Assessing Class Performance and Progress using Grade Self-Estimation in Undergraduate Embedded Systems Courses

Militina Gorobets

Electrical and Computer Engineering
University of Calgary, Canada
mgorobet @ ucalgary.ca

Michael R. Smith

Electrical and Computer Engineering
University of Calgary, Canada
mike.smith @ ucalgary.ca

Emily Marasco

Electrical and Computer Engineering
University of Calgary, Canada
eamarasc @ ucalgary.ca

Abstract –*Embedded systems courses are typically taught with a heavy emphasis on laboratory experience using sophisticated software development tools. There is a considerable mismatch between practical and theoretical knowledge associated with a steep initial learning curve, leading to an initial bi-modal grade distribution. We introduce a self-assessment philosophy to enable the students to use a process improvement technique in order to better understand how to connect their theoretical and applied knowledge, and identify areas for improvement. Students were found to overestimate their performance early in the semester; a tendency that was considerably corrected by the term's end. We evaluate student performance on a number of course components using student self-analysis during final exam questions, as well as progress throughout the term exams.*

Keywords: self-assessment, undergraduate, engineering, embedded systems, performance

1. INTRODUCTION

Embedded systems courses are not typically taught in the same manner as most other undergraduate courses. There is a much heavier emphasis on laboratory experience using an integrated development environment (IDE) as that is the best way to reinforce concepts [6]. For some students, there will be a steep learning curve associated with their first exposure to the complexities of an IDE, where practical understanding of the concepts of registers, interrupts and assembly language programming is needed. On the other hand, others will already have become experienced in IDE usage from previous exposure to microcontrollers through hobbies or University Clubs [1].

As a result of such wide range of prior experience, it can be anticipated that many students will be initially discouraged when they find themselves in the bottom node of a bi-modal grade distribution. This is in complete contrast to the instructor's expectations based on previous experience of teaching such courses. Smith [6] reported that the hands-on aspects of embedded systems courses invoke a high level of interest amongst engineering

students. This typically leads to a convergence to a normal distribution, with a high mean, by the course end.

Smith proposed to overcome the initial student discouragement through the use of an earned mark analysis (EMA) spreadsheet [6]. This is a modification of the earned value analysis (EVA) process management tool used within Humphrey's Personal Software Process (PSP) [3]. EMA is designed to allow the students to track their performance and provide a final mark prediction through a combination of data based on the student's own experience in previous courses not related to embedded systems and the instructor's knowledge of student performance in the current course. Smith [6] provided anecdotal evidence on the success of the EMA approach to encourage students while they overcome the initial steep learning curve associated with senior year embedded systems courses.

Other authors have provided quantitative results on the benefits of self-estimation (even when not directly related to the recorded grade via bonus points). Self-assessed performance is intended to provide deeper understanding of the student's own strengths, progress in the course and identify possible knowledge gaps. Sadler and Good [4] saw that self-estimation improved student learning in their study of grade 7 students. Falchikov and Boud [2] have created an overview study of various approaches to self-assessment and related outcomes for university students. The study incorporated the results of previously published papers and led to important observations. Most studies involved in the overview found that students tend to overestimate their own performance, although experienced students in more advanced courses provided more accurate grade estimates [2, 4].

This paper is organized as follows: first, the philosophy, methodology and the proposed metrics for evaluation of the accuracy of self-assessment within an embedded systems' course are described. This section also discusses the trends and relationship we anticipated seeing in the data. The combined results and discussions section follows, presenting data gathered from two years of embedded systems' classes (referred to as Year 1 and Year 2). The section is split into subsections based on the proposed metrics. Finally we conclude with some general observations and suggestions for future work.

2. METHODOLOGY

In this paper we examine the use of self-assessment in a technical course, *Assembly Code and Interfacing* [8]. This is a required third year course for students in our software engineering degree or computer engineering minor program. It also serves as a technical elective for fourth year students undertaking an electrical engineering degree. Students use knowledge from the course for many Canadian Engineering Accreditation Board (CEAB) final year design projects (Capstone) and in elective courses involving embedded system concepts.

The course philosophy is to use self-assessment as the first stage in providing a process improvement overcome in what we believe to be the three types of knowledge gaps present in a university-level technical course:

- What the students do not know;
- What the students believe they understand, but don't. This possibly generates overconfidence, leading to less review, damage of equipment and the potential of personal harm in a laboratory environment.
- What the students believe they don't know, but do. This leads to unnecessary relearning of knowledge and under-confidence in the use of tools designed to decrease design and development time.

The following summarizes the different approaches used to permit student assessment in ENCM511.

