students on the same criteria in the table and this feedback will also be returned in the same way. There is also a comments box where students and tutors can comment on the evaluation they made. They can choose for their comments to be public (everyone in the group can see them), Personal (only the tutor and staff) and administrative (only the PeEv administrator can see them).

After all students in a project group have scored each other the system displays the results to the students. They can then see per learning objective how others view them compared to their assessment of their own capabilities. PeEv has been programmed in MYSQL and is therefore operating system independent.

Students, Tutors and staff accounts have to be created in PeEv. This can be done manually (one by one) within the system itself or alternatively a batch *.csv file can be created. A macro based spreadsheet has been developed to quickly generate input files.

In PeEv the PeEv administrator creates sessions for the Peer evaluation to be carried out. We recommend 2 peer evaluations during the project. Two evaluation are performed,

at the half way point, (so that problems can be detected and dealt with) and at the termination of the project. The length of a PeEv session can be set within PeEv.

Also PeEv has the ability to email users with their account details and inform them of PeEv opening and closing dates. PeEv can also send out reminders to those who have not yet evaluated. PeEv also has a Password reset email function. Standard emails can be generated in text format.

II. Criteria

Although other criteria can be added, PeEv is based on a matrix of behaviour with the criteria and the different behaviour belonging to such criteria. Current Criteria of job Performance, Attitude, Initiative, Management of Resources and Communication however meet the requirements set by the lecturers of Aerospace Engineering. So far the only change which has been made is that for 3rd year capstone projects the criteria for job performance were split into commitment and technical quality. PeEv is flexible enough to change or add any criteria providing they are qualitative described as in table 2.

III. User Experience

PeEv has now been in use at the Faculty of Aerospace Engineering for 7 years. Eight peer and self evaluations are carried out per year among some 650 freshmen, sophomore and junior students distributed in teams of some 10 students. Other faculties within the university have been adopting the system over the last 3 years.

Experience showed that just having the system in place was not enough. Students need to be motivated to use it, to be explained why it benefits them. If not they will see it as a system with the sole purpose of catching them out rather than a system which teaches them to take a critical look at their own performance.

IV. Advantages and Disadvantages

Advantages are that PeEv is a robust operating system independent peer and self evaluation system that can be used by anyone world-wide. The system and its support files are all written in English allowing for sharing. The system is web-based so student can fill in Peer evaluations from the comfort of their homes should they want to. Secondly it allows student to critically look at themselves and experience receiving peer feedback and getting the opportunity to show during the project that they learned better team skills and show an improvement. From a lecturers point-of-view it is almost as good as having a 24h webcam in a project room without the privacy issues. If filled in seriously it gives you a unique insight into the group dynamics which allows you to coach the group and at the end it will help you grade individual group members more accurately.

The disadvantages of PeEv are that it is a dedicated source code programmed using a MYSQL database. All data going into the system are encrypted such that it is impossible to find a single student's evaluation of just one other single student. Adopting the system will become more difficult in the future as the current builders of the programme are moving on. Another disadvantage is that the system is as good as the lecturers and students using it. Giving quantitative feedback in group based design projects is a skill that requires training and not all lecturers are open to this type of feedback. Also if a group of students collectively decide to not take this system seriously the feedback coming from PeEv will have no value. Again it is up to the lecturer operating the system to properly motivate students.

THE EINDHOVEN SYSTEM [8]

I. Working Principle and Criteria

In Eindhoven Applied Computer Science offers group design projects starting in the first year. The freshmen and sophomores participate in three projects though the year. Eindhoven designed different systems for the freshmen and sophomores. In both cases, the principal aim of awarding different grades is to reduce the number of free riders in the groups. Below we only describe the sophomore system, since it is better suited for the goals we set for the design projects in Delft.

For the sophomore year Eindhoven has devised a paper-based survey. Students rate their group members and themselves anonymously on a one item survey. The one item is dedication. Dedication is not defined, but students tend to understand this concept broadly. For the students this includes corporation, presence, enthusiasm and keeping agreements. A one item survey is quite unusual, however, long surveys are a poor fit with the culture prevalent in applied computer science. The researchers from Eindhoven recognize that set criteria can support the rating, but the best way to give an overall rating is to inquire after the overall impression. The researchers also mention that when you work with more criteria, you should also train students in how to use these. The students in this

field do not appreciate too much non-technical training. Students rate dedication on a 5-point Likert scale, see table 3. Extreme ratings need to be clarified by a written statement.

