STRATEGIES FOR STUDENT EVALUATION IN THE ONLINE CONTEXT

SANDA-MARIA AVRAM

ABSTRACT. In this paper we conducted an investigation on the performance of the students during the second semester of the academic year 2020-2021. We looked at the performance results obtained by students on the laboratory work, practical and final exams while we were forced by the Covid pandemic to move entirely into an online education system. Our focus was to determine the impact of a consistent behaviour (or lack of it) on the final student performance. We determined that, even in an online setting, a good involvement (in terms of attendance and good performance) guarantees good final results. The investigations were performed using the Formal Concept Analysis, which is a very powerful instrument already used by us in previous research in order to detect student behaviour in using an e-learning portal. Another set of results showed that the change of the final mark computation formula to be based in a higher proportion on the lab work was closer to the actual overall performance of students.

1. INTRODUCTION

Nowadays the Covid pandemic forced most of us to migrate to the online educational system. Although in the last decade the online educational systems have shown a rapid development [15], being somehow forced to adopt an only-online education brought to light many challenges and many places where this system needs improvement. It also put everything into perspective and maybe made us to better appreciate the aspects that make traditional education valuable. Online educational systems are the set of techniques, methods and environments that provide access (through Internet) to educational support for students [15]. There are synchronous components (e.g., students may attend live lectures; real-time interaction between educators and students may

Received by the editors: 9 November 2021.
2010 Mathematics Subject Classification. 68T30, 68P20.

Key words and phrases. student evaluation, formal concept analysis, online learning.
exist with the possibility of instant feedback), and there are also asynchronous components (e.g., students may access teaching material in an any-time, any-place, any-pace manner).

From the teaching perspective, the online instruments used to be considered an extension and support of traditional learning. Authors J. Liebowitz and M. Frank define in [12] the concept of blended learning as “blend” between traditional and online learning. The proportion in which each of the two learning systems is used creates different types of blended learning.

Y. Park compares in [13] a discussion-based blended learning model with a lecture-based blended learning model. In the first type of learning model, the students are expected to be actively involved in online forums, while in the second type of learning model the main online activities of students are: submitting tasks or downloading materials. The investigation from the data collected in this study show that there is a linear prediction between online activities and student performance, i.e. the total score that they obtain in the case of discussion-based blended learning course. However, no linear prediction exists in the case of lecture-based blended learning course. Therefore, it is concluded that the type of online activity is important in determining whether the online involvement of students could predict corresponding outcomes.

In our current research, we use standard deviation to see the distribution of marks and Formal Concept Analysis as a technique to discover patterns in the effect that the integral online education system (that was somewhat forced by the Covid pandemic) had over the teaching process and whether the changes we made in our approach had a positive or a negative effect.

The investigation described in this paper was done through an FCA-based analysis performed on the students’ activity results during one semester (i.e., 14 weeks) by considering the marks obtained during and at the end of the semester. The considered activities took place during the second semester of the academic year 2020-2021. The rest of the paper is structured as follows: section 2 presents a very short overview of the “forced” online learning during the Covid pandemic. Section 3 presents the theoretical background for the method/instrument/technique we use in our investigation. The motivation of this work as well as some prerequisites and other useful details are described in section 4. Section 5 presents the actual tests and results. The last section (i.e., 6) contains the conclusions and future work.

