

A Framework for Peer Assessment in Programming Classes

Elias de Oliveira,
Lucas Andrade
Univ. Federal do Espírito Santo
Vitória, Espírito Santo
elias@lcad.inf.ufes.br

Márcia Oliveira
Instituto Federal do Espírito Santo
Vitória, Espírito Santo
clickmarcia@gmail.com

Marcos Spalenza,
Matheus de Araujo Nogueira
vitoriavirtual.com.br
papers@vitoriavirtual.com.br

ABSTRACT

We can find a multitude of strategies for evaluating students in class. However, many of them are not possible to be carried out in practice due to the number of students in classroom. The strategy we find very useful to enforce learning among students is the blind peer evaluations. One drawback is the difficulty of deciding on fair and accurate assessment for the marks, which should also take into account the lecturer judgment evaluation. Therefore, we propose in this work some strategies to deal with this problem by using the Moodle as an interface between the lecturer and our proposed system. In order to create a rich representation of many possible real-case situations in class, we discuss them via simulations. The results have three folds. Firstly, we want to show how we are capable of identifying responsible students more accurately. Secondly, using our methods, we show whose students need closer assistance because they are not able to recognize the right answer, among some others, for the given activity. And finally, for those students of which evaluations are nearby to those also performed by the lecturer, a reward grade should be given.

CCS CONCEPTS

•Information systems → Content analysis and feature selection; Information extraction; •Applied computing → Collaborative learning;

KEYWORDS

Self-review; Peer review; Programming code assessment

1. INTRODUCTION

Programming is becoming a steadily popular discipline in many unconventional areas other than computer science [16, 18, 10, 5]. Nevertheless, our teaching practices are still out-of-date. Many of us are still sheet-of-paper based way of students' assessment on programming skill. Moreover, the only feedback a student receive from any previous practice is their own marked source code solutions. Usually, the student does not get to know what their colleagues thought about the same problem and how their solutions could be compared to among them.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

XXII TISE, November 28, 29, 30, 2017, Ceará, Fortaleza, Brazil

On the one hand, we believe the students would benefit most if they could discuss in pairs, in class, their own solutions [14]. On the other hand, the effort to give the right feedback at the right time to all students while their learning process in class is a tough task to be achieved without some sort of help from modern intelligent processing mining for educational system [19]. Besides, there are some initiatives in the literature [20] to overcome this problem. The use of peer-assessment can bring up the students discussion within class, where a set of codes solutions can be transparently criticized and their features evaluated accordingly to what is designed for that course, or class. In our proposed approach for this work, we claim that the lecturer/teacher/instructor has a key role to play when it comes to decide on the final fair marks [24], when reviewers fail to reach a consensus, or the students are practicing unfair grading. This why in our approach the lecturer's judgment is an important part of our students' peer-assessments model.

Instead of giving the students the responsibility of their own individual evaluation, as it is done in some self-evaluation approaches [13], our approach introduces the idea that each student *must* evaluate some source-code solutions of other students in class. Thereby each student will have the opportunity to assess what and how their classmates expressed themselves with regard to the same activity solution. It is important to say that the source-code solutions are anonymously shown to them. Therefore, each students does not know which programming source code their colleagues have submitted.

In order to deal with all the grades given by each student in class to their assigned set of source codes to be evaluated, we built a system which together with the Moodle¹, a very popular learning management system (LMS), manages all the necessary red-tape work. Our system fetches all the submitted source-code solutions submitted by the students of a class from a given Moodle to our computer server. We format these solutions into a nice readable table of responses to be presented back to the students only the programming source-code solutions, without students' identifications, on another activity created by the lecturer on the Moodle (see Table 1 in Section 4). Having this new activity online, the students now enter their grades for the assigned source-code solutions of their classmates. The rest of calculations are done internally by our system and both the grades for the first and the second activities are sent back to the Moodle.