TABLE 3
EINDHOVEN PEER SCALE

Points	
0	The person skives
1	The person contributes marginally
2	The person is a neutral member of the group
3	The person is serviceable and helpful to the group
4	The person is a pioneer to the group

The students assess the group twice: midterm and at the end of the project. The tutor rates the group members as well. The tutor data is not used in the final score, but is kept on file to monitor group performance and to serve as a check on the reliability of student ratings and complaints. The scores of the midterm assessment are meant for the students to practice with rating each other. The outcomes are used formatively; extreme scores are communicated by the tutor to the person involved. During a group discussion, fellow students can give advice on how such a person could improve his or her performance. The quality of this learning experience is dependant on the skills of the tutor. Extremely low scores can lead to expulsion. However, the project coordinator makes this decision. The decision to expel people from a project is a drastic measure and should be supported by the tutor as well as the group.

The project coordinator performs all the administrative tasks. That way, all the results can be presented anonymously to the students. The students receive their individual ratings and the group rating.

The results of the final assessment are used for summative purposes. The final rating is based on the following formula:

$$G_i = G_g + (AV_i - AV_g)/2 \quad (1)$$

In which G stands for grade, AV for average, i for individual and g for group.

The formula for the final grade is robust. The deviation of the average determines the size of correction. This way, it is hard for students to give each other, or themselves, unrealistically high ratings.

Research [8], [9] on the survey shows that the validity of the survey is higher when the groups are assembled in a random order. The validity of the survey has been scrutinized. It turns out that the inter rater reliability between students and tutors is quite high.

II. User Experience

The students, tutors and teachers are content with this system. It is straight forward and it gives the students the opportunity to award each other grades that recognizes a person's contributions. The reliability of the instrument is quite high, as is shown in statistical research [9]. This research also states that the spread of rating is smaller in the half term surveys than in the final rating. The researchers did not find any evidence of free riding, which would be shown through really

low ratings. This could mean that the system has a preventative effect.

III. Advantages and Disadvantages

The Eindhoven system is simple to implement and straight forward. Disadvantages include the fact that the students are identifiable and that there still is administration to be done by the instructor.

SPARK [10]

Spark is an acronym for Self and Peer Assessment Resource Kit and has been developed by researchers and lecturers from University of Technology Sydney in Australia at the end of the nineties. There were multiple reasons for the staff to start developing Spark. Students frequently complained about the equal group marks for unequal contributions. Free riders and plagiarizers receive a mark that does not represent their efforts. The same goes for better students, who are usually marked down. This is not motivating for group members. Staff members saw this dissatisfaction reflected in complaints from students and low motivation. Paper-based attempts were not able to overcome confidentiality concerns and generated a lot of administrative pressure. The staff needed a system that was generic, easy and reliable.

I. Working principles

Spark is a database programmed in MYSQL that is accessed through the web interface. The interface is designed to be user friendly and easy to comprehend. Spark is a subject set-up operated by the instructor who can batch enroll students and groups from a class list which must include student numbers. However, the students can also form their own groups within the system if this is advantageous. The lecturer sets cut-off dates for group formation and the entering of ratings by students. He also enters criteria for the self and peer assessment process. There is a sample set of criteria in Spark. This set is based on research and generates valid results [12]. However, the designers recommend negotiating relevant criteria together with the students. This could be part of the group formation process and it should be used according to learning objectives. The criteria can be set to be “prompting” (formative) or to “final” criteria that are used in calculations.

A reason to negotiate criteria is to help students become more aware of possible criteria for quality of work. Especially in design related fields, it is important for students to develop explicit criteria and to learn to phrase these. Negotiating criteria can be very helpful for these students, since it requires them to reflect on what criteria add to a well functioning design team and a quality design process. The ratings are numerical. They can range from -1 to 5 according to the lecturers intention or through student negotiation.