2. Online learning in the COVID context

There is a consistent body of research based on this topic, especially in these last couple of years due to the Covid pandemic. Among the papers to mention is [2], where the author presents a fivefold perspective (i.e., “strengths,
weaknesses, opportunities, and challenges”) on what the Covid pandemic crisis brought to the educational system. In terms of strengths there is the time and space flexibility, the customisation of the learning process based on the student’s needs, the possibility of creating an environment that is collaborative and interactive. In terms of weaknesses the author mentions the loss of direct human communication, the technical problems or difficulties (e.g., frustration and confusion generated by the online environment, immature instruments), loss of student’s accountability due to the time and space flexibility. The main opportunity is the “online learning boom” that forces both the teachers and the students to find and try new technical solutions. This comes with the possibility of acquiring new abilities such as critical thinking, adaptability, resilience. The opportunities are also great for the IT community, which could and do help with developing instruments/programs that are tailored to the online education. Maybe the most numerous and important points are touched upon in terms of challenges. Due to this “forced” change from the traditional to the online education, most of the challenges stem from trying to integrate, engage and motivate all the participants in the learning process (e.g., teachers and students). The online environment is not mature yet, therefore there are not as many rules and regulations nor the actual infrastructure as in the case of traditional system. Another important challenge is the additional costs involved with the equipments, trainings and creating online educational content. These costs have also an equity aspect in the sense that not all teachers and/or students have the possibility to acquire the required equipment and/or the Internet connection.

The study presented in [1] investigates the perspective of Pakistani higher education students on the online learning in the pandemic context. The results obtained show that traditional learning is preferred in this case.

Authors of [8] put everything in perspective by discussing the difference between “Emergency Remote Teaching and Online Learning”. They begin their investigation stating that online learning seems to be perceived as having a lower quality than traditional (i.e., face-to-face) education, despite the fact that the research shows that that is not always the case. The view that this study lights upon is that the traditional educational system, because it exists for so long, has developed an “ecosystem” (a consistent infrastructure) in which lectures are only one component, thus, similarly, in time, an effective online educational system will require to build its own “ecosystem” in order to support efficient education.
3. Formal Concept Analysis

Formal Concept Analysis (FCA) stemmed from applied mathematics but it is nowadays positioned in the Knowledge Discovery and Representation field. In the formal (dyadic) context, two sets are considered, the first being a set of objects, and the second being the set of attributes. Between the elements of the two sets there is defined a relation that states “object $o$ has attribute $a$” [7].

Maximal groups of objects that have the same attributes are determined. Such groups are called concepts. All determined concepts form a complete lattice by introducing an order relation between the concepts. The order relation states that between two concepts there is a relation in which one of them is considered subconcept and the other is considered superconcept if and only if all elements of the set of objects of the subconcept are included in the set of objects of the superconcept.

In the triadic format, alongside the two sets considered by the dyadic FCA, a third set is introduced that is called conditions. Therefore, the relation between the three sets now states “object $o$ has attribute $a$ under condition $c$” [11].

In the dyadic format, the relation can be represented as an incidence table on which each object takes up a row and each attribute takes up a column. In the triadic form, however, we have a tridimensional representation of the relation between elements of the three sets, so that for each condition an incidence table as the one described for the dyadic form can be used to describe the relation.

A triadic formal concept is also called a triconcept and consists of maximal groups of objects that have a specific set of attributes under a specific set of conditions [11].

The set of objects form a formal concept is called extent, while the corresponding set of attributes is called intent, both in the dyadic and triadic context. The corresponding set of conditions is called modus in the triadic context. [11].

There are several tools that have been developed for FCA. The one we used in our investigation is the FCA Tools Bundle [10, 9] as it offers a user-friendly visualisation for contexts and in addition to this, it enables navigation in the triconcepts. The concepts of a dyadic context can be visualized as a concept lattice. Triconcepts cannot be viewed in a concept lattice like in the case of a dyadic context. However, exploring triconcepts can be done by deriving dyadic contexts from them by projecting along one of the dimensions of a triadic concept (i.e., the extent, the intent or the modus).
Our own previous contributions use the FCA instruments in order to investigate into the behaviour of students [6, 4, 5] while using an educational portal called PULSE [3].

