SELF-ASSESSMENT AND REFLECTION IN A 1ST SEMESTER COURSE FOR SOFTWARE ENGINEERING STUDENTS

Jacob Nielsen, Gunver Majgaard and Erik Sørensen
The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

ABSTRACT

How can student self-assessment be used as a tool and become beneficial for both lecturers and students? We used a simple self-assessment tool for pre- and post-testing on a first-semester engineering course. The students graded their knowledge on human-computer interaction based on their ability to understand and explain specific concepts. The assessment tool was inspired by the SOLO-model.

The assessment tool promoted practice reflections on the academic concepts. In the pre-test the students became aware of specific academic expectations in the course and they got to grade the gap between their own current knowledge and what would be expected at the end of the course. The lecturer could right from the beginning sharpen the academic semester plan based on the assessment results.

The post-test could be used by the students as a tool for grading their knowledge in preparation for the final exam. From the lecturer's perspective the post assessment was useful in optimising the course for the next throughput.

1. INTRODUCTION

How can student self-assessment be used as a tool and become beneficial for both lecturers and students? From an educational perspective, pretesting students is an important part of tailoring a course to fit with the students' prerequisites. When evaluating the students' learning at the end of a course it makes sense to relate that to their prerequisites in order to measure how much the curriculum of the course has affected their learning. Making the students self-assess their knowledge level at the beginning and the end of a course is one approach at making the students actively reflect upon their own learning.

Other resent research has investigated pre- and post-tests of students' self-assessments (Schiekirka et al., 2013), which is based on self-assessment of specific learning objectives related to factual knowledge. They conclude that their tool is easy to implement and assists teachers in identifying strengths and weaknesses on the level of specific learning objectives for a particular course.

Within our 1st semester human-computer interaction (HCI) course, we wanted to promote our students' HCI competences and analytic competences. We wanted the students to create prototypes, be reflective and articulate their design process. We believe that dialog based on the academic theory and their design experiences reinforces the learning process. This is theoretically supported by Schön (2001) and Bateson (2000).

Furthermore, we wanted to make the students aware of the main topics in the course before we got started. We wanted them to take in and reflect on what they were going to learn this semester. This was done through a pre questionnaire. In the pre questionnaire the students became aware of what was expected of them on the conceptual level. This type of learning activity was of a reflective nature. This awareness could also help the students to focus on important concepts during the semester.

The other type of learning activity we promoted in the course was active participation while constructing interactive systems. The knowledge achieved by the students was expressed in actual designs of prototypes and reflections on these. Part of the knowledge expressed in action and in the design of prototypes will often be difficult to put into words and can be described as tacit knowledge (Schön, 2001; Agyris, 1978).

The concept of knowledge-in-action alone is not sufficient in a learning process or in a field practice. This must be supported by the more retrospective forms of reflection - reflection-on-action. Reflection-on-action helps the students to articulate conceptual knowledge on app-programming and HCI. In the retrospective reflection process their own experiences are connected to emerging conceptual knowledge. And conceptual knowledge is used in the professional communication amongst peers.

In the classroom we wanted both knowledge-in-action and reflection-on-action. It is in the interplay between these forms of reflections that the skilled designer unfolds his potential. It is in this interplay that innovative processes evolve. It is also in this interplay that the students achieve a good learning depth.

To put extra focus on the reflection-on-action the students made a self-assessment at the end of the course similar to the one at the beginning. In this assessment the students again were to reflect on their understanding of core concepts in HCI. This type of reflection was not so much linked to optimising their interactive designs, it was more linked to their understanding of the academic concepts in HCI. This type of reflection is important and productive in higher education. The students must understand the academic concepts in depth in order to develop a critical and analytical approach to them.