- *Earned mark analysis (EMA)*: At the start of the term, students generate an instructor prediction of their final grade based on their performance in laboratories and exams relative to the class averages in other course [6]. This prediction is updated by the students as their actual performance in laboratories, quizzes and exams becomes known.
- *In-Exam Estimation*: Students get the opportunity to estimate their grade for questions during the exam.
- *Post-Exam Estimation*: Students submit their grade estimate after receiving the instructor's solutions to the problems on leaving the exam room.

In order to improve the motivation to participate in grade estimation during the exams and reduce blind guessing, bonus points were provided if the student provided an estimate that was accurate to within 10% of the instructor's mark. The bonus constituted approximately 10% of a short in-term quiz / exam mark and 5% of a longer final exam.

The students were encouraged to make changes in their personal development practices on the basis of certain observations that can be made based on the estimation outcomes:

- Over-estimation might indicate that they did not understand the concepts as much as they had thought.
- Under-estimation might indicate they were spending more time relearning material they actually knew.

2.1 Proposed metrics

Overall midterm examination metrics: The midterm, being typically the first major exam during the semester. It can be a good indication of what the student expectations are, and whether they are realistic given their current knowledge of the material. The initial quizzes and midterm were deliberately delayed until later in the term compared to other courses to overcome issues of a steep initial learning curve associated with embedded systems.

Overall final examination metrics: Final examinations typically consist of a variety of questions covering areas ranging from basic knowledge to fairly open-ended design questions. The estimation is done in-exam.

Individual final examination question metrics: Given the sheer scope that a cumulative final examination has to cover, it is possible that certain questions will be done better than others. Likewise, certain questions will create the distinction between the stronger students and the rest of the class. The students' performance relative to their estimations can provide a valuable outlook of the question difficulty and appropriateness. Student self-assessment has repeatedly proved a useful cross-check for instructor marking of individual open ended questions.

Midterm and final exam relationships: It should not be unexpected that the individual's response to their performance on a midterm exam can govern their future grade estimations in the class. With this in mind, we quantitatively investigate the relationships between performance on the midterm and the final.

3. RESULTS AND DISCUSSION

Due to space limitations, only select data from two different ENCM511 classes are presented. This secondary use of anonymized data was made under the University of Calgary Conjoint Faculties Research Ethics Board approved protocol. All data has been normalized to show percentages. The bubble plot circles are proportional in size to the number of students in the group. As discussed in [6], it is important for students to understand the instructor's planned relationship between scores and GPA, especially when that relationship differs from that in other courses. For ENCM511, the score / GPA relationship was fixed at the beginning of the term as

$$\begin{aligned} A+ &> 87 > A > 83 > A- > 80 \\ 80 &> B+ > 76 > B > 73 > B- > 69 \\ 69 &> C+ > 65 > C > 60 > C- > 55 \\ 55 &> D+ > 50 > D > 45 > F. \end{aligned}$$

3.1 Overall comparison between midterm and final exam estimation accuracy

Figure 1 shows the histogram representation of the Year 2 class grade distribution. At this point in the term, the students have only engaged in laboratories (in which most will have been interested enough to put in the necessary effort to complete and score full marks) and a single quiz. The actual midterm mark distribution, Fig. 1A, shows two distinct peaks at 50-55% and 70-75% with no student mark above 90%. This reflects the bimodal distribution seen in past years [6]. Figure 1B shows students' estimations significantly skewed towards higher grades.

The final exam takes place after all laboratories and shorter exams (midterms and quizzes) have been completed. At this point, the students should be more aware of their relative strengths and weakness. It can be anticipated that the students will be able to play to their own strengths given that the final exam is longer than the midterm, and provides question choices. This is born out in Figs 1C and 1D, actual and estimated final exam marks respectively, with distributions occupying nearly the same region with approximately the same distribution shape. There still appears a small peak around the 50% point, but the majority of students are now situated between 65% and 85%, with some even reaching as high as 95%.

Figure 2 provides a more detailed comparison of actual and estimated marks for the midterm and final examinations for Year 1 and Year 2. Here, groups appearing below the diagonal represent overestimation, and those above are underestimation.

Students had a tendency to overestimate their own midterm performance in both analyzed years. In Year 1 (Fig 2A) more students estimated grades in the region 80-

100% than in Year 2 (Fig 2B), which could be a sign of an easier quiz that year created a skewed comparison. In Year 1 (Fig. 2C) as many students overestimated their final exam marks as overestimated their midterm marks. In Year 2 students made a far better job at estimating their final marks than their midterm marks, Fig. 2D. This may be a reflection of the difficulties both the students and the instructor experienced in a year where a transition was made to a new embedded system IDE, leading to a lower, more realistic student evaluation of their performance.