4. Defining the problem

The students’ activity investigated in this paper was done online. We used Microsoft Teams for video conferencing with students. The laboratory work was evaluated after an interview conducted in a one-to-one (i.e., student-to-teacher) manner. The practical exam was done using Moodle [14] quizzes oriented more on the practical concepts. The final exam consisted also in Moodle quizzes randomly selected from a large set of questions based (in a balanced manner) both on theory and practical aspects. The students involved in the investigation belonged to two sections (with slightly different study objectives) which we are going to refer further on as S1 and S2. During the final exam S1 students were supervised using video on Microsoft Teams and with shared screen on https://meet.jit.si/, while S2 students were supervised only using video on Microsoft Teams. Due to concerns that online examination setting is more prone to cheating possibilities we changed the formula of computing the final mark by decreasing the weight of the exams and increasing the weight of the laboratory work.

As a teacher you would like to convey as much information to students as possible and perfect your methods/skills each year. And as much one would like that all students would acquire the maximum level of information, one would also like to have a fair and an honest evaluation system. One would prefer the marks to reflect as closely as possible the level of information that students have acquired.

We are trying, therefore, to take a closer look at the entire activity of students and more specifically their activity during the semester (i.e. their laboratory work where they have to implement the new concepts in order to solve some given problems) and the practical and final exams.

We analyse the students’ marks in order to determine patterns in their behaviour which can have an impact on their overall performance. We considered in our investigation a mandatory subject. The students’ evaluation included their marks obtained during the entire semester (14 weeks - in which they were supposed to complete 8 assignments), a practical examination and a final exam.

Table 1 presents the number of the students involved in the investigation. The students study this mandatory subject in their first year of study for students in S1 and in their second year of study for students in S2. During
the last week of the semester there is a practical exam, and at the end of the semester there is a final examination.

<table>
<thead>
<tr>
<th>Table 1. Details about the number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sections</td>
</tr>
<tr>
<td>S1</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Total number of students</td>
</tr>
<tr>
<td>First-time enrolled students</td>
</tr>
<tr>
<td>Re-enrolled students</td>
</tr>
<tr>
<td>Students who passed</td>
</tr>
<tr>
<td>Students who failed</td>
</tr>
<tr>
<td>Promovability rate</td>
</tr>
</tbody>
</table>

The students which do not pass the subject during this phase have another chance within a re-examination. All the students which do not pass this subject after re-examination have to re-enroll within one of their next years of study. Therefore, the “Re-enrolled students” from Table 1 are (in this case) students in their third (final) year of study.

In the Romanian education system the marks are given on a scale from 1 to 10 (10 being the maximum), a mark equal or above 5 denotes passing the exam/subject. We are going to coarse the range of the marks following the qualification system applied in our primary educational system, but also in other countries. The four qualifications we use are depicted in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Qualification classes used for student results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualification name</td>
</tr>
<tr>
<td>insufficient</td>
</tr>
<tr>
<td>sufficient</td>
</tr>
<tr>
<td>good</td>
</tr>
<tr>
<td>very good</td>
</tr>
</tbody>
</table>

We consider that a student has a consistent activity when the marks obtained vary only slightly. In order to determine the consistency of students activity we used the standard deviation applied on the marks they obtained as detailed in the following bullet list. We are analysing these values considering the two thresholds (i.e., $th_1$ and $th_2$) that demarcate the 3 classes of acceptable (acc), big, and respectively too big values, as depicted in the Table 3:

- *labActivDEV* is the standard deviation for the 8 lab marks (any unhanded laboratory work was marked with 0). The first threshold for this value is 1.5 and the second threshold is 3.
• `semActivDEV` is the standard deviation for the evaluation of the activity done during the semester (lab average and practical exam). The first threshold for this value is 1 and the second threshold is 2. We considered that having here only two values they are proper to be closer and therefore we have lower values for th1 and th2.

• `examsDEV` is the standard deviation for exam results (practical and final exam). The first threshold for this value is 1 and the second threshold is 2.

• `averagesDEV` is the standard deviation for lab average, practical and final exam. The first threshold for this value is 1 and the second threshold is 2.

