Student Self-assessment in a Programming Course Using Bloom’s Revised Taxonomy

Satu Alaoutinen
Lappeenranta University of Technology
P.O. Box 20
53851 Lappeenranta, Finland
satu.alayoutinen@lut.fi

Kari Smolander
Lappeenranta University of Technology
P.O. Box 20
53851 Lappeenranta, Finland
kari.smolander@lut.fi

ABSTRACT
Making programming teaching meaningful is a challenge for a teacher. This study presents a simple student self-assessment tool, that can be used to motivate learning and to follow progression. The tool is a survey questionnaire, that uses Bloom’s Revised Taxonomy as a base of its scale. The results show, that the students can quite well place their knowledge to the Taxonomy based scale, and they feel that it could help their learning. It also tells the teacher the level of knowledge gained more objectively than the general scales. The scale is designed for a programming course, but with modifications it can be used in other classes too.

Categories and Subject Descriptors
K.3.2 [Computers and Education]: Computer and Information Science Education—Self-assessment

General Terms
Experimentation

Keywords
Self-assessment, Bloom’s Revised Taxonomy, Programming

1. INTRODUCTION
In our university students are being pushed to finish their studies faster and programming, along with mathematics, has been a topic that causes problems for many students. In many cases basic programming is the last course that keeps them from graduating, while it was one of the first courses taught in our university. We have been seeking ways to improve our programming teaching. The language taught in our fundamental programming course has already been changed, and the technical infrastructure improved, so now it is time to concentrate on teaching aspects [8].

One key issue in teaching is motivation of the students. By giving students goals to achieve and means to follow their progress, learning can be meaningful, especially in compulsory fundamental courses that just have to be passed. Recently, many universities have started to require learning outcomes [9] for their courses. Learning outcomes are based on the level of knowledge and define quite clearly what is expected from a student, so they can be used to set the goal for the students. Learning outcomes also give a base for course assessment [1].

One form of assessment is self-assessment. Studies show that when students evaluate their own or others work, their results get better [10]. Problem is, however, how well the students’ evaluations match their actual skills and knowledge. If the students could evaluate their own learning using a simple survey instead of a more complicated tool, more teachers could use it. The students could use self-assessment as a means of following their own development, and when combined with learning outcomes, to set clear goals for the course. We were interested to know, whether the students are able to evaluate their own learning and to recognize the corresponding level from the Bloom’s taxonomy. If they can do that, it would be easy for a teacher to follow the learning process, by asking the students to fill in a simple self-assessment survey a few times during a course. Combined with the course learning outcomes, the survey could give the students a way to follow their own development.

2. BLOOM’S REVISED TAXONOMY
In many studies the reference point or framework is the earlier mentioned Bloom’s Taxonomy [3] (referred later as the original Taxonomy) or its revised version [1] (referred later as the revised Taxonomy). This taxonomy describes knowledge hierarchically and cumulatively. The higher levels require mastering of levels below. The original Taxonomy defines three domains of knowledge (cognitive, affective, and psychomotor) and it separates domains into six levels. Since there was a debate if the order of the three highest levels was correct, a new revised version of the Taxonomy was created by Anderson et al. They combined the cognitive process dimension to the knowledge dimension, which consists of factual, conceptual, procedural, and metacognitive knowledge, as depicted in figure 1. Quite often the knowledge dimension is left unused and the cognitive process dimension is used.

The simplest form of knowledge is rote learning, but it is a foundation of deeper knowledge. There are always some facts that need to be remembered in order to deeper cognitive processes to take place. Similarly, things need to be understood before they can be applied. Creating is higher
level applying and it requires analyzing and evaluation skills. For example in a programming project, choosing a sorting algorithm need evaluation skills and finding reasons for program malfunction needs analysis of the code.

