

Five Myths of Assessment

Mats Daniels, Anders Berglund, Arnold Pears

Department of Information Technology
Uppsala University
PO Box 337, 751 05 Uppsala, Sweden

{matsd, anders.berglund, arnoldp@it.uu.se}

Sally Fincher

Computing Laboratory
University of Kent
Canterbury, Kent, CT2 7NF, U.K.

S.A.Fincher@ukc.ac.uk

Abstract

This paper describes some issues concerning assessment and the corresponding motivation for students to work in a desired manner. The issues came from studying assessment in the Runestone project, but are, as we see them, of general interest. Our findings illustrate the need to not take the effects of assessment, nor what it measures, for granted. It is our intention to promote Computer Science Education research as an essential area for improving our education, in this case by exposing myths about assessment as myths.

Keywords: Computer Science Education, assessment.

1 Introduction

To be a teacher is an important and, at times, very rewarding profession. To “see” enlightenment in students is an extremely fulfilling experience, but we would lose our credibility if we wrote “I saw enlightenment/understanding” in the grading report. We need to find alternative ways to assess our students, and to do the assessment as fair as possible. How to assess, and the closely related issue of the effects of assessment on students, have been discussed extensively in most staff rooms as well as documented in literature [Boud, D. 1995, Gibbs, G. 1996, Hult, H. 1998, Jacobsson, P. 1995, Trowald, N. 1997, Wiiand, T. 1998]. These discussions are important parts of forming our profession and it is essential that these issues are addressed based on knowledge and not on “common sense” or ill-funded myths. Undertaking discipline oriented educational research concerning these issues is thus vital [Clancy, M., Stasko, J., Guzdial, M., Fincher, S. and Dale, N. 2001, Daniels, M., Petre, M. Berglund, A. 1998,

Holmboe, C., McIver, L., and George, C. 2001, and Pears, A. and Daniels, M. 2003].

We will investigate five myths about motivating and assessing students and expose these myths as being just that, myths. We base our arguments on studies of a collaborative course offered by Uppsala University (UU) in Sweden and Grand Valley State University (GVSU) in MI, U.S.A.. The course is known as the Runestone project [Daniels, M., Petre, M., Almstrum, V., Asplund, L., Björkman, C., Erickson, C., Klein, B., and Last, M. 1998, Daniels 1999, Last, M., Almstrum, V., Daniels, M., Erickson, C., and Klein, B. 2000] and many assessment issues have emerged and been studied during its lifetime.

2 Runestone

The Runestone course has existed since 1998, and was initially funded by The Swedish Council for Renewal of Higher Education. The course has subsequently been sustained by the institutions themselves, through changes of teachers as well as heads of departments on both sides, as a regular course offering; which provides some confirmation of the value it gives to the institutions involved. Studies of the course have been funded from several other sources over the years.

Several characteristics of this course make it unique: Runestone is based on a 10-week intensive group project. What makes this course unique is that the “groups” which have to work together are extremely distributed. Each team is composed of six individuals, three from Sweden and three from the US. They are not only distributed by virtue of geography and time zone, but by background, educational expectations and environment and, of course, language and culture.

The task that the teams have to work on is the same for them all – and, indeed, the same over time. It might be supposed that this is a fertile ground for plagiarism and particularly longitudinal plagiarism. However, this has been explicitly addressed during the course evolution. Some years students have been encouraged to use solutions from previous years as their starting points; some years this has been required of them. This works because the task is very large and open-ended [Daniels,

M., Faulkner, X., and Newman, I. 2002, Newman, I., Daniels, M., and Faulkner, X. 2003].

The supervisory staff (one at UU, one at GVSU) each supervise half the teams. That means that they will never meet half the students they will be responsible for assessing.

The students are being marked under two separate grading schemes. It was – and is – the case that the US students are graded according to their own institutional norms (one of 5 letter grades, A-D pass, or F fail) whilst the Swedish students are awarded “pass” or “fail”.

Although the actual project is the driving force in the course, it is also clear that a unifying factor across all of these variables – two sets of students, two supervisors, two cultures, two sets of grading rules, one project with several teacher-led IRC-based meetings, called milestone meetings during this course and a final presentation - is being assessed. Across such diversity how do you make it work the same everywhere? In this sense, Runestone exposes some assumptions about the expectation of, and execution of, assessment which can remain tacit within a less extreme environment. Equally, because of the communication restraints, which mean groups have to work in a textual medium (e-mail and Internet Relay Chat (IRC)), group dialogues that are normally invisible are here necessarily made manifest. The milestone meetings are central for the assessment and are thus of high interest in this context.