• `activDEV` is the standard deviation for the entire activity (lab marks, practical and final exam). This means that we have here 10 distinct marks in the standard deviation computation. The first threshold for this value is 1.5 and the second threshold is 3.

Table 3. The use of thresholds to demarcate the classes of standard deviation values

<table>
<thead>
<tr>
<th>Classes of values</th>
<th>The use of thresholds to demarcate the classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>acceptable (acc)</td>
<td>$0 \leq val &lt; \text{th1}$</td>
</tr>
<tr>
<td>big</td>
<td>$\text{th1} \leq val &lt; \text{th2}$</td>
</tr>
<tr>
<td>too big</td>
<td>$\text{th2} \leq val$</td>
</tr>
</tbody>
</table>

Another aspect that we considered is the attendance. Our faculty enforces a rule that states that students have to have a 90% laboratory attendance in order to be allowed to enter the examination and/or the re-examination. Having not met this rule, any student is considered to have failed the subject. Thus, in our investigation, students that have the `attendance` attribute have met this requirement.

5. Results and Discussions

5.1. Formal Concept Analysis Results. First we have built a dyadic formal context, where we considered the students as objects and their lab results and their attendance as attributes. A simplified example of such a context is depicted in the Table 4.

For the simplified example we have 3 objects (the students) and 5 attributes (lab marks qualifications and attendance) and a total of 6 concepts:

```python
[["Stud1","Stud2","Stud 3"],[]],
```
Table 4. Example of simplified dyadic context

<table>
<thead>
<tr>
<th></th>
<th>lab i</th>
<th>lab s</th>
<th>lab g</th>
<th>lab vg</th>
<th>attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stud1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud2</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud3</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A concept is actually all the rectangles in Table 4 which are completed with ‘x’ obtained by moving columns and/or rows. Each such concept is then represented in a concept lattice. The resulting form (i.e. lattice) for this simplified example is depicted in Figure 1. Here, each node represents a concept. That is why, having 6 concepts, the lattice representation has 6 nodes.

Figure 1. FCA-based lattice representation for the simplified example
The lattice in Figure 1 is read as follows: starting from a node, the objects (which in our case are students) for that node are collected going downwards and the attributes for that node are collected going upwards. The attribute-labels are placed above a node and all nodes below it (directly or indirectly connected with that node but only by following descending arcs/links) have that attribute and object-labels are placed below a node and all nodes above it (directly or indirectly connected with that node but only by following ascending arcs/links) contain that object. Thus, for the node that has the label “attendance” we can see that we have 2 students, i.e. “Stud 3” and “Stud 2”, as we go downwards to collect the objects, which in our case are students.

For the node that has the label “Stud2” in order to determine its attributes we go upwards observing that such attributes are “lab g” and “attendance”. The label “lab i” placed on the lowest node indicates that no student has that attribute.

The complete data set considered had all of the students (189) and the 5 attributes mentioned above. The resulting lattice is depicted in Figure 2.

![FCA-based lattice representation for the first results](image-url)
We were not able to list all the students as labels below the nodes because the lattice would not be intelligible with a large number of labels under some nodes, but the colour intensity of the nodes denote the number of students (the top node containing all students). As it can be deduced from Figure 2, all the students with good (lab g) and very good (lab vg) lab marks have attended at least 90% of labs (i.e., have “attendance” attribute), while part of students which have sufficient (lab s) and insufficient (lab i) lab performance did not have a satisfactory attendance rate.

That is because, for instance, the node connecting below the nodes with labels “lab s” and “attendance” represent/contain all the students with sufficient lab performance and an acceptable attendance. There are also students that have a sufficient lab performance but not an acceptable attendance. They are depicted in the node with the label “lab s” alongside those with acceptable attendance. The node with the label “attendance” contains all the students with an acceptable attendance regardless of their lab performance. We can conclude then that this rule enforced by our faculty has merit and it is justified by the results we have obtained.