The testing tool – the questionnaire - used in this work was based on an idea of one of the authors. The primary goal was to make it measure the learning using the SOLO Taxonomy (Biggs and Tang, 2007), which divides learning into a hierarchy of 5 levels: prestructural (misses point), unistructural (identify, do simple procedure), multistructural (enumerate, describe, list, combine, do algorithms), relational (compare/contrast, explain causes, analyse, relate, apply) and extended abstract (theorise, generalise, hypothesise, reflect). The SOLO Taxonomy is often used to form intended learning outcomes for any given course using this hierarchy of verbs, but here we use it to form the grading system presented in the methods section.

Our main objectives for this study were to design and evaluate a simple self-assessment tool for pre- and post-testing in a first-semester engineering course.

2. METHODS

The research methodology was based on Design-based Research and action research (Barab and Squire, 2004; Lewin, 1946; Majgaard, 2010; Majgaard et al, 2011). Design-based Research is a branch of educational research that uses the iterative design of educational interventions to exemplify and develop theories of learning. It also brings a change in the behaviour of the target group into focus and allows emerging goals. Experiments and critical reflections are the core of the research method, allowing learning from and through practice. In this research the experiments covered self-assessment before and after the course.

The target group of this study is first-semester Software and IT engineering students. The main academic learning objectives of the 5-ECTS Interaction and Interaction Design course are that the student at the end of the course should be able to:

- Plan a user-centred design process
- Investigate the users, their needs and their practice through interviews and observations, and present the results in ways suitable for making design decisions.
- Involve users in design and evaluation in suitable ways
- Design interactions to fit with the users' needs and practices
- Apply fundamental design rules for user-friendly designs
- Describe different types of interactions
- Use selected types of interactions
- Plan and conduct evaluation of interaction design and present the results
- Develop simple digital prototypes
- Reflect on interactive design processes and the meaning of good design

The book used in the course is Interaction Design: Beyond Human - Computer Interaction. (Rogers et al, 2011). The course was followed by approximately 70 engineering students and run in parallel with their 10-ECTS semester project on user-centred design allowing them to work more thoroughly with the theories and methods of the course. Besides the pre- and post-tests the students had an oral examination.

The questionnaire, which is used for both the pre and post student self-assessment in this course is based on a likert-type (Rensis, 1932) SOLO-inspired (Biggs and Tang, 2007) grading of the students' assessments of their own learning. The questionnaire is comprised of 50 relevant terms within the subject areas: interaction design and user-centred design. The terms are chosen to broadly cover the learning objectives of the course as well as the content of the course book and can be seen from Figure 1. We are not going for a specific number of questions, but rather want to make sure we cover the essential parts of the course material. With our questionnaire, we have three aims: the first is to get an overall measure of the students' prerequisite knowledge of the selected terms. The other is to get an indication of the learning taking place during the course through a comparison of the results of the pre and post questionnaires. Our third aim is to get the students think about their own knowledge of the different subjects in this course through self-assessment reflections.

The succession of the terms of the questionnaire are closely related to the succession of the subjects of the book and the course plan, and as such lower-numbered terms are presented early in the course and high-numbered terms at the end of the course.

The questionnaire is handed to the students in paper form at the beginning of the first of a total of 12 lessons within the course. The students are asked to fill in the form grading each of the 50 terms using the following numbers:

1. I have never heard of this before.
2. I have heard of this before, but do not really understand what it means.
3. I have an idea about what this means, but I don't want to have to explain it.
4. I have a clear idea about what this means and I am able to explain it.
5. I know exactly what this means and I am also able to relate it to other subjects.

The data of the filled-in and anonymous questionnaires are collected and entered into a spreadsheet, where the average grading of each of the subjects is computed and visualised in a bar graph. The average grading results are analysed and from these eventual changes can be made to the course curriculum.

The post questionnaire is handed out at the last lesson of the course before the examination preparation period and the procedure is exactly the same as for the pre questionnaire. The average grading of each of the terms of the pre and post questionnaires can now be compared and we use this as an indication of the teaching and learning results of the course.

3. RESULTS

With the pre- and post questionnaires, we attempted to measure the student's self-assessed learning at the beginning and at the end of the course.