The project was reviewed in 2001, and at that time we found that there were differences in how the team viewed the process. In discussion these differences with Runestone staff and other colleagues and researchers we came to characterize these as “myths”.

3 Myth 1: All staff perceive assessment criteria in the same way

One of the motivations for the project review was to more completely specify staged assessment criteria for the students. This looked like a good idea, but the process of agreeing on assessment criteria and the actual meaning of them was however far more difficult than expected. For example the two teachers interpreted their joint decision about how to judge the working process in relation to the product of the work differently. While one of them in practice put a heavy emphasis on the process, the other one regarded the outcome as an important factor.

In the 2001 instance of Runestone regular milestone meetings were introduced with the idea that staff might be able to detect, and have a chance to help, teams that were in trouble. Staff-student contact was earlier changed in other ways: each supervisor as well as being responsible for their own teams, was available to *all* students for specific technical areas (in which they had particular expertise) and in which they set milestone assessments and for which they assigned grades. As a corollary, staff workload was reduced, as we removed the need for local expertise in all aspects of the project.

4 Myth 2: All students perceive assessment criteria in the same way

Assessment is often described as the means by which one leads students into a desired way of working. This is partly based on the assumption that students perceive the assessment criteria in a uniform way.

In the 2000 Runestone instance students had to work to a “minimum requirement” to pass. When this was implemented, there were complaints from the students about the interpretation of this. Indeed, on closer inspection, we saw that in the US this was interpreted as 50%, “the minimum to achieve a passing grade” and in Sweden it was thought of more in terms of “doing the job”, i.e. working throughout the whole project, which in effect would resemble a grade A on the American scale.

We also undertook a qualitative study, aimed at discerning, analyzing and describing the students’ experience of being graded [Berglund and Booth, 2002]. We used a phenomenographic approach, as phenomenography [Marton and Booth, 1997] offers the intellectual tools to describe the different ways in which a cohort of individuals - such as a group of students - experiences a phenomenon; which might be as diverse as the act of programming or the experience of being graded. Data was collected in semi-structured interviews: in this study 15 interviews were conducted (both in the US and Sweden) at the beginning of the course and again shortly after its end.

Although a phenomenon can be understood, or perceived, in numerous ways, phenomenographic results allow a researcher to organize the different perceptions present within the cohort into a limited, often rather small, set of categories, where each category summarizes and describes a particular way of understanding a phenomenon. The task of the researcher is thus to analyze the data in the interviews into a set of qualitatively different categories. The results from a phenomenographic research project describe the perceptions that exist within the whole group, without relating them to a specific individual or group of individuals.

The interviews were part of a larger research project, and issues other than grading were also discussed [Berglund, 2002]. The opening question about grading was normally “*You have different grading systems in the two countries. Does this influence your collaboration?*” with follow-up questions concerning other aspects. In the full group of students, three qualitatively different ways the students’ experience being graded were identified:

Grading as something important

This category describes the experience of grading as something of importance, or the point of taking the course. The focus is here on me, as an individual, on my grade and the relation between my achievement and my grade (or the group, the group grade and the grade as recognition of success).

Grading as something not relevant

Experiences in this category are related to the team and its work: these are perceived as more important than grading. The focus is on a personal sense of obligation to contribute to the work of the team, or for the team to do well in relation to other teams.

Grading as an obstacle

Where the aim is to learn, or to do a good project, then there is a perception that it is better to work towards these goals, than to work on tasks that are graded or rewarded in the grading. In this way, grading can be perceived as an obstacle.

In this section we have presented descriptions of three qualitatively different ways in which students experience grading. The meaning of grading expressed in these three categories is distinctly different. Additionally, the relationship *between* the categories can be defined by the scope of the students' concerns. The first category is entirely focused on the individual, whether the issues are concerned with an individual or with the individual within a group; while the third category encompasses issues outside the group, such as the aims of the course. There is thus a wide variation in how the students in this setting perceived the assessment, and this is not likely to be unique to this course.

5 Myth 3: If we set assessment criteria that focus on x [process] that is what students will focus on

As discussed earlier (myth 1) staged assessments were introduced in the belief that it would lead the students to a certain behavior, i.e. to work steadily in a professional manner with a focus on planning and process.

The project is defined by seven milestone deliverables, typically a week apart, which are also assessment points.

Week 1: Team set up. Each person is to interview one of their remote team-mates and produce a web page about them. Everyone works to establish the team web site.

Week 2: Design & Specification. Which describe the individual objectives to be achieved by the team and their analysis of the problem presented in the project description.

Week 3: Motor Control. These detail the server functions which are what actually controls the stepper motors on the maze game.