When students log on to the system, they can access a FAQ section. When students enter the assessment area, they can opt for a practice sheet or for entering their ratings. They can change these until the cut-off date. Students are also asked to assess their own contributions. When the cut-off date has

passed, the lecturer exports a spreadsheet of all the students and groups, which include SPARK factors for Self and Peer Assessment (SPA) and for Self Assessment compared to Peer Assessment (SA/PA). The lecturer then adds the group marks and uses the SPA factor multiplied by the group mark to calculate individual marks. It is up to the instructor to publish the grades.

II. Criteria

The standard set of criteria in Spark is based on a set developed and tested by Goldfinch. Table 4 shows the Spark criteria.

TABLE 4
SPARK CRITERIA

Category: Efficient functioning of group
Helping the group to function well as a team
Level of enthusiasm and participation
Organizing the team and ensuring things get done
Performing tasks efficiently
Suggesting ideas
Understanding what is required
Category: Leadership
Chasing and coordinating
Deciding who does what and when
Integrating everything at the end to answer the problem
Category: Number crunching
Analysis and cross checking
Data and formula entry and formatting
Finding out how to solve problem
Getting new data
Category: Writing report
Editing format, style, grammar, spelling
Getting extra references and appraising their usefulness
Producing diagrams, figures, tables

III. User Experience

The Spark system has been in place for about 6 years now. Students and teachers are satisfied with the system. Especially the feature to negotiate the criteria is very powerful to gain the acceptance from students. The lecturers have learned to set the cut-off date one week after the deadline of the project. This way the students have time to reflect on the team effort and result.

The instrument shows that students in groups that receive good group marks tend to underestimate their own contributions and rate themselves lower than their peers rate them. In groups with lower group marks, students tend to overestimate their contributions to the group mark. Thanks to the formulas used to calculate the final, individual grades this is not a problem. The adjustment of the grade is relative to the group mark. This means that someone with an outstanding contribution in a lower scoring group can still receive a higher grade than a student in a well performing team who has contributed on an average level.

Thompson reports that one time the students opted to include the -1 and 0 rating in the assessment. Research [10], [11], [12] has shown that the students use these ratings carefully and effectively.

IV. Advantages and Disadvantages

For the students the use of multiple assessment criteria can better reflect the variety of group contributions; the flexibility of changing and customizing criteria can encourage deep approaches to the learning process; it improves confidentiality as they can access the system from home; it extends time for reflection and re-evaluation as they can change their ratings as often as they wish until the pre-determined cut-off date. Apart from the advantages for staff as in decreasing administrative workload, the Spark system has the potential to improve student learning from group work tasks, it produces data which can be used fairly for both summative and formative assessment purposes. Disadvantages lie in the area of lack of transparency on how the grades are affected by the outcomes.

COMPARISON OF SYSTEMS

PeEv and Spark have many features in common; they are both system independent and can be installed on any server. They both have an option to add to or change criteria. In Spark there is an option to negotiate the criteria with the students. Within Spark the scale can also be adjusted. Within PeEv the criteria give students directions on how to improve their contributions to the group. Both generate data in files that can be exported and that are easy to read. They both have a neutral interface that is easy to understand. In both systems individual ratings cannot be traced back to a single student.

A major difference is that Spark can be used for both summative and formative purposes. However, it does not give students any clues on how to improve their performance. In either case, it is very important to train lecturers how to give feedback to the group. Another difference is in the representation of the results. In PeEv students do not see final scores, they only see graphics or text that corresponds with the text in the matrix. In Spark, it is up to the instructor to publish the results. However, because of transparency reasons, the instructor should do this. Also PeEv is not intended to result in an automatic grade. The results from PeEv can be used in grading but not in a direct numerical form. The overall responsibility of what the grade should be still lies firmly with the lecturer.

The Eindhoven system is in a different league. The Eindhoven system does not ask students to rate on multiple criteria, but on just one criterion that looks at the overall impression. This system is not online and therefore not completely anonymous. It is a straight forward system, which is easy to implement. The use of this system for formative purposes strongly depends on the input from the tutor. Eindhoven is the only system where expulsion from the project is open for discussion and where this is imbedded in the system.

CONCLUSIONS & RECOMMENDATIONS

Peer evaluations have been proven to be a very useful tool in helping monitoring and assessing students in project based learning. However whichever system is used does require that tutors are properly trained in the proper didactical use of a

peer evaluation tool and the thinking behind project based learning. Also it requires acceptance by the students that it is a system that benefits their learning, rather than punishing them.