3.1 The Pre Questionnaire

62 students answered the pre questionnaire, and looking at the results of the average grading of the terms in this test (the dark grey bars of Figure 1), we see that the responses to the 50 terms are widely spread from just above 1 up to ~4.6. The average is 2.47, stating that if we look at all the student answers, the students would rate their own knowledge as "I have heard of this before, but do not really understand what it means".

3.2 The Post Questionnaire

43 students answered the post questionnaire, and now the average grading of the terms (the light grey bars of Figure 1) range from just below 2.50 up to 4.60. The average is 3.82, which means that overall the students would rate their understanding of the course as "I have a clear idea about what this means and I am able to explain it".

3.3 Comparison

The question terms that show the smallest gap between pre and post are e.g. "Questionnaires", "Brainstorming" and "Evaluation". The question terms that show the largest gap between pre and post are e.g. "Wizard of OZ – model", "High Fidelity / Low Fidelity" and "Life-cycle model of Interaction Design". The average rise from pre to post is 1.36.
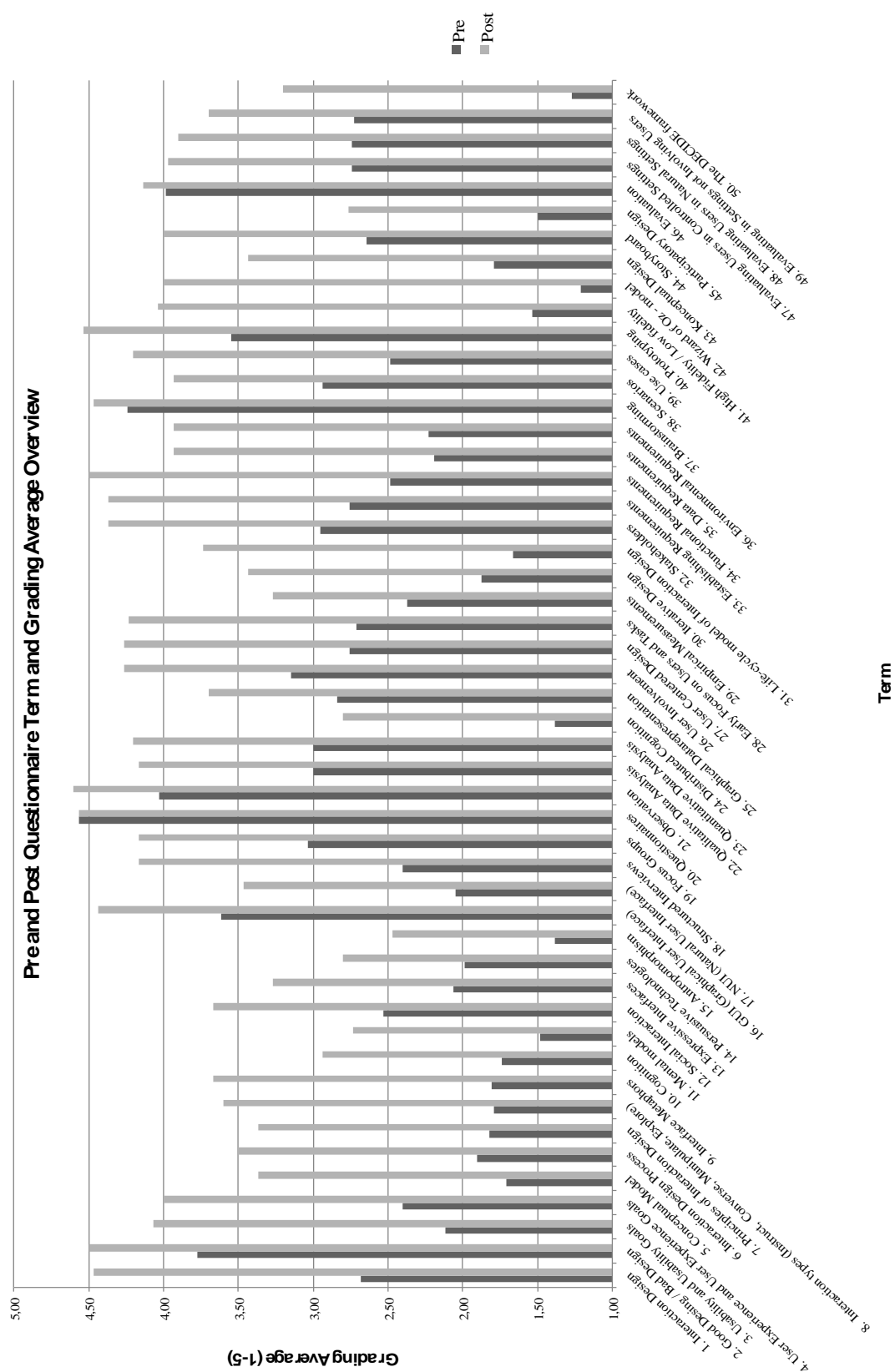
Figure 1. Results of Pre and Post Questionnaires in the Course Interaction and Interaction Design, 2012