Our system is so far configured to provide the user with two types of evaluations. The first is that when the student's grades *ranking* are similar to that given by the lecturer. This shows that the student is able to assess the importance of some answers over others as quite well as the importance also expressed by the

¹ <http://www.moodle.org>

instructor. In this case the exact values given to grade the assigned source codes are not so much the goal.

Another type of evaluation is that when the student is asked to guess the *correct values for grades* given by someone experienced such as the lecturer. In this case the student receives greater grades as much as their own grades are closer to that given by the lecturer. In our opinion, this type of evaluation is harder than the previous one.

The literature show that peer assessments are also good strategies to improve student motivation, stimulates students to reflect, discuss, and collaborate in their learning process [2, 24]. We show in our experiments how to visualize that some students were able to grasp quite well the overall lecturer criteria of evaluation, whereas some others are those who need greater help on the subject.

This work is structured as follows. We discuss, in Section 1, we discuss among some literature works some similarities with the current work conducted in this paper. The architecture of our system is discussed in Section 3, which is responsible to extract the data from the learning management system, process it and send it back to the user interface. In Section 4, we describe how our simulation experiments were performed and the results yielded, and the conclusions are presented in Section 5.

2. SOME RELATED WORKS

Evaluations play important role both on the students learning process and for the lecturer's improvement of their teaching's strategies. As many more different forms of evaluations the better, as the students can perform better in some than other types. This way, by observing the students on their multidimensional form of learning, the lecturer can isolate the key variables which could be hindered the students learning process [11].

In other words, the evaluations serve to the two sides of the process: to the learner and also to the lecturer. For the former to correct their learning procedures in order to achieve increasingly better results. For the later, to rearrange its strategies in order to improve their capacity of precisely helping the learner in their individual needs. Therefore, the scientific progress in class can only come by the continuous refinement of the route of pedagogical strategy [17].

One of the problems the lecturers are facing these days is to keep pace with the great number of students in class that the educational environment is impinging, especially in the e-learning environments [3]. Thereby some strategies to reduce the lecturer effort, without neglecting the quality, have been proposed in the literature. We discuss in this section only two types of these strategies, for the sake of brevity. The a) self-assessments [12] and b) {blind, or not} peer-assessments [6], whereas we consider to be *traditional* that approaches where the assessment is done only by the lecturers themselves. Furthermore, in this work we are considering both the self and peer assessment, based on given grades, used in conjunction with the lecturer grades assessment as the baseline [1] for the final results.

In this work we will discuss three sorts of assessments: the self, the peer and a mutual strategy. The latter is a sort of a combination of both previous type of assessment strategies.

Although it is an important tool of evaluation, some lecturers may resort to the self-evaluation strategy as a way of diminishing their workload. On the contrary, this is a great opportunity for

acquiring the cognitive state-of-mind of the students. In this strategy [7, 26] the students are asked to look into their own activities and, based on the course prior accorded criteria, engage in their own learning, learner responsibility, metacognitive skills and a dialogical, collaborative model of teaching and learning. In other words, this is a powerful way to bring up the student awareness of their active role in the learning process. According to [1], self-assessment should include two main elements: a) *making decisions about the standards of performance* expected and then making b) *judgments about the quality of the performance in relation to these standards*. The main problem, besides involve students in both of these aspects, is how to introduce this practice without blowing up the workload the lecturer/instructors which are already doing many other day-to-day duties [1].

One issue to tackle as a result of any self-evaluation strategy is that pointed out by the authors in [23, 8]. Their works showed that the students' perception of themselves is not as accurate as their actual performance. In the light of these studies, thus, we need to find a way to balance this strategy with some other approaches to produce the learning desired improvement intent for these students. The authors in [12] claim that when

Correctly implemented, student self- assessment can promote intrinsic motivation, internally controlled effort, a mastery goal orientation, and more meaningful learning.

