SO, STUDENTS HAVE BEEN GIVEN FEEDBACK, NOW WHAT?

Trond Morten Thorseth, Magnus Strøm Kahrs
Norwegian University of Science and Technology (NORWAY)

Abstract

Formative assessments or “assessments for learning” are challenging to perform in higher education, given the lecture form and the number of students present. We have developed and tested an assessment procedure where large student groups were given a two-hour assessment with immediate feedback, six times throughout one semester. However, the consequences of this immediate feedback are largely unknown, beyond making the students realise any misconceptions towards the specific assessment items. Ideally, formative feedback should provide each student with information on how to proceed, both with regards to content matter and to their own learning strategies. Our statistical analysis suggests that we are far from reaching this goal.

We have followed 109 students through the first introductory mathematics course at the bachelor engineering education at NTNU. By using an in-class assessment procedure, we gave students regular feedback on their progress throughout the course. In addition to working on and responding to questions during the class assessments, the students reported an expectancy measure before each assessment and reported the degree of satisfaction after each assessment session. The students also reported the hours of effort spent on the subject the last week before the assessment. They were also challenged to reflect over their own individual learning process.

The six assessments, as we have performed them in class, predict only 58% of the final exam results. Measurements of effort do not correlate with the result from each assessment. However, we see a weak but significant correlation for the sum of self-reported effort with the results from the final exam, .293 significant p=0.04.

We conclude that good assessments are not enough if we don’t provide information about how to improve learning strategies. Both learning of the subject and acquiring new strategies for learning takes time to master and hence require continued effort from both students and academic staff beyond a single semester.

Keywords: Feedback, reflection, learning strategies, change strategies, digital assessment, mobile technology.

1 INTRODUCTION

Social cognitive theory is grounded in the idea that learning and behaviour occur in social environments [1] [2]. Social cognitive theory of self-regulation [3] and in particular the theory of self-regulated learning [4] emphasizes he learner as an agent in the learning process [5]. Social cognitive learning theory distinguishes between active learning and vicarious learning. When a student follows lectures, he or she passively observes demonstration of knowledge, what Bandura calls vicarious learning. Observational learning through modelling, occurs when students observe a teacher lecture subject of math [6]. Students taking a course in mathematics, learn the subject from a teacher lecture, a book, or an online video lecture. To learn the material properly requires further work with active learning where the students must engage with the material. But they learn how to study by modelling others, regarding how to use the information provided, attitude towards the course and the lecturer, and how to interpret the feedback given either in class or during mandatory exercises.

For a student to be able to demonstrate knowledge in a final exam, he or she needs to have encoded new information, stored the information, and finally accessing the material through what is called a retrieval process [7]. It is quite common that our students do this encoding of new learning, only by attending lectures. To improve their learning, we have mandatory exercises that provide active learning. It has been demonstrated that the process of retrieving information from memory strengthens the memory of the retrieved information [8, 9]. A possible reason for this so-called testing effect is the effort put in retrieval, the amount of processing that is required to find, retrieve, and make sense of the retrieved information to solve the problem. It is known that the amount of effort in the retrieval strengthens the effect and hence a full problem solving test will work better than a multiple choice [10].
The effect of practice testing has been shown to be higher in mathematics than other subjects where this effect has been measured [11].

Feedback is considered a central part of learning. In the learning process, feedback is a consequence of a student’s performance [12], but not as an end product but as part of the learning process. The delivered response to a production [13] can be any information that the learner can use to confirm, add, overwrite, tune, or restructure information in memory. The idea is to provide information to bridge the gap between actual performance and the learning goals. One part is to provide feedback on the subject being taught. The other aspect of good feedback practice is to provide feedback on the process of learning [14].

The main idea of our assessment system is to efficiently provide formative assessments. Pele is a digital assessment system where the students respond using their own device to respond. With the system, we can utilize the testing effect, provide immediate feedback, and provide various forms of feedback both on the subject and the learning process. Verification of all responses are considered important, since feedback improve retention even if the problems has been answered correctly [15]. For questions that have been solved correctly by the majority of the students in the class, we give feedback by confirming correct answers. Given problems that have caused challenges for the class, a more thorough review of the correct solution is called for by the teacher.