Week 4: Video Processing. Analysis of the video from the camera, to identify the location of the marble on the maze game board (before sending it to the remote player)

Week 5: Server. Integration of the video processing and stepper motor control with communication to a control interface.

Week 6: Interaction/navigation. Enhancing the server to provide a feedback control loop designed to steer the marble path over the game board.

Week 7: Client applet. Export of the control interface as a GUI (Graphical User Interface) that is available to the player via a web browser.

The attitudes expressed above (myth 2) "Grading as something important", "Grading as something not relevant", and "Grading as an obstacle" ran counter to the staff expectations. That is, the assessment criteria were set up to emphasize the "process", but there was still an overriding focus on the "product", i.e. a functioning system, among some of the groups. We also saw that the assessment criteria was subordinate to the risk of "looking stupid and/or disloyal", for especially the Swedish students, in terms of why the students focussed on "process".

6 Myth 4: Different assessment criteria for different students within a single team will cause different levels of participation/contribution

This is perhaps not a common situation, but it is likely to be a criticism raised as more and more alternative ways to run courses emerge.

There was considerable discussion prior to the 2001 instance of Runestone about the effect within teams of their work being marked under two separate grading schemes. The concern was that the American students would feel that their grades could be negatively effected by the Swedish students' lack of motivation to work past the "pass" limit.

There was a range of opinion on the issue. At one extreme the opinion was that this was not a problem at all: at the other extreme, it was thought that the problem was of such magnitude that the course would have to be abandoned. In-between there was a spectrum of concern: "...its not fair to the GVSU students to put them in teams with people who have different motivations because of the way grades are recorded.", and ".. final grading means that in some sense Swedish students have less incentive to work to the maximum of their ability, and this is further reinforced by their exposure to the USA grading system through the collaboration process..", and "...perception is reality. . if one group perceive that their partners are not contributing as much as they should because of a different grading, then this is their reality and it could poison the Runestone waters." (These discussions were carried out in e-mail, and involved past Runestone instructors as well as other members of both departments responsible for teaching and its administration).

In response to this issue, we introduced a formal Peer Evaluation mechanism, which would count for 10% of an individual's grade. In this way, we hoped to gather information related to how the students' view their work and the work of their team-mates. We thought that the pass/fail" system might be detrimental to team performance in one of two ways: if the Swedes were

actually less motivated to work beyond their passing grade we would see the US students recognizing their lesser contribution; if the US students were strategically motivated to maximize their own grade we would see the US students award themselves higher. In either case the result would be that US students awarded each other more “points” than they awarded their Swedish team members.

Using data gathered from 93 students in 2000, 47 Swedes and 46 Americans, we calculated how each group of students perceived its own contribution and the contribution of its international collaborators [Pears, A., Daniels, M., Berglund, A., and Erickson, C. 2001]. Each student was given USD 100.- to distribute over all members in the team. With the uneven distribution of students, the average for a Swede would be USD 17,23 and the average for an American would be USD 17,17 if everyone spread the money equally among the team members. The table below shows that the difference between the two cohorts was clearly higher in favor of the Swedish students than the uneven distribution would explain. It is especially interesting to note that the American students on average gave the Swedes 40 cents more than to themselves. That is, they perceived and acknowledged that the Swedes contributed more in the project.

Average to Swede	USD 17.53
Average to American	USD 16,97
Average from Swede to Swede	USD 17.66
Average from American to Swede	USD 17.22
Average from American to American	USD 16.88
Average from Swede to American	USD 16.91

Given the staff discussions of the “two grade” system, reported above this allocation pattern was unexpected, indeed quite the reverse of what was feared.

7 Myth 5: Assessment criteria drive motivation (motivation can be manipulated via assessment)

Myth 5 envelopes the previous myths. That assessment criteria guide students' engagement in the subject matter has a reputation as the “the holy grail” of assessment, and serves as a guiding principle for many educators.

One, perhaps to many, surprising result of our investigation around the consequences of different marking schemes for the two student cohorts in Runestone is that the marking scheme was of subordinate importance for what the students did. We find this encouraging as we see it as a step away from what could be seen as a "marking game", i.e. for the teachers to construct a marking scheme that coerces the students into doing the things they believe are important, and for the students to figure out what behavior will give a good grade. We do not claim that marking schemes are irrelevant to what the students do in a course, there is plenty of evidence that they are [Gibbs, G. 1996, Wiiand,

T. 1998], but we do want to point out that they are not as powerful as we think..

There are however other education approaches, e.g. Problem Based Learning [Kolmos and Algreen-Ussing 2001], that claim to motivate students to engage in the subject without needing to resort to close guiding through assessment schemes. Other examples of guiding, and motivating, students from our department range from involving the students in grading and presenting each others code [Berglund, A., Daniels, M, Hedenborg, M., and Tengstrand, A. 1998] to including real users in project oriented courses [Daniels, M. and Asplund, L. 1999, Daniels, M. and Asplund, L 2000], and international collaboration on a small scale [Clear, T. & Daniels, M. 2001].