Figure 3: Individual final exam questions predictions and actual grades (Year 2). Diagonal lines represent the split between overestimation and underestimation. (A) Question 1 – Basic Knowledge, (B) Question 2 – Lab Experience, (C) Question 3 – Testing, (D) Question 4 – Extended Concepts, (E) Question 5 – Mandatory Design Question, with (F) shading information to allow comparison to high (dark) or low (light) final course grades..

3.2. Detailed Analysis of Estimation Accuracy across Final Examination Questions

The final exam consists of different questions which aim to test various aspects of the students' knowledge. The first question (grades for which are shown in Fig 3A) was intended to test the basic knowledge acquired in the course. The answer required some logical thinking and written answers, but not knowledge of obscure or detailed content. As such, it would be expected that the class

would do well. However, we see that most students' expectations varied from reality by a fair margin. For instance, majority of those who expected to get grades between 90% and 100% in reality scored below 70%.

Possible conclusions drawn from this observation are:

- Part of the question (not all, as most students still achieved >50%) could have in fact covered material more obscure than had been intended by the instructor who composed the exam e.g.. based on an important part of the lectures that was only briefly addressed and not reinforced;

- The question component contexts were interpreted differently from that intended;
- The (graduate) markers assigned to mark these short exam question components were unclear on how to award fractional marks for the wide range of partially correct answers;
- The students did not consider these basic concepts important (examinable) when preparing for the exam.

With the second question (Fig 3B) we see a good agreement between estimates and actual grades. There is an almost equal split between underestimation and overestimation, even for weaker students. This question was based on laboratory sessions, where the problems required the students to effectively identify their weakness and gaps in knowledge in order to find a solution. As such it is logical to expect that the student would likely be aware of his / her abilities in that area, and be capable of generating a reasonable mark estimate.

The third question (Fig 3C) addressed embedded Agile testing, a software engineering concept that the instructor had adapted and adopted for both research and teaching purposes in a computer engineering, embedded system environment [5, 7]. The performance in terms of estimated/actual grades isn't nearly as focused around the line splitting underestimation and overestimation as it was for question two. At the same time, there doesn't appear to be any distinct grouping of students in one region or another – rather the population is scattered. For instance, students who estimated their performance to fall into the range between 70% and 80% in reality were between 40% and 100%. Since this question addresses the more complicated topic of software testing in an embedded system context, and because the concept was required for several laboratories, we would have expected to see better estimation quality from the class as a whole. The results for this question are also different from question one (Fig 3A) in that there isn't a clear trend to under- or overestimation, in fact it appears more as guessing. The possible reasons here include:

- Unfamiliar context – the students might simply not have known how to approach the problem in the given setting, resulting in blatant guessing of their own performance;
- Lack of the experience that was expected to have been acquired. Perhaps the in-laboratory experiences required only very basic knowledge of preparing software tests, and few students went above to learn more. There would then be students who missed key points, damaging their overall performance.

In question four (Fig 3D), we see a similar behavior as in question three. Here the spread between predicted and actual grades is fairly large. However, extended concepts (some of the more obscure / complex embedded systems topics) were tested here, and thus we would expect variations in terms of student confidence and their predictions (stronger vs weaker students in particular).

The fifth question (Fig 3E) was a design question, testing the understanding of the “big picture” required in embedded systems. For the first time, we see a focused population appear near the top right of the plot (80-100% range). Figure 3F is a re-interpretation of Figure 3E using the x-y coordinates to show actual and estimated performance for individual students on this question. The shading represents a third axis indicating the student's overall final exam performance – heavy and light shading representing high and lower final exam performance respectively. The concentration of dark shading in the top right corner is taken as representing the fact that strong students in the class can be expected to do well on design specific and general topics in the course. There is far more scattering of the student population in terms of estimated and actual performance in the region of 30%-70%, for students in almost all grade brackets (again, see Fig 3F).

3.3. Comparison of Estimation Performance during Midterm and Final Examinations

Figure 4A shows that students in the Year 2 class of this embedded systems' course estimated their midterm grades higher than the final. This, of course, makes sense since we discussed in Section 3.1 the overall overestimation that was evident in the midterm. Figure 4B looks at the effect of the actual midterm grade (which most students overestimated) on the estimation during the

final exam. There is a slight trend towards students generally estimating their final exam performance to be worse than the actual midterm grades. This could be interpreted as most students perceiving final exams to be more difficult (due to the increased scope, length and complexity of questions) than midterms, and hence assuming worse performance. We could also link this to students overcompensating for their initial overconfidence in their own performance during the midterm.