In their work, the authors pointed out a schematic to explain the meaning of self-assessment where three aspects are depicted: a) self-judgment [25]; b) learning targets and instructional correctives; c) self-monitoring. All these aspects are the important engine to improve the students' learning.

The peer assessment has another cognitive approach. Rather than trying to be critical with their own work, in a peer-assessment strategy they are now impelled to look into not only their own respectively activities, but their classmates' activities as well. This can be done by a variety of means.

Peer learning is actually part of our development from the earliest years of life and the centrality of the lecturer makes us lose sight of this. We need to transform our educational environment into a place where one can see the birth of critical thinkers, who can evaluate the pros and cons of different ideas, or point of views, and *etc.* [22].

Differently of what we are interested in this work, some authors [24] dealt with the problem of arbitrating the conflicts among of group assessment, in what they called the *non-consensus*, *i.e.* when two or more students do not agree with a reasonable evaluation of a particular activity. They say that *non-consensus is a common challenge in this type of approach that makes the reliability of peer assessment a primary concern in practices*. Their proposed solution is based on the use of *review deviation and radicalness* to identify non-consensus in peer assessment. One of the suggested strategy by the authors was to award that student who gives a score close to the mean value of its review group scores, whereas should be penalized who gives a score far away from the group's mean value.

The addition of the lecturer as an active actor in our proposed approach either in the self and the peer assessment, we are capable of giving more stability to the grade results [20]. Most of the previous authors pointed out some problems to blindly accept

both self and peer assessment [6] as the final grade. According to this hypothesis, in our proposed approach, we will always use the lecturer assessments to guide the final results. In this work we will consider two strategies: a) the actual grade value similarity between the student and the lecturer at each activity and b) the ranking grade between the student and the lecturer at each activity. Nonetheless, it is possible to consider many other ways of combining the lecturer given grades with the students', for instance, the mean of students' grades can also be considered when evaluating the student individual assessment [24]. Note that, by grading each other activities a great deal of metacognitive skills is worked out by the students.

In order to implement both of these discussed strategies, we apply the Pearson and Spearman correlation [4, Cap. 4]. The first when we want that the student learn to give the right value to the activity, whereas the second when we are satisfied enough with the order of the grade given by the student when compared to the order given by the lecturer.

3. THE SYSTEM ARCHITECTURE

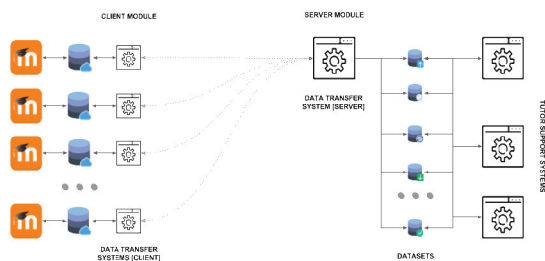


Figure 1: System architecture for LMS data extraction.

Data mining tools are nearly part of our everyday lives. Embedded in many modern systems to help to acquire knowledge about the user to provide this user with better experience. In education area this still quite an exception, although some developments can be found in the literature and few others are actually applied in traditional classrooms [19,9,15]. A great opportunity is rising with the increase use of learning management system and the popularization of open and Mass Online Courses – MOOC's, where a huge amount of data is produced which can review old needs now possible to be quickly discovered by new methodologies [21].

Research in Educational Data Mining (EDM) [19] seeks to count for the data demands which come from LMSs either when in online learning, or in classroom. In order to meet the day-to-day production of data and to provide the suitable support to the instructors of the modules it is necessary to design some intelligent systems that act specifically to acquire enough information from data so that it can be able to act very similar to what would be done by the specialist of that class. Therefore, to this end, we extract the data from the learning platforms and take them to our local servers. This is done by the system module called *Tutor Support Systems* (TSS), ensuring that this transference meets some security requirements.