5.2. Triadic Formal Concept Analysis Results. Next we modelled our data in the form of triadic contexts \((G, M, B, I)\) where the object set \(G\) consists of students (as in the dyadic setting), the attribute set \(M\) contains activity qualifiers obtained (i.e., \(i, s, g\) and \(vg\)) while the condition set \(B\) contains the activities for which the students obtained the qualifiers, meaning the lab average, the mark for the practical exam, the mark for the final exam and the final mark obtained by the formula:

\[
60\% \times \text{lab average} + 20\% \times \text{practical ex} + 20\% \times \text{final ex}
\]

A small selection of this triadic context is depicted in Tables 5(A) and 5(B). We have here a \(2 \times 4 \times 2\) triadic context, the “slices” being labeled by condition names.

There are exactly 7 triconcepts of this context, i.e., maximal tridimensional cuboids full of incidences:

```bash
[
    ["Stud 2","vg","practical exam"],
    ["Stud 2","g","lab average"],
    ["Stud 1","g","practical exam"],
    ["Stud 1","s","lab average"],
    ["Stud 1","Stud 2","i","s","g","vg"],[],
    ["Stud 1","Stud 2"],[],["lab average","practical exam"],
    [],["i","s","g","vg"],["lab average","practical exam"]
]
Table 5. Example of triadic context

(A) a. lab average condition slice

<table>
<thead>
<tr>
<th>lab average</th>
<th>s</th>
<th>g</th>
<th>vg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stud1</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud2</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

(b) b. practical exam condition slice

<table>
<thead>
<tr>
<th>practical exam</th>
<th>s</th>
<th>g</th>
<th>vg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stud1</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud2</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

The first four of these triconscepts are proper, meaning that they have all the sets (i.e., objects, attributes and conditions) non-empty.

For our tricontexts we considered all students that had at least one activity, eliminating the students which did not submit any laboratory work. Therefore, for section $S_1$ we have 67 students (objects) and for $S_2$ we have 84 such students.

Having three sets (i.e. objects, attributes and conditions), the representation of a triadic context is tridimensional, and therefore it can be represented as a trilattice, which can be hard to navigate. Therefore, the FCA Tools Bundle allows analysing triadic contexts by projecting along one of the dimensions of a triadic concept (i.e., the extent, the intent or the modus) in order to obtain a dyadic context which can then be represented as a bidimensional lattice which is easier to navigate. This is done starting from one triadic concept and “locking” on one of the dimensions by setting that dimension (i.e., the extent, the intent or the modus) with the set of values within that triconcept. By right-clicking on the nodes (representing concepts) of a lattice obtained in this manner, one can see the triadic concept associated with it. From this point on, one can lock on a different dimension in the triadic concept and generate another bidimensional lattice. One can analyse thus the triadic context by repeating this process. In Figure 3 we locked to see only the very good performances in one or more activities. As it can be seen in Figure 3(A), all of the $S_2$ students that had a very good performance on final exam ($ex$) did also very good during laboratory activity ($lab$) and also had very good final marks ($fin$). Moreover, a very good performance within the practical examination ($prac$) reflected in very good final results ($fin$) only for students that have done very good also in laboratory activity (this is depicted within the lattice by the common node next to the node having the label “ex”). For the students in $S_1$ (results depicted in Figure 3(B)) all the students with very
good lab performances (lab) had very good final marks (fin). The results for students in S1 are different from the results for students in S2 considering the perspective that students with very good lab performances (lab) had very good final marks (fin), but not necessarily a very good performance on final exam (ex). That is due to the fact that (as mentioned in Section 4) students in S1 were more closely monitored during the exam, and as it usually happens some students do not perform well under stress.

![Diagram](image)

(A) Results for students in S2  (B) Results for students in S1

**Figure 3.** Results for the students that had a very good performance

From these investigations results a strong correlation between the work involvement of students and their final results.

Another triadic context considered was the one where we have students as objects, classes of values depicted in Table 3 as attributes and the standard deviations described in Section 4 of the paper (i.e., labActivDEV, semActivDEV, examsDEV, averagesDEV and activDEV) as conditions. The results showed that there is a strong correlation between labActivDEV and activDEV.