3. RELATED RESEARCH

Assessment is a vital part of an educational system, and has been studied quite much. Self-assessment is a good way to improve learning results and it also gives motivation for learning [10]. Despite of its effectiveness, self evaluation is not widely in use as an instructional tool. Reason for this could be the extra work it requires and the effort required to create a reliable self-assessment system. For example, Sung et al. [12] have developed a web-based self- and peer-assessment system. According to their study, the students are quite skillful in assessing, both own and others’ work. They report that by practice, the assessment skills improve, and the students were able to improve their work via assessment.

Some studies try to verify the reliability of self-assessment. Crews and Turner [5] compare the students’ telecommunications knowledge against their own assessment. They test the actual knowledge using 100 randomly selected multiple-choice questions from a course text book sorted into 5 dimensions of telecommunications: 1) networks and infrastructure 2) hardware and software 3) management and skills 4) performance and security 5) standards and protocols. The students evaluated their level of knowledge in every dimension above using a five step Likert scale from very low (1) to very high (5). A small sample size prevents any generalization, but a trend is that the students know more than they imagine.

The creators of the taxonomies, both original and revised, give examples of how to apply it. The examples cover many disciplines, but unfortunately computer science is missing. However, some teachers have created proposals of how to use them for example in programming. Thompson et al. [13] have created guidelines of how to apply the revised taxonomy in programming assessment. They give hints of what kind of questions are needed in which level, and some concrete examples. [6] proposes suitable programming tasks for the levels from comprehension to synthesis in the original Taxonomy.

Another study [4] maps software maintenance tasks to the original taxonomy, and presents how knowledge of software may differ at various levels of the cognitive process dimension. Xu and Rajlich [14] study the cognitive process during program debugging. They found that searching for a reason of an error requires all levels of cognitive knowledge in Bloom’s Taxonomy. Debugging has always been difficult for beginning programmers, and this explains why. They suspect that novices have to spent time on absorbing knowledge and understanding it.

In [7], Bloom’s taxonomy has been used as a reference point to make sure that all levels of thinking are included in programming learning activities for a new web-based programming learning environment that includes also a possibility for peer-assessment.

There are also other taxonomies for assessment purposes, which are more or less derivatives of the Bloom’s taxonomy, but they are not widely used. The SOLO-taxonomy [2] instead has been used in a computer science assessment, but it concentrates on the content of the responses instead of the cognitive performance. It is also difficult to apply in self-assessment, because it requires knowledge about correctness of own answers.

In summary, there are many tools for student self-assessment, but they are rather complex for our purposes. Some studies have used simple methods, like Likert scales for self-assessment. They are simple enough, but lack objectivity, because same things can have different meanings for different people (i.e. I manage recursion well). Students’ actual knowledge has been tested in different ways. A laborious way is to use special test questions, while an easier way is to use exam questions. For our purposes, using exam results as the verification method suits best. In order to determine the level of students’ knowledge, Bloom’s Revised Taxonomy seems to be an applicable method, because it makes the distinction between different levels of knowledge. In addition, there are experiences available of using it to study programming related processes.

4. RESEARCH METHOD

This is a survey study, which uses a web-based questionnaire. The study was conducted in one programming course, where 87 students were active. Students were from different departments and a mixture of freshmen and advanced students.

As was noted earlier, the creators of the revised Taxonomy did not give any examples how to use it in the computer science discipline. However, two articles were found, where the researchers have fixed this shortcoming. In [6] and [13], there are example problems and questions about programming for different levels of the Taxonomy. The assessment scale used in the survey generalizes those questions and problems.

Using a web questionnaire, the students were asked to give their student id and rank their skills in different programming topics by answering the following question:

Evaluate on every item, into which level (1..6) your knowledge best fits. The scale 1..6 means:

1. I can list related commands/concepts.
2. I can explain, what that command/concept means. I can apply an example to a similar problem.
3. I can list cases, when that command/concept can be used. I can apply an example to a different problem.
4. I can explain the meaning of that command/concept in its context. Why is it there.
5. I can ensure the correctness of use of that command/concept.
6. I can use that command/concept in problem solving without an example.