As an example of the amount of data produced by LMSs, we depict here the statistics yielded by the Moodle. According to its

official website ² there are 72 thousand systems registered in 232 countries around the world. Currently with 95 million users, Moodle is a great example of academic productivity for teaching professionals who individually easily assess all students. Therefore, by extracting the data from a system as such we can process it with modern data mining algorithm and turn back richer knowledge to the lecturer of a course provider of the data.

In order to orchestrate our proposed approach, a *Data Transfer System* (DTS) interacts directly with the LMS platform database to fetch the data uploaded by students. Using the LMS *web service*, the DTS connects clients and servers generating results on demand. Working with these tasks, the required TSS generates and sends report to all participants. The feedback is given by system using an upload method of DTS. Like download, upload function uses transference services to insert data directly on learning platform. For this application cycle, a developed TSS can become a tailored service which can be activated by the lecturer as wish during the configuration of course assignments. The architecture is presented in Figure 1 in order to keep services in high availability for all LMS.

In the next section, we show some of the results of our strategy applied to some of our datasets and show that our approach is actually able to resolve the problem in the self- and peer-assessment approaches of being fairness and accurate.

4. EXPERIMENTS AND RESULTS

In order to turn clearer the use of our proposed framework, in this section we describe: a) how students perform the combined peer assessment activities within our LMS interface; b) some of the cognitive classes we have already identified, and c) what would be some possible actions toward reaching out the students in their specific need. As we said in Section 1, we are going to focus here in only two out of many other ways of assessing the grades given by the students when self, or peer evaluating.

In our approach one needs to consider two basic steps. In the first step, the lecturer assigns an activity where all the students will hand out their answers via our LMS. At the end of this first step, we will have all the student's activity answers in our system and the lecturer can now assign their grades, we are considering possible values from 0 to 100. But before reviewing the grades to the students, the lecturer opens another activity. This is the second step of our methodology. Now the students will be asked to give their own grades for their classmates' answers including her own answer. The activities to be assigned to the students in this second step is chosen automatically by our framework software from those already submitted by the students in the first step, so the lecturer does not need to worry about it. Each activity will have its unique ID (See Table 1). Therefore, the students will analyze their anonymous classmate's answers and, then, assign a grade to each of them. For instance, consider the example shown in Listing 1. The Listings from (1) to (3) are some arbitrary chosen answers for a worked activity in one of our classes. The description of this activity is as the following:

The Activity Statement

² <https://moodle.net/stats/>

Write a program to get the number of P points from three teams in a football league, according to the following mathematical expression:

$$P = 5G_p - G_n + 3V_f + 2V_c + E$$

In this expression, G_p is the number of positive goals, G_n is the number of goals taken, V_f is the number of wins away from home, V_c is the number of wins at home and E is the number of draws. In the end, the program must show, according to the number of points obtained by a team, the champion and the runner-up of the championship.

Listing 1: (ID=544-8855) A Not Good Code Example

```
#include<stdio.h>

main()
{
    int x,P,Gp,Gn,Vf,Vc,E,aux;
    printf(" digite na ordem o numero\n");
    scanf("%d",&x);
    while(x<=3 && x>=1)
    {
        printf(" digite na ordem o numero de gols positivos,
negativos,
vitorias fora de casa, dentro de casa, e empates do 10
times na ordem começando pelo primeiro\n");
        scanf("%d%d%d%d%d",&Gp,&Gn,&Vf,&Vc,&E);
        P=5*Gp-Gn+3*Vf+2*Vc+E;
        scanf("%d",&P);
        x++;
        aux=P;
        if(P>aux)
            printf("%d\n",P);
    }
}
```

This code was submitted by one of the students as the answer for one of our activities. However, this code was submitted by one of the students as the answer for one of our activities. The main reason the lecturer marked as zero (0.0) for this code was because it did not express the objective, evidencing that the student did not understand the problem statement. Besides, this code does not compile. We thus aspect the students should have the cognitive feeling, if not the same, a similar sense of value when asked to give their grades for this answer