Spark and PeEv are both robust systems which can and should continue to be developed and grow with the changing needs of student learning. They are flexible enough to adapt to changing needs and the fact that they are server and electronic learning environment independent means that they are easy to use for any university no matter how small. The Eindhoven system appears fit-for-purpose for the Eindhoven needs but the lack of anonymity may reduce its effectiveness.

The vital differences between SPARK and PeEv are that PeEv allows for room for improvement of a student's team skills during the project and the fact that Spark can be used for formative purposes. The former is something deemed very important in project based learning in Delft.

We recommend the use of either SPARK or PeEv in project based learning as practical and easy to use tools. Having reflected on both SPARK and PeEv the decision for now is to continue to work with and develop PeEv as it is the system that currently best meets our needs.

ACKNOWLEDGEMENT

This paper was written as part of ongoing research for a national project called PROOF in which Delft University of Technology takes part in cooperation with the Free University of Amsterdam and Utrecht University, funded by SURF. The project aims to investigate current online peer review systems in the Netherlands and upscale and facilitate peer review.

The authors also wish to thank Rob in 't Groen and Jan Smulders for their vital support in creating PeEv over the last seven years as well as all their colleagues, particularly Barry Zandbergen and Aldert Kamp, and the students at the Faculty of Aerospace Engineering whose feedback over the last 7 years have made PeEv to what it is today.

REFERENCES

- [1] Saunders-Smits, G.N., and Graaff, E. de, "The development of integrated professional skills in Aerospace Engineering through problem-based learning in design projects", *Proceedings of the 2003 American Society of Engineering Education Annual Conference and Exposition*, Nashville, Tennessee, 2003, session 2125.
- [2] Brügemann, V.P., Brummelen, E. H., Melkert, J.A., Kamp, A. , Reith, B.A., Saunders-Smits, G.N. and Zandbergen, B., " An example of Active Learning in Aerospace Engineering" in E. de Graaff, M. Nieweg and G.N. Saunders-Smits (eds), *Research and Practice of Active Learning in Engineering Education*, Amsterdam University Press, 2005
- [3] Graaff, E. de, A. Kolmos, History of problem-based and project-based learning in E. de Graaff and A. Kolmos (eds), *Management of change, implementation of problem-based and project-based learning in engineering*, Sense Publishers, Rotterdam and Taipei, 2007
- [4] Hofstede, G., "Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations Across Nations", Newbury Park, CA: Sage Publications; Second Edition; February 2003
- [5] Hofstede, G., www.geert-hofstede.com - Official website Geert Hofstede on cultural dimensions.
- [6] Saunders-Smits, G.N., www.peev.tudelft.nl – Online Peer Evaluation system of Delft University of Technology.

- [7] Andernach, T. and Saunders-Smiths, G.N., "The use of Teaching-Assistants in Project Based Learning at Aerospace Engineering", *Proceedings of the 36th ASEE/IEEE Frontiers in Education Conference*, San Diego, 2006.
- [8] Perrenet, J.C., C.M. Morgan, *Individual Differentiation in the Assessment of Group Work*, DBL booklet no.5, Educational Service Center TU Eindhoven, Eindhoven, 2003.
- [9] Perrenet, J.C., Wederzijdse beoordeling bij groepswork, in *Onderzoek van Onderwijs*, 2003, vol. 32, nr. 1, pag. 7-10, (In Dutch).
- [10] Thompson, D. & I. McGregor, Self and Peer Assessment for Group Work in Large Classes in P. Kandlbinder, (ed), *Making a difference: 2005 Evaluations and Assessment Conference Proceedings* (pp. 41-52), Sydney: University of Technology, Sydney, 2005.
- [11] Thompson, D., <http://www.educ.dab.uts.edu.au/darrall/sparksite/> - Official SPARK Website
- [12] Thompson, D., Online learning environments with group-based assessment: evidence of significant benefits from the integration of web-based projects with self and peer assessment in a visual communication undergraduate degree program. *Paper presented at CLTAD Conference London, April 10-12 2002: Enhancing the curricula: exploring effective curricula practices in art, design and communication in Higher Education*, 2002.