4. DISCUSSION

4.1 Analysis of the Results of the Pre Questionnaire

As anticipated from the questionnaire design, the students rate the terms that are commonly known to people at a higher level of understanding. "Questionnaires", "Observation" and "Brainstorming" are rated above 4 followed by "Good Design / Bad Design", "GUI", "Focus Groups", "User Involvement", "Prototyping" and "Evaluation" which are all between grading 3 and 4. We were surprised to see "Focus Groups" and "User Involvement" rated at this level, since we did not expect first-semester students coming from pre-university educations to have any knowledge about this.

At the lower end of the self-assessed level of understanding, we see subjects such as "Wizard of Oz", "Anthropomorphism", and "Distributed Cognition". These are clearly very specific subject terms and looking through the data there are only a few students that rate these higher than level 3. Overall the students have never heard of these terms before.

The result of the pre questionnaire only led to a few changes in the curriculum of the course. First of all, there were only few surprises going through the statistics, and secondly, the spreading of the results of the individual questionnaires was for most subjects large enough that it would still be necessary to let the students work with a given subject in order to be able to reach the same level of understanding. What we did change based on the pre-test was, however the focus on brainstorming within the projects. Here, a large majority of the students rated this high enough to make us assured that they would be able to do proper brainstorming in their project groups. And since the project groups consisted of 5-6 students, we believed that those who did not have any experience in brainstorming would gain this from their fellow students within the groups. If the students scored four in brainstorming, they agreed on being able to explain the concept. This isn't the same as being able to implement a brainstorm and it isn't the same as knowing specific brainstorming methods. But from a debate in the classroom we came to the conclusion that brainstorming was something that most students had actually worked with beforehand and the project groups were all quite confident on how to proceed with this.

From the lecturer's perspective the pre questionnaire resulted in minor adaptations of the course. This can be compared to Schöns reflection-in-action where the lecturer adapts on the fly his curriculum (Schön, 2001; Argyris 1978).

From the students' perspective the pre questionnaire gave them a possibility to reflect on the important academic concepts in the HCI-course. The students also obtained a small insight into the gap of what they already knew and the knowledge they were supposed to be familiar with at the end of the course. They might also have obtained some of the same knowledge while browsing in the semester book on HCI. In addition, the questionnaire gave them time to reflect on their own academic knowledge level and the expected knowledge level. This can be described as specific knowledge gaps (Angelo and Cross, 1993).

4.2 Analysis of the Results of the Post Questionnaire and Comparison from the Lecturer's Perspective

The calculated averages of the post questionnaires show increased levels of self-assessed understanding for all of the terms except one. Luckily, no negative learning was observed at any of the terms. This was first and foremost of great reassurance in terms of the lecturer of the course, but it also gave a good overview of where the students had learned the most and the least and how this would potentially change the curriculum of the course this year and the following. From the results obtained, it is quite obvious that the students have been honest about their learning. Preferably, the lecturers would have liked to see an average for most of the terms just around or above 4, stating that the students would be confident enough with each term in order to be able to explain it and maybe relate it to other subjects inside or outside the curriculum. In the following we have picked out a few of the most interesting terms from Figure 1 and analysed them further.