Listing 2: (ID=169-8700) A Better Code Example

```
#include<stdio.h>

main()
{
    int i, c, v, GP, GN, VF, VC, E, pontos, campeonato=0,
viceCampeao=0 ;
    for(i=1;i<=3;i++)
    {
        printf("Digite o numero de gols positivos, o numero de
```

```
negativos, o numero de vitorias fora de casa, o
numero de vitorias em casa, e o numero de
empates do
time %d: ", i);
scanf("%d %d %d %d %d", &GP, &GN, &VF, &VC,
&E);
pontos=(5*GP- GN +3*VF +2*VC +E);

if(pontos>campeao)
{
    campeonato=pontos;
    c=i;
}
else if(pontos>viceCampeao)
{
    viceCampeao=pontos;
    v=i;
}
}
printf("O campeão é o time %d com %d pontos,e o vice-
campeao é o time %d
com %d pontos ", c, campeonato, v, viceCampeao);
}
```

The code presented in Listing 2 is another example. This is a better code than that shown in 1. An important aspect valued by the lecturer in her assessment on this answer was that this student showed a higher cognitive level of understanding when used a decision instruction to choose the runner-up of the championship. This was an important characteristic to bring the grade to the maximum of the total grade.

Listing 3: (ID=175-8500) The Best Code Example

```
#include<stdio.h>
main()
{
    int GP,GN,VF,VC,E,P,valc=0,valu=0,timec,timev,ntime=0;
    do
    {
        printf("entre com o numero de gols positivos e gols
negativos
respectivamente \n");
        scanf("%d%d",&GP,&GN);
        printf("entre com o numero de vitorias fora de casa e numero
de
vitorias dentro de casa respectivamente \n");
        scanf("%d%d",&VF,&VC);
        printf("entre com o numero de empates \n");
        scanf("%d",&E);
        P=5*GP - GN + 3*VF + 2*VC + E;

        ntime++;
        if(P>valc || P> valu)
        {
            if(P>valc && P>valu)
            {
                timev= timec;
                valu= valc;
                timec= ntime;
                valc= P;
            }
        }
    }
}
```

```

        if(P> valv && P< valc)
        {
            timev= ntime;
            valv= P;
        }
    }while(ntime<3);
    printf("O time campeao e %d com %d pontos\n",timec,valc);
    printf("o time vicecampeao e %d com %d pontos
\n",timev,valv);
}

```

Finally, in Listing 3, differently of that previous code, this code was successfully compiled and yielded some results, but not yet the correct one. However, the student showed a reasonable cognitive level of understanding the content discussed in class about the elements surrounded this activity even though she used the logical expressions in a confused way. This student was awarded with a maximum grade.

In the second step, we constructed a report with all the students responses, as shown in Listings 1–3. The anonymous ID of the respondent is placed within the code listing. After having examined the list of codes in the report, students are then asked to give grade to each of the answer in the list of codes. We specified a protocol for entering the grade values in to our LMS, as follows. Each line written by the student is initiated by a "#" hash symbol, followed by the ID of the respondent, as shown in Table 1. A semicolon is used to separate the ID and the given grade. For instance, four lines of a student peer assessment evaluation is depicted in Table 1:

Table 1: The grades given by a student into the LMS

A Student	
ID	Grade
#169-8700;	50.5
#172-8498;	45.00
#175-8500;	100
⋮	⋮
#544-8855;	0

In the first column we can see all the IDs (#169-8700, #172-8498, #175-8500, . . . , and #544-8855) of the selected answers assigned for a student to grade. In the second column, after the semicolon separation character, is placed the respectively grade values for the answers, for instance, the student decided to grade a slight superior of the medium value, 50.5, for the #169-8700 answer, whereas graded a maximum value for the #175-8500 answer.

The grades entering by the students are then automatically paired with that assigned by the lecturer, in the first step, and the specified correlation formula for that activity is automatically calculated to be transformed into a new grade to each respective student for the peer assessments.