However, understanding how the assessments fits into the students learning process, how to interpret the results and what advice to give the students is not immediately easy. With the electronic assessment system and the implementation of practice, we aim at providing feedback in such a way that the test acts as a formative assessment of student learning. Here, we want to explore the results and try to understand how we should advise our students in how to use the given feedback. In this paper, we primarily want to understand how our form of assessment fits with the students’ learning process. Can the assessments in the form given, predict the final results?

2 METHODS

This study was conducted in a first-year engineering mathematics course at Norwegian University of Science and Technology (NTNU). Participants were 109 logistics, material, and chemistry students, studying for a bachelor engineering degree. The subject was introductory calculus given with mainly lectures and mandatory exercises. However, approximately every second week, a mandatory in-class assessment was conducted, six in total.

The classroom exercise was a two-hour event where the first hour resembled a classroom test. The students got a set of typically six problems they had to solve. The assignment was handed out on paper. 50 minutes into the assessment, they got a second paper with possible response alternatives. They got the last five minutes to submit their results digitally using their own smartphones, after which the students had a 10-minute break. Results were gathered with the One2Act software Pele developed at NTNU. This digital assessment system allows the teacher to control a set of webpages where the students can submit their answers asynchronously. As a controller of the assessment, the teacher can immediately see how the class has performed and start planning how to review the results in a joint discussion. Dependent on the students’ submitted responses, the teacher should decide on what way to provide feedback. A method of feedback that can potentially generate the most cognitive activity should be preferred. A typical procedure has been to start with the problem causing most problems for the students. If the problem is complex and the correct score is low, a review of the solution might be one alternative. If more than half of the class has given a correct answer, a peer discussion might be a good alternative for learning, rounded up with a second submission for the task at hand. Being able to resubmit new answers after a discussion, allows for a double feedback; a peer feedback and a verification from the teacher at the end. Sometimes the problem is easy and a small hint might be the thing that the students need to solve the problem. The aim of the second part of the classroom assessment is that the students get immediate feedback on the problems they have worked on. For each problem that the teacher reviews, students get some minutes to compare with their own solution. Students are encouraged to reflect over why they responded as they did, and what misunderstandings they themselves have identified.

Before the assessment, we asked each student about their self-efficacy on specific subjects. After they had made their submission, we also encouraged a post-assessment reflection towards the learning process in general. Here, the students were asked about the degree of satisfaction with the
On order to be allowed to use personal data from the final exam, we applied to Norwegian Centre for Research data (NSD Project 43993). From the teacher, we got data on each student about each problem and the final degree. For each candidate, we have a score from 0 to 1.0 that was translated to a final mark by the teacher.

### 3 RESULTS

#### 3.1 National pre-knowledge test

As a part of this study, 79 of the 109 students took a national pre-knowledge test in mathematics, a test given nationally by the Norwegian Mathematical Council very second year. Our students scored on average 21.5±6.3 which is close to other engineering students nationally 21.95±0.54 [16].

#### 3.2 Self-expectations

Before each assessment exercise, the students were asked about their expectations for the assessment. "The assessment today contains two problems regarding differential equations. How sure are you that you will be able to solve this problems today?". The students responded on a scale from "Very sure" (coded as 1) to "Not at all sure" (coded as 5). This measurement was conducted on assessment 2 to 5. The reported values are similar for all the assessments but the level of insecurity was slightly larger on assessment 4. The self-expectation data were significantly internally correlated, except for assessment 6, which was the last assessment before the final exam.

This expectancy measure is some sort of self-efficacy measure and the data shows an internal consistency. Students who are unsure on assessment 1 are in a consistent manner sure on assessment 5. However, this internal consistency might be an indirect signature of the consistent way of working with the material and hence a consistent self-efficacy. What is interesting to notice is that the self-expectations and assessment results are uncorrelated, except for the third assessment where the correlation is significant 0.378 p=0.001.

