Evaluation of the VL Logic (342.208-9) 2018W End of Semester Questionnaire

David M. Cerna

Abstract

In this technical report we cover the choice of layout and intentions behind our end of the semester questionnaire as well as our interpretation of student answers, resulting statistical analysis, and inferences. Our questionnaire is to some extent free-form in that we provide instructions concerning the desired content of the answers but leave the precise formulation of the answer to the student. Our goal, through this approach, was to gain an understanding of how the students viewed there own progress and interest in the course without explicitly guiding them. Towards this end, we chose to have the students draw curves supplemented by short descriptions of important features. We end with a discussion of the benefits and downsides of such a questionnaire as well as what the results entail concerning future iterations of the course.

1 Introduction

2 The Questionaire

In Figure 2 one can find the questionnaire we provided to the students prior to the first minitest of the third Module. The most striking element of the questionnaire is the space on the upper lefthand side left for the students to draw curves depicting their understanding and interest in the course over the semester. Furthermore, we provided space for the students to highlight important features of the curve and their relation to the course content (In particular Question 2a and 4a of the Questionnaire). If we isolate one of these empty coordinate planes (see Figure 1) one will notice that the provided space is divided into three sections, one for each Module of the course. Information concerning the particular meaning of these divisions was not provided, though we intended one curve which spans the three divisions. Nonetheless most students drew a spanning curve rather than an individual curve for each Module. As we will discuss in a later section, we took this choice as an indication that the course material, per Module, is distinct enough that some students considered them as entirely separate entities, something we would like to change in future iterations. Before, discussing other aspects of the Questionnaire, we will focus on one particularly pressing issue, namely, how does one interpret the imagination of the student in a formal and meaning fully way.

3 Interpreting the Curves

Because we gave the students the ability to express themselves rather than filling in a traditional questionnaire with fixed answers, we ended up with very high participation,

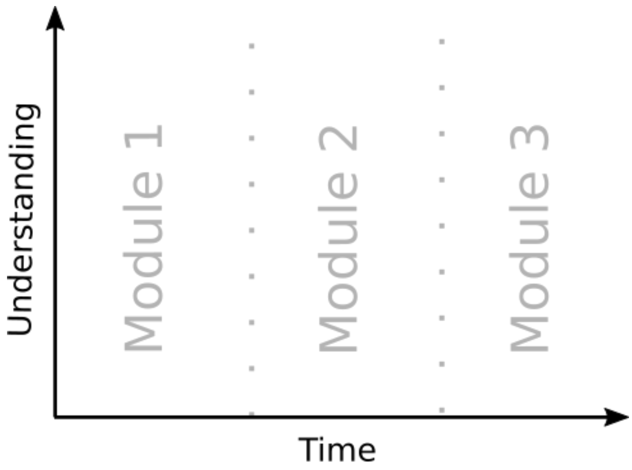


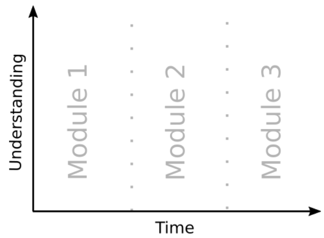
Figure 1: The coordinate plane provide to students for illustrating their understanding as a curve.

134 students filled out the Questionnaire and of those, 131 drew curves depicting their understanding and interest in the course over the semester. Though this freedom of expression had the caveat that students could draw curves which challenged our ability to interpret them in a meaningful way. Moreover, those with more intricate drawings more likely had something meaningful to say about the course. To better illustrate what we are referring to consider the curves depicted in Figure 3. These represent curves which are easy to interpret. Obviously, the curve on the lefthand side of Figure 3 denotes high understanding through out the course while the curve on the righthand side denotes difficulties within the second Module.