One possible configuration is to set the activity to transform into grade the value resulted from the *Pearson correlation* between the student grades and the grades given by the lecturer for the same answers. The greater the correlation with the lecturer grades, the greater the grade this student will be assigned. Another possibility

is, alternatively, to set the *Spearman correlation* for the activity, in this case the student is asked to correctly rank the grades among the answers. We also provide the user with the possibility of specifying an expression to combine some correlation formulas to produce the final grade for the student’s activity.

Figure 2 shows the correlation results of all students grades together with the lecturer grades L , called here as the *baseline*.

The values shown in Figure 2 is a result of a simulation to help us to discuss some categories of grading assessments and see how they are visually presented in tools such as that heat map depicted in the figure. In the experiments we have four types of students: a) r_i – those who act more like randomly assigning grades to their classmates’ activities; b) aa_j – those who are very close to that grades also assigned by the lecturer; c) ab_l – those who are very close to that grades also assigned by the lecturer, but not as close as the aa_j group of students; d) ac_m – those who are close to that grades also assigned by the lecturer, but farther than the ab_l group of students, given that $i = 1, 2, \dots, M; j = 1, 2, \dots, N; l = 1, 2, \dots, O; m = 1, 2, \dots, Q$. M, N, O , and Q are integer values to express the quantities in each category of students.

Thereby on the bottom side of the Figure 2 there are all the labels starting from the right side with the L label, the lecturer baseline grading, and ending with ac_{10} at the left side of the figure. Moving up one line, we find again the correspondent L label at the right side of the figure. The cell at the first line, combined with the first column at the left side, we can see that the correlation with L between itself could not be different but 1.

Based on the Figure 2 it is quite simple to spot those students who are not aligned with the criterion the lecturer is applying to assign grades to the activities. Although these students may assign quite very similar grades as those also assigned by the lecturer, we claim that these students are not fully understanding whatsoever the course content, or the lecturer criterion of judging the activity assessment.

On the other hand, among those students of the category ac_m , who are not much aligned to the lecturer way of grading as much as the aa_j -individuals are, we can see that they are very alike. Their color in the figure is quite similar to that very light blue also showed for the correlation between L and L . This is at the left column on the top of the figure. Note that the correlation among categories such as r_i , aa_j , ab_l and ac_m are darken then L and ac_m , this is placed at the second quadrant of the figure, at the bottom right side of the figure. This is to show that they are clearly different among them and, what is more important, we can easily spot them.

In the light of this, we may say that the lecturer should pay a great deal of attention to those r_i group of students, whereas aa_j group are clearly as good as the lecturer. Although the ac_m class of students also need some care and attention, the necessary action for the correction of the course is simpler than that necessary for the r_i group of students.

The tool showed in Figure 2 can even tackle, what is considered by some, one of the common problem when dealing with mutual assessment on online LMS, which is the possibility of the students to copy somebody else’s evaluation. This can be detected by the high level of correlation between pairs of students grading [14]. Hence, the lecturer can now act in order to clarify those situations where pair of students have a high correlation between them on one or more activities.

Based on the heat map of Figure 2, we can now easily grade the students on their correlation achievements. One approach is to assign the correlation straightforward as the grade, adjusting only the correlation value to the range of the LMS grade interval. For those who did not respond, or are uncorrelated to the lecturer, grading them to zero is a possible approach.