#### 3.3 Pele test

Six in-class assessments were conducted throughout the course. The student could score from 5 to 11 points that varied according to the number of problems they had to solve. The variables A1 to A6 contains the score on assessment 1 to assessment 6, respectively. In all only 36 of 109 students participated in all the tests.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>88</td>
<td>.0</td>
<td>6.0</td>
<td>2.261</td>
<td>1.5865</td>
</tr>
<tr>
<td>A2</td>
<td>88</td>
<td>.0</td>
<td>6.0</td>
<td>3.557</td>
<td>1.4532</td>
</tr>
<tr>
<td>A3</td>
<td>78</td>
<td>1.0</td>
<td>6.0</td>
<td>3.474</td>
<td>1.0410</td>
</tr>
<tr>
<td>A4</td>
<td>81</td>
<td>.5</td>
<td>11.0</td>
<td>6.753</td>
<td>2.6126</td>
</tr>
<tr>
<td>A5</td>
<td>89</td>
<td>.0</td>
<td>7.0</td>
<td>3.360</td>
<td>1.1604</td>
</tr>
<tr>
<td>A6</td>
<td>55</td>
<td>.0</td>
<td>6.0</td>
<td>3.582</td>
<td>1.3429</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 3.4 Self-reported effort

The students were asked: “How many hours have you spent working with mathematics in addition to mandatory activities?” This question was asked together with assessment 2 to 5. The number of reported hours varies from 0 hours to 20 hours. Average reported values are between 2.4h to 2.9h as shown in Table 2.
Table 2. Descriptive statistics for the self-reported effort before assessments 2 to 5.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>85</td>
<td>.0</td>
<td>20.0</td>
<td>2.461</td>
<td>3.0161</td>
</tr>
<tr>
<td>E3</td>
<td>62</td>
<td>.0</td>
<td>20.0</td>
<td>2.944</td>
<td>3.2820</td>
</tr>
<tr>
<td>E4</td>
<td>61</td>
<td>.0</td>
<td>10.0</td>
<td>2.533</td>
<td>2.5898</td>
</tr>
<tr>
<td>E5</td>
<td>68</td>
<td>.0</td>
<td>8.0</td>
<td>2.412</td>
<td>2.1317</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The distribution of the effort data shows that a significant number of students rely on following the lectures and do the mandatory work. A lot of the students report 0 hours of effort. Of the self-reports here, 16% report 0 hours, 36% do not report anything, and 46% report spending between 1 and 20 hours of effort.

Figure 1 shows a histogram of all the self-reported number of hours spent working with the subject beyond mandatory activity.

We see from Figure 1 that 78 of the self-reported hours of effort, about 16% percent, report not doing any effort beyond mandatory work. We also have one student reporting an effort of 20 hours. The same student reports a similar effort before test 3. Later this single student reports 3 hours on test 4 and 0 hours on test 5.

If we look at internal correlations in the effort variables, there are significant internal correlations. There is a significant correlation between effort between test 2 and 3 (0.58, p<0.01), between test 2 and 4 (0.379, p=0.03) and between test 3 and test 5 (0.479, p=0.01). Even if the correlations are weak, they are significant and hence indicate low consistence in work pattern. If we look at the correlations between reported effort the week before the assessments and the subsequent assessment scores, there is one case where we have a weak correlation .253, barely significant p=0.049. In all the other cases, we find no correlation between the self-reported effort and the assessment scores.
When we look at the self-reported effort vs results on the final exam (Figure 2) we see two candidates with more than 90% score on the final exam. But we see that the accumulated self-reported efforts were quite different. We also see students reporting relative high effort but still fail (score below 0.35).

The one candidate that has reported the highest effort can be treated as an outlier in this analysis since that point has a strong influence the calculated correlation. If we treat the accumulated effort of 44 hours, as an outlier and recalculate the correlation, it gets weaker; 0.221 but still significant p=0.034.