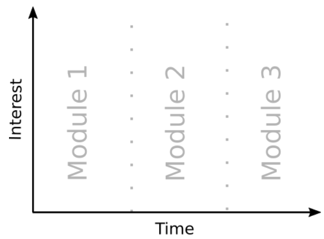
However, even in the case of these “good example” there are interpretation problems. For instance, we know that the curve on the lefthand side of Figure 3 represents high understanding, but how high with respect to the righthand side. Furthermore, how low did the students understanding (referring to the righthand side of Figure 3) go in Module two with respect to Module one and three. These issues are easy to deal with by enforcing a scale with finite values upon which all illustrations can be measures. However, the real issue arises when we look at the more “creative” illustrations of students as depicted in Figure 4. While the lefthand side of Figure 4 is in an abstract sense similar to the righthand side of Figure 3, there is so much fine structure which makes it difficult to enforce a scale on the illustration. This problem is push to the extreme in the illustration found on the righthand side of Figure 4 where not only is there a lot of fine structure, but it all seems significant.

These issues forced us to develop a categorization of the illustrations defined by the amount of structure present in the image. As one can see, the lefthand side of Figure 3 represents the complete lack of fine structure while the righthand side of Figure 4 represents an overwhelming amount of structure. Our categorization, from simple to complex is as follows: **Constant**,**Linear**, **Parabolic**, **Saw Tooth**, and **Other**. It is quite easy to identify the categorization of lefthand side of Figure 3 and the righthand side of Figure 4, **Constant** and **Saw Tooth** respectively, but for the other two, a choice has to be made. Do they, abstractly, represent the same type of object. In our analysis, we will refer to both as **Parabolic** being that, as one abstracts away the fine structure, one ends up with a definite parabolic shape in both cases.

End of Semester Questionnaire



1) Illustrate your **understanding** of the course as a curve.
2) Think of something offered by the lecturers that added to your **understanding** (software, exercise, etc.)
   a) Place on the curve using a • labeled by (I, II,or III):
      I
      II
      III
   b) Can you suggest something which would have aided your **understanding**?



3) Illustrate your **interest** in the course as a curve.
4) Think of something offered by the lecturers which increased your **interest** in the course material? (e.g. software, exercise, examples, etc.)
   a) Place on the curve using a • labeled by (I, II,or III):
      I
      II
      III
   b) Can you suggest something which would have increased your **interest** in the course material?

5) Did you attempt bonuses? **YES** **NO**
   a) If **YES**, How many?
      $\backsim \frac{1}{4}$ $\backsim \frac{1}{2}$ $\backsim \frac{3}{4}$ **all**
      and why?
   b) If **NO**, why?

6) Which software did you use? (mark all that apply)
   Limboole RISCAL Theorema SMT Solver
   a) Which software was helpful (why)?
   b) If you had particular problems with a piece of software please explain below.

7) Would you like the introduced software to be part of your other courses in the curriculum? **YES** **NO**
   a) If **YES**, which ones?

8) Would you like to see more examples such as the box game in the course? **YES** **NO**
   a) If **YES**, for which topics?

Figure 2: The end of the Semester Questionnaire provided to students.

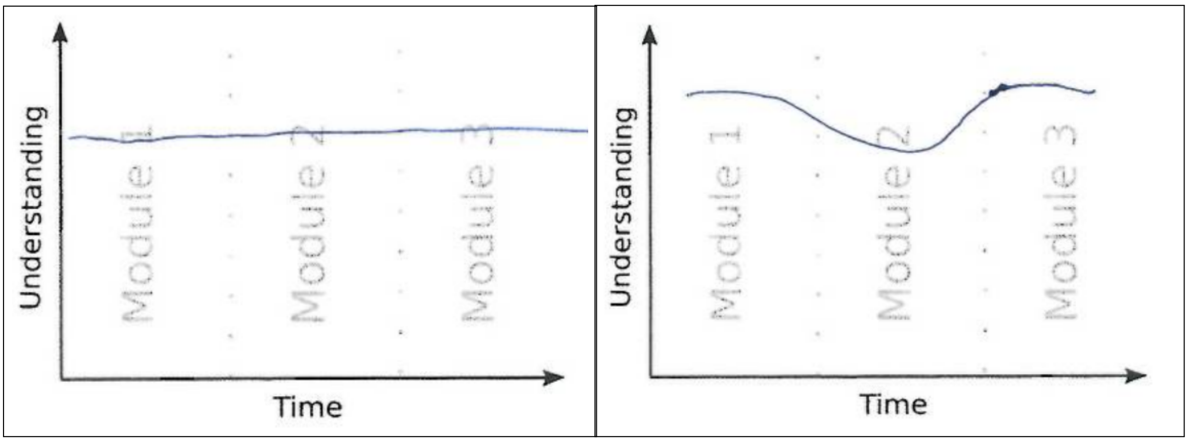


Figure 3: Examples of student curves which are easy to interpret.