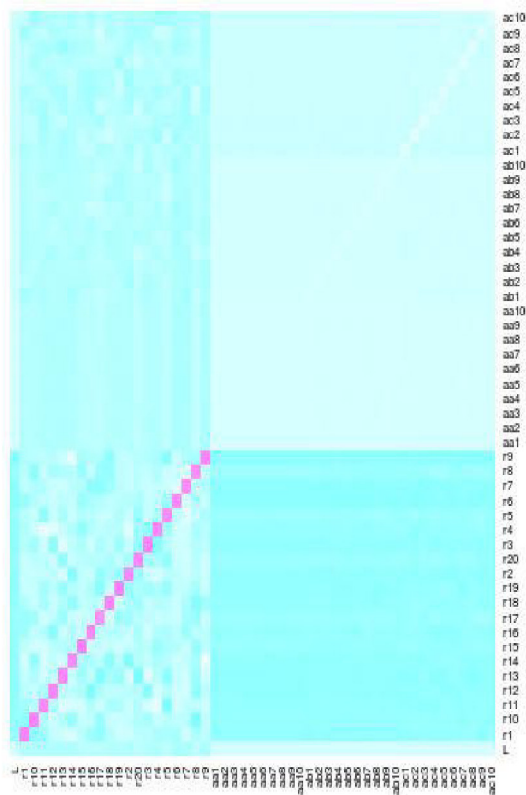


Figure 2: The correlation between the students and the baseline grades

To sum up, in addition to helping the lecturers to motivate even more the students in their own learning process, we can extract from this tool the information when the students are actually learning or need more attention to the topic discussed in class on the subject regards to the evaluated activity.

5. CONCLUSIONS

Everybody who teaches has to spend lots of time on inspecting and assessing the activities carried out by each student in order to show their learning progress [24]. In this paper we discussed two strategies for weighting students' grades given by themselves to their respectively classmates' programming source codes. Our methods are also good strategies to assess the student's cognition when some of the answers – programming source codes, developed by their colleagues in class are presented back to each one of them. By our proposed method, each student has to assess its own programming code, and also give grades to some selected programming source codes submitted by their classmates. These

selected source codes are chosen accordingly to each student in order to stimulate their evaluation cognition.

In addition, our experiments showed that the implemented tool eased the performance of the automatic mutual assessment, which otherwise this process would be nearly impracticable. Therefore, these tools are ready to use and help lecturers to mediate the students to a better learning. The approach proposed in this work open up the possibility to build a set of reports where we can point out the lack of cognition expressed by some students on certain topics of the learning process.

We believe that student self-assessment, defined as a dynamic process in which students self-monitor, self-evaluate, and identify correctives to learn, is a critical skill that enhances student motivation and achievement. Therefore, this should be more used in classes. We believe our proposed framework provide the community with a good tool for turning peer-assessment strategies a more popular practice in our learning process.

In future work, we are planning to consider to include a cleverer way to select source codes for a particular student. A way in which one could exploit the current cognitive status of the student and then help this student to improve their learning experience.

6. REFERENCES

- [1] BOUD, D. *Enhancing Learning Through Self Assessment*. Kogan Page, 1995.
- [2] BOUZIDI, L., AND JAILLET, A. Can Online Peer Assessment Be Trusted? *Journal of Educational Technology & Society* 12, 4 (2009), 257.
- [3] BRESLOW, L., PRITCHARD, D. E., DEBOER, J., STUMP, G. S., HO, A. D., AND SEATON, D. T. Studying Learning in the Worldwide Classroom: Research into EdX's First MOOC. *Research & Practice in Assessment* 8 (2013).
- [4] BUSSAB, W. O., AND MORETTIN, P. A. *Estatística Básica*, 8 ed. Saraiva, São Paulo, 2013.
- [5] CIOFFI-REVILLA, C. *Computation and Social Science*. Springer International Publishing, Cham, 2017, pp. 35–102.
- [6] DE GREZ, L., VALCKE, M., AND ROOZEN, I. How Effective are Self-and Peer Assessment of Oral Presentation Skills Compared with Teachers' Assessments? *Active Learning in Higher Education* 13, 2 (2012), 129–142.
- [7] EGAN, P. A. *The Relationship Between Student Self-Regulation Strategies and Increased Student Achievement: A Study on How the Explicit Integration of Self-Regulation Strategies Impacts Student Reading Achievement in the Elementary Classroom*. PhD thesis, University of St. Francis, 2017.
- [8] FALCHIKOV, N., AND BOUD, D. Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research* 59, 4 (1989), 395–430.
- [9] FLORIAN, B., GLAIN, C., DRACISLER, H., SPECHT, M., AND GESA, R. F. Activity-Based Learner-Models for Learner Monitoring and Recommendations in Moodle. In *European Conference on Technology Enhanced Learning* (2011), Springer, pp. 111–124.
- [10] HOSSEINI, R., AND BRUSILOVSKY, P. A Study of Concept-Based Similarity Approaches for Recommending Program Examples. *New Review of Hypermedia and Multimedia* 0, 0 (2017), 1–28.