### 3.5 Results from the exam

The pre-test was correlated to the final exam. We find a weak correlation .235, p=0.04 between the final exam and the pre-tests. When we correlate the assessments with the final exam, we see (Table 3) that each assessment has correlation to the exam.

![Cumulative effort in hours versus results at final exam.](image)

Table 3. Correlation between assessments A1 to A5 and the final Exam.

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam</td>
<td>.486**</td>
<td>.470**</td>
<td>.347**</td>
<td>.582**</td>
<td>.338**</td>
<td>.404**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.002</td>
<td>.000</td>
<td>.001</td>
<td>.002</td>
</tr>
<tr>
<td>N</td>
<td>85</td>
<td>84</td>
<td>75</td>
<td>80</td>
<td>86</td>
<td>55</td>
</tr>
</tbody>
</table>

If we add up all scores from all assessments and look at the correlation with the final exam it is 0.762, p<0.001.

Figure 3 shows how exam results are correlated to the accumulated assessment scores. Given the correlation of R=0.762 p<0.001 we see some candidates that have got the message that “this seem to work” from the assessment feedback but fail on the final exam (marked with red circles). Some students have similar score on the assessments and have got top grade on the exam (marked with blue circles).
Figure 3 shows the assessment score versus the final exam. Both scales are normalized.

If we examine correlations between the self-reported efforts and the exam, we see a weak correlation between effort on the second test and the exam. The correlation is weak 0.232 but significant p=0.033. For all other assessments, we find no correlation between self-reported efforts and the final exam. However, when added together all the reported hours of effort and correlate the accumulated hours of effort with the exam, there is a weak 0.293 but significant p=0.04 correlation between self-reported cumulative effort and the exam. 8% of the results at the exam can be explained by the reported effort.

4 DISCUSSION

When we look at the correlation of the in-class assessments and the final results we see that the knowledge assessed during the course, explain only about 50% of the final results. The self-reported effort is close to insignificant and the pre-test explain 5% of the result. How can we understand this and how do we help students to understand the kind of feedback that they have been provided?

We must consider the fact that when students perform this assessment, they are at various stages of learning the material. From our data on self-reported effort, we infer that some students have little or no practical interaction with the material, beyond spending time in lectures trying to understand what the teacher explains, what Bandura [5] calls vicarious learning. The assessments are meant to be a significant part of the learning process that will add to the level of understanding that the students gain at the end.

Another important factor that we have to consider is that we know that a lot of learning happens in the last weeks before the exam. This intense effort in the last week is for some students the most important part of the learning period where the material is matured and is fitted together. The results show that at this stage about 50% of the learning is still to come.

Given the degree of uncertainty that the students report, is just another reminder that feedback is essential. It is also a reminder for the teacher that the time working through the problems and giving feedback is essential and useful for their own learning, even when they answer correctly. This degree of uncertainty is a nice reminder to the students that this assessment is a significant part of their active learning of the material. The variable degree of certainty is also a clear indication that each assessment is testing the material in an early stage of the learning process. Few students have the sensation of having learned the material at this stage and hence the need for feedback is essential.

We try to make the assessment itself a significant part of the learning process by introducing active learning with immediate feedback in an efficient way. Explaining how to work and how to use these tests are essential to motivate students to participate in these mandatory assessments for the sake of
learning. A lot of our students report what we can call an automatic stress reaction to any activity that resembles a test situation. Working with the meta-knowledge of learning and the testing effect is therefore one way to go. From data not presented in this paper, we have reason to believe that there is a need for strengthening students’ meta-knowledge of learning.