Figure 4 shows for instance that $S_2$ students with similar class of $labActivDEV$ and $examsDEV$ had only either big or acc (acceptable) standard deviations. The label “too_big” placed on the lowest node denotes that no student had both $labActivDEV$ and $examsDEV$ too big. All the students that do not have the two deviations either “big” or “acc” are contained in the topmost node.

The same scenario as the one depicted in Figure 4 takes place when we group any of the five standard deviations considered and/or if we consider them all.

Other such similarities can be deduced using FCA. But the conclusions drawn from the result showed that general consistency in the students’ work generated predictable outcomes.

Finally, we wanted to see how close is our final average (computed as $60\% \times lab\ average + 20\% \times practical exam + 20\% \times final\ exam$) to the actual average of all evaluations (i.e., 8 lab marks, 1 practical exam, 1 final exam) by computing the standard deviation between the two. Almost all the values obtained were less than 0.5 for this standard deviation, the actual percentages are detailed in Figure 5. We were also interested in how good the decision was to change the previous final average formula (computed as $20\% \times lab\ average + 40\% \times practical\ exam + 40\% \times final\ exam$), denoted as “prev” in Figure 5. On the $Oy$ axes we have the percentage of students and on the $Ox$
axes we have the value of the standard deviation. As it can be seen, the current computation formula for the final mark for both sections reflects better the students’ performance than the previous one.

![Comparison between the previous formula of computing the final mark with the current formula by measuring (through standard deviation) how close are they to the student evaluations](image)

**Figure 5.** Comparison between the previous formula of computing the final mark with the current formula by measuring (through standard deviation) how close are they to the student evaluations

6. **Conclusion and Future Work**

During this Covid pandemic we were forced to have our activity online. This new arrangement brought new perspectives and new challenges, both of which have as well advantages as disadvantages.

As advantages we can mention:

- faster and easier communication
• more electronic teaching materials which can be consulted at any time - any place - any pace

Disadvantages and challenges that we observed were:

• students are not as focused as in a classroom
• there is non-verbal communication that does not apply in a virtual meeting (for instance an experienced teacher can detect at a glance if the concept presented was understood or more time/effort/examples are required).
• evaluation tests can not be monitored as closely as in a classroom, as there are students (especially in our domain of activity - which is computer science) who are really creative.

In this paper we wanted to address the main question: Did the consistency or inconsistency during the semester (i.e., in the lab work) affected the students performance on the exams (practical exam and/or on the final exam)? In order to answer this question, our investigation was fourfold.

First, we investigated the students’ involvement from the perspective of their attendance by using dyadic FCA. The results showed that all students with good and very good lab performance meet the attendance requirements.

Second, we used triadic FCA to see the correlation of students’ activities in the case of very good performances. Our results showed that all S1 students with very good lab performance obtained very good final marks, while all S2 students with very good performance in the final exam did also very good in lab and had very good final marks.

Third, we used triadic FCA on standard deviations of students’ activities. From this investigation we observed that a standard deviation that we considered to be “too big” cannot be observed consistently though all activities (i.e., lab, practical exam and final exam).

Fourth, we observed that the current formula used to calculate the final mark (changed during the Covid imposed online period) gives a better appreciation of the students’ performance than the previous one (used in the traditional face-to-face setting).

The work presented here was mainly focused on good and very good performances. Further investigations can be focused on sufficient and insufficient performance to try to determine other causes for such results.

Moreover, in order to address the disadvantages mentioned above, we would like to conduct a comparison with pre-Covid results and further on maybe to post-Covid ones.
REFERENCES


Faculty of Mathematics and Computer Science, “Babes-Bolyai” University, Cluj-Napoca, Romania

Email address: sanda.avram@ubbcluj.ro