Tick the appropriate box on each item.

The whole list of course topics is visible in results section (figure 2). The list has actually 22 items of C-programming language and programming in practice, covering data types and operators as well as dynamic memory allocation and version control.

Because the assessment scale is ordinal, but the distance between the steps is not known, for each of the students a median of all his/her 22 answers was calculated, so that it could be compared to his/her exam grade. We used exam
grade as a verification point simply because it was easily available and gave the results we needed.

The students were also asked two open-ended questions:

1. How was it to answer this questionnaire?
2. If this questionnaire was published in the beginning of a course combined with the learning outcomes, would it help learning? Can you think other ways to use the questionnaire?

5. RESULTS

50 students answered the survey and 48 answers were valid. Two answers were missing the student id, so they could not be combined with their exam result. In addition four more answers were removed. Two were considered as outliers (boxplot), one was a row of ones with a comment that there was no time to concentrate on that, and one student did not make any effort in the exam. So, it left 44 answers to be used, about 51% of the active group.

Figure 2 shows how the students have answered in different items. The y-axis presents the level of the Taxonomy from remembering (1) to creating (6), x-axis lists the topics handled in the course. There are two set of markers lines, a median from what the students ticked in each item, and the lowest assessment. The lowest scores are not all from the same student. The topics in x-axis are in the same order as they were taught. As can be seen, the fundamental things, like data types, operators, input, and output are well mastered by most of the students. The median is at the highest level, creating. When going to the right on the x-axis the level of mastery gets lower. Things get more difficult and there is not as much time to practice. The most difficult things in C-language, pointers and dynamic memory handling, stick out at their own level. Most of the students think that they can use pointers and allocate memory, if they have supportive examples available. This corresponds quite well to the experiences the teachers have from the class exercises.

Figure 3 is a scatter plot of each students’ survey median and his/hers exam result. The dots represent the students. Each student has a survey result/exam result pair, and many students can have the same results. The relative size of the diamond tells how many students got just that grade-median combination. For example, of those students who got 4 from exam, 2 got 4 as a median from the self-assessment, 9 got 5 from self-assessment, and 3 got 6. Pearson product moment correlation between the survey median and the exam grade is 0.5429 and the p-value of that correlation is 0.0001. The figure shows a rising trend in self-assessment while the grades get better. When looking the results closer, older students seem to be better in assessing themselves. In our test course the threshold was between second and third year. For the first and second year students the correlation was 0.58 with a p-value 0.0039 (N=26) and for the older students 0.84 and 0.0000 (N=18) respectively.

The two open-ended questions were analyzed by simple open coding [11]. The first question dealt with answering process itself. 32 out of 44 (73%) answered. 21 considered it was easy to answer. 13 did not give any opinion. 2 mentioned that the scale was clear, but 6 considered the scale was peculiar, too general, difficult, or confused. Our students have used to use ordinary Likert-type scales in assessing courses or their own learning. This new taxonomy-based scale is new to them, as one comment stated: "Slightly
extraordinary choices”. There are other answers too, which indicate that the difficulty comes from novelty. In addition, one student commented bluntly, that “it is difficult to evaluate one’s own knowledge”. Based on the comments, it looks like the scale is understandable for most of the students, and the lack of experience has caused many of the difficulties.

To the question, if the questionnaire would help learning, the students answered as table 1 shows. The question described the intended use of this questionnaire and therefore guided the responses. Two thirds of the answers were positive or neutral and one third considered that the questionnaire would not help learning. One even stated that the questionnaire would not help learning, they described an alternative way of use, which was our idea put in other words: “assessment scale could be used along the course as a table for progress of own skills”. Therefore we do not consider this question and its answers fully reliable. The second biggest group of actual suggestions was that it would help to see what the course contains. This does not fully make sense, because the same list of topics is already available on the course web site. The biggest group agreed that it would help in goal setting, if the required levels are known from the beginning of the course.