It is often assumed that failure in a course can be explained by a teacher not providing adequate feedback [17]. In our opinion it is also the issue that students have a variable knowledge and skills to recognize, understand and reflect upon the feedback given. As stated by Higgins, Hartley, & Skelton [18] feedback is also about how the students make sense of the feedback that they are provided and in the next step actively engage with the feedback [19]. We see that only providing feedback is not enough. We have to train the students in how to use the feedback. However, even though we provide feedback in the way we do, they seem to rely on the first exam, as one student said, as the first “real feedback”. We suspect that the kind of feedback without consequences are not taken as real, and hence we can assume that they have not understood how to utilize the provided feedback. During this course, we have given lectures about learning and how feedback fits into their learning process. But the only way that these assessments differ from tests is the immediate feedback, work-through and the continuous nudging about the learning cycle. They are asked to notice reactions to the assessment, the results, to monitor their way of thinking during the assessments and the reactions to the results. They are also asked to write down goals for the next week’s work. The results to these reflective questions however, portray a group of students who are not willing to enter the level of reflection about their own learning needed to make any change.

Self-reports on effort is also a questionable measure. An hour of learning effort may vary from student to student; reading, intense problem solving, self-testing or discussing with peers (or a combination of these together with frequent visits on social media). The effectiveness of these different activities, and hence the learning outcome, may vary. This makes the intrinsic uncertainty of this measure hard to correlate to other factors such as the test results and the final exam. However, the general results were that the accumulated effort was related to the final exam but the variation is large and the correlation is weak.

Finally, we have to consider the assessment context. We suspect that for some students the assessment conditions that we provide are not perceived as formative. We see this from the emotional reports that the students have provided together with the questions added for reflection on the learning process. A lot of the students associate problem solving within a limited time as a (summative) test. Instead of trying to assess themselves and rely on the testing effect and the learning that may come with it, some strive to submit the right answer, even to the point of cheating, and thus lose both the possibility of appropriate feedback and learning from the testing effect.

It may be that the assessment context, including the feedback provided, shares more similarities with a summative rather than formative assessment. The only actual feedback the students got from the teacher was an individual score. If specific test items turned out to be difficult for about half the class, the students got to discuss their own solution with their peers, thus offering an opportunity for more direct feedback. However, all other feedback was only indirectly accessible, as the students were encouraged to write comments on their own submission sheets when the teacher reviewed the solutions to the different test items. Ideally, formative feedback should provide each student with information on how to proceed, both with regards to content matter and to his or her own learning strategies. From this, there seems to be a need to critically examine both the content, form, and context of the feedback provided to the students, if we are to provide the students with sufficient information on how to proceed.

The students used their own smartphone to respond to the system. This meant that various means of communicating could be used during the assessments. When students apparently had solved the problems, and scored high some were discontent with their own results. When we took a closer look, we could see that the submitted answers were not in agreement with what they had worked out on paper. We interpret this as an indication of students not having understood the purpose of the assessment. This assessment was not about submitting the right answer, but testing their own learning and retrieving feedback on the learning process.

We believe that this kind of assessments requires a significant effort in teaching how to use the feedback provided. If the feedback is to be used, it has to fit into a general goal striving effort. However, some students are not regulating their effort and hence the feedback does not fit in. We see some students struggle to understand the conditions and hence the purpose of a formative assessment. It requires some effort from the students and the teachers to fully understand their own
learning process and how this feedback fits in longer perspective. We saw that spending time to reflect over the observed automatic stress reactions and being able to verbalize the reactions, made some students aware reactions and some learned to cope with them.

5 CONCLUSION

We must understand what these assessments do to the students' learning and how the feedback is both provided by the teacher and used by the students. If we want the students to use the feedback they have to both perceive and understand the feedback effectively. We see that some students get feedback that tells them that "you are doing ok!" and later fail. There is a lot of effort needed to teach the students how to learn effectively and in general how they could use feedback. A significant part of the training must be to explain how the assessment and the testing effect is effective in their own learning. As it is now, a lot of students do not seem to understand this. The perceived effort and correlations with the final results show that the results are not immediate. It is easy for students to make their own perception of the realities, what works and what does not.

ACKNOWLEDGEMENTS

This work has been financed by NTNU in the project “Grep om egen læring”, 2014-2016, project number 20180.

REFERENCES