8 Conclusions

As a cautionary tale, we conclude with a moral:

Don't take for granted that you understand what drives students to pass a course.

9 References

- Berglund, A., Daniels, M., Hedenborg, M., and Tengstrand, A. (1998): Assessment to Increase Students' Creativity: Two Case Studies, *European Journal of Engineering Education*, Vol. 23, No 1.
- Berglund, A. (2002): *On the understanding of computer networks*, Licentiate thesis 2002-02, Department of Information technology, Uppsala university.
- Berglund, A. and Booth, S. (2002): "Are you guys really concerned about the grades?" *On the experience of grading systems as contextual to learning in an internationally distributed computer science course*. ISCRAT2002, Amsterdam, Netherlands.
- Boud, D. (1995): Assessment and learning: contradictory or complementary?, In: Knight, P. (ed) *Assessment for Learning in Higher Education*. Staff and Educational Development Series. Kogan Page, London.
- Clancy, M., Stasko, J., Guzdial, M., Fincher, S. and Dale, N. (2001): Models and areas for CS education research, *Journal of Computer Science Education*, 11(12).
- Clear, T. and Daniels, M. (2001): *A Cyber-Icebreaker For An Effective Virtual Group?*, ACM Conference on Innovations and Technology in Computer Science Education, Canterbury, UK.
- Daniels, M., Petre, M., Almstrum, V., Asplund, L., Björkman, C., Erickson, C., Klein, B., and Last, M. (1998): *RUNESTONE, an International Student Collaboration Project*. IEEE Frontiers in Education conference, Tempe, AZ, USA.
- Daniels, M. (1999): *Runestone, an International Student Collaboration Project*. NyIng report No 11, Linköping.

- Daniels, M. and Asplund, L. (1999): *Full Scale Industrial Project Work*, IEEE Frontiers in Education conference, San Juan, Costa Rica.
- Daniels, M. and Asplund, L.: (2000) *Multi-Level Project Work; a Study in Collaboration*, IEEE Frontiers in Education conference, Kansas City, USA.
- Daniels, M., Faulkner, X., and Newman, I. (2002): *Open Ended Group Projects, Motivating Students and Preparing them for the "Real World"*, IEEE Conference on Software Engineering Education & Training, Covington, USA.
- Gibbs, G. (1996): *Improving student learning – through assessment and evaluation*. The Oxford Centre for Staff Development, Oxford, UK.
- Holmboe, C., McIver, L., and George, C. (2001): *Research agenda for computer science education*, Psychology of Programmers Interest Group.
- Hult, H. (1998): *Examination inom ingenjör- och civilingenjörutbildningarna*, (in Swedish), Institutionen för systemteknik, Linköpings Tekniska Högskola.
- Jacobsson, P. (ed.) (1995): *Quality Review F. Review of the Swedish and Finnish M Sc Programmes in Engineering Physics*, Kungliga Tekniska Högskolan.
- Kolmos, A. and Algreen-Ussing, H. (2001): *Implementing a Problem-Based and Project Organized Curriculum - A Cultural Change*, Das Hochschulwesen, No. 1, Luchterhand Verlag, pp 15-20.
- Last, M., Almstrum, V., Erickson, C., Klein, B., and Daniels, M. (200): *An International Student/Faculty Collaboration: The Runestone Project*, ACM Conference on Innovation and Technology into Computer Science Education, Helsinki, Finland.
- Marton, F. and Booth, S. (1997): *Learning and Awareness*. Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Newman, I., Daniels, M., and Faulkner, X. (2003): *Open Ended Group Projects a 'Tool' for More Effective Teaching*, ACM Australasian Computing Education Conference, Adelaide, Australia.
- Pears, A., Daniels, M., Berglund, A., and Erickson, C. (2001): *Student Evaluation in an International Collaborative Project Course*, SAINT, San Diego, USA.
- Pears, A. and Daniels, M. (2003): *Structuring CSEd Research Studies: Connecting the Pieces*, ACM Conference on Innovation and Technology into Computer Science Education, Thessaloniki, Greece.
- Trowald, N. (1997): *Uppfattningar om examination - en intervjustudie av högskolelärare*, (in Swedish), Högskoleverket, Sweden.
- Wiiand, T. (1998): *Examinationen i focus*, (in Swedish) Rapport nr 14, enheten för utveckling och utvärdering, Uppsala University, Sweden.