- [11] LIEBERMAN, D. A. *Learning: Behavior and Cognition*. International Student Edition. Wadsworth Publishing Company, Belmont, CA, 1990.
- [12] MCMILLAN, J. H., AND HEARN, J. Student Self-Assessment: The Key to Stronger Student Motivation and Higher Achievement. *Educational Horizons* 87, 1 (2008), 40–49.
- [13] NICOL, D. J., AND MACFARLANE-DICK, D. Formative Assessment and Self-Regulated Learning: a Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education* 31, 2 (Apr. 2007), 199–218.
- [14] OLIVEIRA, E., AND SPALENZA, M. Self and Peer Assessment Strategies. In *Computer on the Beach 2017* (Florianópolis, SC, May 2017), SBC.
- [15] OLIVEIRA, M. G., CIARELLI, P. M., AND OLIVEIRA, E. Recommendation of Programming Activities by Multi-Label Classification for a Formative Assessment of Students. *Expert System and Applications* 40 (2013), 6641 – 6651.
- [16] OUSTERHOUT, J. K. Scripting: Higher Level Programming for the 21st Century. *Computer* 31, 3 (Mar 1998), 23–30.
- [17] PERRENOUD, P. *Avaliação: Da Excelência à Regulação das Aprendizagens – Entre Duas Lógicas*. Artmed Editora, Porto Alegre, RS, 1998.
- [18] PRENSKY, M. Programming is the New Literacy. *Edutopia magazine* (2008).
- [19] ROMERO, C., VENTURA, S., PECHENIZKIY, M., AND BAKER, R. S. J. D., Eds. *Handbook of Educational Data Mining*. CRC Press, Boca Raton, FL, Oct. 2010.
- [20] SHIBA, Y., AND SUGAWARA, T. Fair Assessment of Group Work by Mutual Evaluation Based on Trust Network. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings* (Oct 2014), pp. 1–7.
- [21] SIEMENS, G., AND BAKER, R. S. J. D. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (New York, NY, USA, 2012), LAK '12, ACM, pp. 252–254.
- [22] SPILLER, D. Assessment Matters: Self-assessment and Peer Assessment. Tech. rep., Universidade de Waikato, Nova Zelândia, 2009.
- [23] TOUSIGNANT, M., AND DESMARCHAIS, J. Accuracy of Student Self-Assessment Ability Compared to Their Own Performance in a Problem-Based Learning Medical Program: a Correlation Study. *Advances in Health Sciences Education* 7, 1 (2002), 19–27.
- [24] WANG, Y., LIANG, Y., LIU, L., AND LIU, Y. A Motivation Model of Peer Assessment in Programming Language Learning. *Computing Research Repository – arXiv.org abs/1401.6113* (2014).
- [25] ZIMMERMAN, B. J. Becoming a Self-Regulated Learner: An Overview. *Theory into Practice* 41, 2 (2002), 64–70.
- [26] ZUMBRUNN, S., TADLOCK, J., AND ROBERTS, E. D. Encouraging Self-Regulated Learning in the Classroom: A Review of the Literature. *Metropolitan Educational Research Consortium (MERC)* (2011), 1–28.