4.2.1 "1. Interaction Design"

This term represents the overall subject of the course, and as such it is important that the students leave the course with a high degree of knowledge about what this term means and also an ability to explain it to others, and in the case of their semester project, an ability to to work with the different parts involved in doing interaction design. The students enter the course with a knowledge grading of this at an average around 2.70, which means that they probably read the study plan before going to the first lesson, so that they have an idea about what interaction design means and what the contents of the course are going to be. At the last lesson this grading is very close to 4.50, which is considered very good indeed and which assures the lecturers that at least an overall understanding of the course contents has been achieved.

4.2.2 "4. User Experience and User Experience Goals"

Setting goals for and assuring specific user experiences are some of the most important aspects of doing interaction design today, so a good understanding of these terms is a must and a necessary prerequisite if you want to work at a professional level with interaction design. The pre questionnaire gives an average grading of these terms around 2.4 and the post questionnaire 4.00. Both usability and user experience have been central subjects in the teaching of the course and it is thus reassuring that the students rate themselves at a high level of understanding at the last lesson of the course.

4.2.3 "6. Interaction Design Process"

Looking at the learning objectives of the course the first objective is for a student at the end of the course to be able to "Plan a user-centred design process", and the other objectives mention process keywords such as Design, Prototyping and Evaluation. In the course we have spent a lot of time teaching the four key elements of the interaction design process (Establishing requirement, Designing alternatives, Prototyping and Evaluation) and outlining that doing interaction design also naturally means doing user-centred design. It therefore comes as a surprise that the students do not reach an average grading of at least 4 for this subject at the end of the course, and therefore we will increase focus on this subject next semester. One analytical, but unsupported comment to the grading of this subject may be, that as the post test was run at the last lesson of the course – before the students' preparations for the exam, and since the course material about the interaction design process is placed at the beginning of the course the students had less memories about this term and had troubles relating it to the actual content of it.

4.2.4 "20. Questionnaires"

As stated above, it was surprising for us to see "Questionnaires" rated at this level (4.56) at the pre questionnaire, and it did not leave much space for improvement in the post questionnaire (4.57). This clearly indicates one weakness with this type of measurement. When the students indicate a high knowledge level at the beginning of the course it is difficult to measure all the new material that they have worked with during the course (e.g. types of questionnaires, different ways of asking questions, measuring quantitatively or qualitatively etc.).

4.2.5 "26. User Involvement" and "27. User-Centred Design"

These two terms are central to the learning objectives and very much related to the overall interaction design process and within the teaching there has ben great focus on stating the importance of the user and the necessity of involving them in the design process in order to assure usability and the right user experience. It was nice to see these as rated above 4 with the post questionnaire.

4.2.6 "42. Wizard of Oz – model"

The Wizard of Oz term rates, as expected, as unknown to the students in the pre questionnaire. This does, however, change dramatically during the course, and it receives an average rating of 4 at the end being the one term for which the students' average knowledge level changed the most. Wizard of Oz is a term for a model of how to simulate interaction with very simple and usually non-interactive prototypes, and thus the term cover a very important part of doing initial prototype evaluation at a stage of design where it is still too uncertain to start spending time developing technologies. This term is therefore regarded an essential part of interaction design and therefore it is very important that it is highly rated in this questionnaire.

4.2.7 "45. Participatory Design"

Within the course and especially the students' semester project we did not ask the students to involve the users directly in the design of the prototypes. Therefore, and as an undesired result, the students regarded this term as less important within the curriculum even though it actually is an important part of interaction design that the designer has knowledge about and is able to choose between different methods of user involvement. Learning of this term will as a consequence be reinforced within the next semester course.