4. CONCLUSION AND FUTURE WORK

The presented work demonstrated and discussed data gathered over two years from a third year (senior) introductory course to embedded system interfacing. The goal was to analyze the effects of student self-estimation in examinations on the overall class performance and progress. We investigated trends and relationships between student self-assessment during midterm and final exams and on different course components.

The investigated metrics were intended to provide an outlook on the relative difficulty of the exams and questions to each other, and the students' sense of preparedness. We observed the commonly reported tendency of students to overestimate their grades, in particular early on in the semester (i.e. midterm exams in this study). We also saw that students self-assessed their final exam grade to be lower than their actual midterm grade, perhaps speaking of overcompensation for their overconfidence of their midterm mark.

In general, the actual student grades converged to a skewed normal-like distribution in the final exam, as compared to the fairly bimodal one appearing in the midterm. This convergence was quantified by an increase in the mean, meaning that the weaker students improved during the semester.

Analysis of individual final exam questions showed that the class not only performed relatively poorly on a straightforward question – but also that the students thought they did well. On the other hand, it provided good feedback for gained lab experience, though not necessarily the more complex aspects of it (Agile development). We also saw that stronger students performed better at the design questions, which could be related to them being more willing to invest more time into preparation and understanding of the lab exercises. Certain questions, on the other hand, offered no particular pattern or distribution, leading to a conclusion that students were guessing their performance.

Some future work can be done to investigate gender difference in terms of estimation quality and abilities. However, the vast majority of students in electrical engineering are traditionally male, which means that performing gender-dependent studies would not

necessarily yield statistically significant results. There exists potential to extend the study into lower years or even other engineering departments in order to gather gender-related data on self-assessment during heavily laboratory oriented classes.

Acknowledgements

M. Gorobets was supported through scholarships from the Canadian Natural Sciences and Engineering Research Council (NSERC) and from Alberta Innovates Technology Futures. M. Smith was funded through an NSERC Discovery grant, the University of Calgary and the Analog Devices University and Teaching Support program. Thanks to Dr Behjat, University of Calgary, for assistance and advice in preparing this paper.

References

- [1] Engenuics. (2014). *MPG – The microprocessor group* [Online]. Available: <http://engenuics.com/mpg/>
- [2] Nancy Falchikov and David Boud, "Student Self-Assessment in Higher Education: A Meta-Analysis," *Review of Educational Research*, vol. 59, no. 4, pp.395-430, 1989. Available as of April 6, 2014 from <http://www.jstor.org/discover/10.2307/1170205?uid=3739392&uid=2&uid=3737720&uid=4&sid=21103648300181>
- [3] Watt S. Humphrey, *A Discipline for Software Engineering*. Addison Wesley Professional, 1995. {ISBN 9780201546101}
- [4] Philip M. Sadler and Eddie Good, "The Impact of Self- and Peer- Grading on Student Learning," *Educational Assessment*, vol. 11, no. 1, pp.1-31, 2006. Available as of April 6, 2014 from <http://www.cfa.harvard.edu/sed/staff/Sadler/articles/Sadler%20and%20Good%20EA.pdf>
- [5] James Miller and Michael Smith, "A TDD Approach to Introducing Students to Embedded Programming," in *Proc. 12th Annual Conference in Innovation and Technology in Computer Science Education, ITiCSE*, (Dundee, Scotland, U.K.; June 25-27, 2007), pp. 33-37, 2007.
- [6] Michael Smith, "Student Prediction and Tracking of Learning Progress during an Assembly Language Programming Course" *Proc. American Society of Engineering Education Conf, ASEE 1999*, (Charlotte, North Carolina, USA; June 1999), 1999. Available as of April 6, 2014 from [http://www.asee.org/search/proceedings?fields\[\]=title&fields\[\]=year&search=Student+Prediction&commit=Search](http://www.asee.org/search/proceedings?fields[]=title&fields[]=year&search=Student+Prediction&commit=Search)
- [7] Michael Smith, James Miller, Lily Huang, and Albert Tran, "A case for taking a more 'Agile' approach in the development of embedded systems," *IEEE Software Magazine*, pp. 50-57, May 2009. Available as of April 6, 2014 from <http://www.computer.org/csdl/mags/so/2009/03/mso2009030050-abs.html>
- [8] Michael Smith. (2013). *ENCM511 Home Page – 2013* [Online]. Available: http://www.enel.ucalgary.ca/People/Smith/2013webs/encm511_13/