There was no connection between the two answers. A student could feel the questionnaire difficult to answer but saw that it could help learning, or vice versa. Neither did exam results connect to any specific answers.

6. DISCUSSION

Traditional scales used in self-assessment like in [5] are subjective and those studies indicate that such self-assessment surveys are not reliable. Replacing Likert type scales with a taxonomy based scale makes assessment more objective, by defining the requirements for each level. We wanted to use a taxonomy scale in assessment to get the knowledge level, but only two taxonomies, Bloom and SOLO, have been widely used in assessment purposes and only Bloom’s Taxonomy concentrated on the depth of knowledge. Because of this we decided to use Bloom’s Revised Taxonomy as a basis for our self-assessment scale.

This study shows that the students in general seem to be quite accurate in assessing their own knowledge. There is a statistically significant correlation between the self-assessment results and the course exam results. Especially the lower levels of the taxonomy seem to be easy to distinguish. Even though the correlation exists, it has to be noted that there are plenty of students, especially among younger students who over or underestimate their skills. For them this scale does not give a realistic image of their knowledge.

Table 1: Question 2 answer categories

<table>
<thead>
<tr>
<th>Answer category</th>
<th>Number of answers</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>No answer</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>No help at all</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>Helps seeing what exactly the course contains</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Helps in preparation for the exam</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Tells the teacher, what has been difficult</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Start-end comparison</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Goal setting</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>

Figure 2: Knowledge level medians of course topics.
What causes the mismatch is hard to say. One reason is probably lack of experience. We gave this survey to our students without any prior preparation or explanation. The older students have more experience in assessing and reflecting their own learning, so they could grasp this new kind of assessment scale better. But, as it was also pointed out in [12], assessment skills improve by practice. If the scale was explained before first usage, the results will most likely be more accurate.

The topic-based grouping of answers (figure 2) seems to give a realistic image of students’ learning in general. The easiest topics as well as the most difficult ones are clearly visible and correspond to our view. Therefore, it could be a useful tool when a teacher wants to follow the students’ progress. A teacher can, for example, ask the students to mark their knowledge levels every three weeks and follow how the knowledge level rises after practice. If some topic stays in a low knowledge level, when it should not, then something needs to be done. In addition, when learning outcomes have been defined, the right target level is easy to define.

Some of the students felt that the scale was not clear. That is not surprising, because some sentences at higher levels of the scale were difficult to formulate. The students’ suggestions of how the tool could be utilized were partly expected and partly confusing. Their suggestions covered goal setting both for a course and in individual level, start-end comparison, and class advancement monitoring. These answers tell us that we are on the right tracks. Yet, it would have been interesting to know what some of the students meant by suggesting “Helps seeing what exactly the course contains”. It seems that they have not visited the course web site at all. This would have needed a follow-up question.

The generalizability of this study can be improved by extending the sample size and taking several courses and universities in the sample. In this study the sample size was relatively small and so far this scale has only been tested in one programming course.

7. CONCLUSIONS

In this study we created a general assessment scale for student self-assessment in a programming course. The levels of that scale were taken from the Bloom’s Revised Taxonomy by generalizing the level descriptions of the taxonomy and applying them to programming concepts. The results showed that the scale in this form gives a quite good general picture of students’ knowledge level.

The topic-based categorization of answers is a useful tool for a teacher. It tells clearly what topics are difficult and need further actions. In addition, it also tells precisely what level of knowledge most of the students have achieved. This is an improvement compared to the assessment scales ranging for example from poor to excellent, which are fully subjective.

After a rewording, and with proper guidance, the scale could be a useful tool for the students also. The students could set goals for themselves and follow their development or use it as an aid while preparing for an exam to see what is required in which level. The over and under estimations are causing some trouble, but practicing assessment helps in that matter.

For the teacher it will work as it is, but requires learning outcomes alongside it to be useful for the students. Some teachers already use learning outcomes at their courses and for the rest they should also be recommended. This tool is a motivation for using learning outcome descriptions as well.

8. REFERENCES