The above was just a small selection and examples of terms and subjects that have allowed us to evaluate this course through the pre and post questionnaires and the students' self-assessment. The following is a list of numbers of the terms and subjects that will receive increased focus in the course next semester: 5-15, 29-31, 43 and 45. Hopefully this will be possible without influencing some of the other terms in a negative direction. From the lecturers' perspective the post assessment was aimed at optimising the course for the next throughput. E.g. the lecturer wanted to increase focus of concepts: 5-15, 29-31 etc. This can be compared to Schön's reflection-on-action which is the subsequent reflection and evaluation on the process that has happened, and its potential consequences (Schön, 2001; Argyris 1978). It is precisely this type of reflection you want before running this new course for the second time.

4.3 Analysis of the Results of the Post Questionnaire and Comparison from the Students' Perspective

From the students' perspective the post questionnaire gave them a possibility to reflect on their current academic level at the end of the course. They could also compare their current level with their initial level and reflect on their progress. Furthermore, the test provided the students with a chance to evaluate their own academic level before preparing for the final exam. These reflections are of adaptive nature and can be compared to Bateson's 2nd level of learning which provides for good and normal learning (Bateson, 2000). But the students also had the possibility to use the test results as a foundation for preparing their exam. In this way the results made the foundation for a possible layout of a learning strategy. A change in a student's learning strategy can be regarded as level-3 learning. This type of learning provides a more enhanced form of learning. Students in higher education can benefit from reflections on own learning strategies (Qvortrup, 2006; Gleerup, 2003).

4.4 Evaluation of the Questionnaire

As the above analysis of selected terms demonstrate it has been quite easy to reflect on the contents of the interaction design course using the calculated and graphically represented results (Figure 1) of the pre and post self-assessment questionnaires presented in this paper, and we can see no reason why this should not apply to other types of courses as well. Using the SOLO Taxonomy for grading knowledge levels of specific course terms through self-assessment is to the best of our knowledge a new approach on measuring learning and evaluating teaching. Others (reference) have used pre- and post-test student self-assessment for measuring teaching quality but with different test methods.

4.4.1 Strengths

What our investigations showed was that this method is easily implemented also in large classes where it becomes a lengthy and often tedious process to pre-test students using traditional assessment techniques. The production of the questionnaire with relevant course terms and subjects can be prepared at the same time as the course plan is made. An electronic version of this kind of questionnaire will be examined next semester, which will speed up the handling of the data.

The process of comparing the results of the pre and post questionnaires becomes easy because with this method we are able to hand out the exact same questionnaires in both cases and the results can be directly compared and evaluated. With non self-assessment tests, such as e.g. multiple-choice having the same test two times during a course would potentially affect the students' answers and as a consequence influence the validity of the test.

With regards to the students' self-assessment there is theoretical background (references) for saying that it is going to be beneficial for their understanding of the course material that they are reflecting on their own learning of the most relevant course terms and subjects two times during the course. We do, however, have no measurement of the effects of this yet.

4.4.2 Limitations

The results showed in this paper prove an increase in the measured levels of understanding for the different terms stated in the questionnaire, which indicates the validity of our method, but further research needs to determine this. We do not yet know if the SOLO-inspired rating used in the questionnaire is optimal for this type of self-assessment. Would it e.g. be beneficial to have more or less levels? Would ratings inspired by Bloom's taxonomy (Bloom and Krathwohl, 1956; Anderson and Krathwohl, 2001) result in more valid results? Also, retropectively the self-assessment should focus more on one's ability to work with a given subject, because knowing exactly what a term means does not necessarily make you able to actually use it and implement that knowledge e.g. in developing new interactive technologies. In order to effectuate this, the wording of the grading system will have to be changed.

Another limitation of our method is that we do not capture the increased knowledge that also takes place with the terms that are rated high (4-5) from the beginning. This has been mentioned previously and is something we will have to take into consideration when we formulate our questionnaire next semester. We may avoid using commonly known terms in the questionnaire, or at least put them in the context of course-relevant terms.

A last limitation in the type of self-assessment that students do through our questionnaire is that it becomes difficult to measure any misconceptions the students may have regarding the different learning elements. Both at the beginning and at the end of the course.

4.4.3 Comparison with Other Studies

When comparing our study to the study mentioned in the introduction (Schiekirka, 2013), the two studies clearly differ in the grading system used. They use 6 levels of agreement to the understanding of a given learning objective, while we use SOLO-inspired 5-level grading system to measure understanding of a given course term. Both methods seem valid, but in our opinion the SOLO-inspired grading that we use is specifically formulated for each grading and make it easier for the students to pinpoint the exact grading and better reflect on their learning.

5. CONCLUSIONS

In this paper we explored how student self-assessment can be used as a tool and become beneficial for both lecturers and students. We used a simple self-assessment tool for pre- and post-testing in a first semester engineering course. The students graded their knowledge on human-computer interaction based on their ability to understand and explain specific concepts.

Generally the assessment tool promoted practice reflection both reflection-in action and reflection-on-action. In the pre-test the students became aware of specific academic expectations in the course and they got to grade the gap between their own current knowledge and what would be expected at the end of the course. The lecturer could right from the beginning sharpen the academic semester plan based on the assessment results.

The post-test could be used by the students as a tool for grading their knowledge in preparation for the final exam. From the lecturers' perspective the post assessment was useful in optimising the course for the next throughput.

The lecturer has decided to use the self-assessment tool again with minor adjustments. The grading system has to be adjusted into being more application oriented. The students should be able to grade their ability to use, apply, analyse, create or design according to specific concepts. This could easily be done within the scope of SOLO.

REFERENCES

Anderson, L. W. and Krathwohl, D. R., 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon. Boston, MA, USA.

Angelo, T. A. and Cross, K. P., 1993. *Classroom Assessment Techniques: A Handbook for College Teachers.* Jossey-Bass

Argyris C. and Schön A. D., 1978. *Organizational learning: A theory of action perspective*. Addison-Wesley, Reading, MA, USA.

Barab, S. and Squire, K., 2004. Design-Based Research: Putting a Stake in the Ground. *The Journal Of The Learning Sciences*, vol. 13, No. 1, pp. 1–14.

Bateson, G., 2000. *Steps to an Ecology of Mind, Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. University of Chicago Press, Chicago, Illinois, USA.

Biggs, J. and Tang, C., 2007. *Teaching for Quality Learning at University (Society for Research Into Higher Education)*, Open University Press.

Bloom, B.S. and Krathwohl, D. R., 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain.* Longmans, Green, NY, NY, USA.

Gleerup, J., 2003. Gyldighed, oprigtighed og ærlighed – om viden og læreprocesser. *I Hermansen Mads (red.), Læring – en status*. Klim.

Lewin, K., 1946. Action research and minority problems. *In Journal of Social Issues*. Vol. 2, No. 4, pp. 34-46.

Majgaard, G., Misfeldt M., og Nielsen J., 2011. How Design-Based Research and Action Research Contributes to the Development of Design for Learning. *Designs for Learning*.

Majgaard, G., 2010. Design based action research in the world of robot technology and learning. *The Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning: DIGITEL 2010*, pp. 85-92. IEEE Press.

Rensis, L.,1932. A technique for the measurement of attitudes. *Archives of Psychology*, Vol. 22, No. 140, pp. 1-55.

Rogers, Y. et al, 2011. *Interaction Design: Beyond Human-Computer Interaction*, John Wiley and Sons, Ltd.

Schiekirka, S. et al, 2013. Estimating learning outcomes from pre- and posttest student self-assessments: a longitudinal study. *Academic medicine : journal of the Association of American Medical Colleges.* Vol. 88, No. 3, pp. 369-375.

Schön, A. D., 1983. *The reflective Practitioner: How professionals think in action*. Temple Smith, London.

Qvortrup, L., 2006. *Knowledge Education and Learning – E-learning in the knowledge society*. Samfundslitteratur Press.