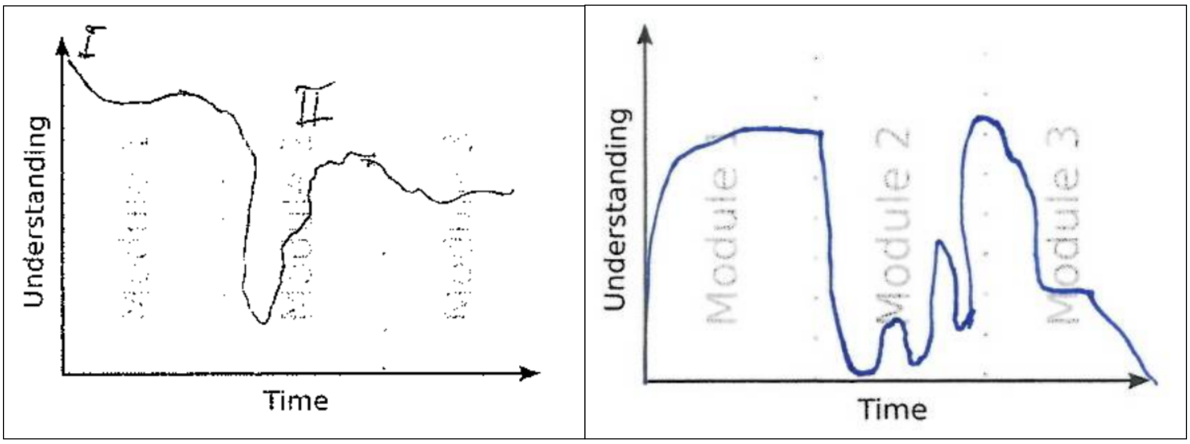


Figure 4: Examples of student curves which are hard to interpret.

To some extent this deals with many issues concerning categorization of the students depictions, however, in the process we have removed much of what the student wanted to tell us. While there is no argument that, abstractly, the righthand side of Figure 4 is a **Saw Tooth** curve, it is not **Saw Tooth** in every Module of the course. Moreover, the student's understanding increased significantly during the first Module and decreased significantly during the third Module. It is only when observing the curve holistically or during the second Module that the **Saw Tooth** behave manifest itself. Thus, we further distinguish the curves by a classification of their behavior on a per Module basis.

Distinguishing curves on a per Module basis can be easily done by repeating our classification, but concerning the three modules individually. This would classify the righthand side of Figure 4 as **Linear**, **Saw Tooth**, **Linear**. This captures finer structure but one again we lose some essential parts of the message the student tried to communicate to us. Yes, both Module one and Module three are **Linear**, but with the opposite slope! Thus, rather than repeating our curve classification at the Module level, we take note of particular properties of the curve within that Module, namely, the **Slope** (either **-1**,**0**, or **1**), the **Maximum** (either **0**,**.25**,**.5**,**.75**, or **1**), the **Minimum** (either **0**,**.25**,**.5**,**.75**, or **1**), **Jump** (either **yes**,**no**), and **Drop** (either **yes**,**no**).

Concerning **Slope**, we are not interested in the particular variations within a given Module, but rather the overall change from the start till the end. Let us consider the situations illustrated in Figure 5. Notice that a curve may locally be assigned any of the possible values but it is only the slope between the start and end position which determine the value of **Slope**. We are attempting to capture the tendency rather than the path from start to finish. The particularities of the curve are captured by the other four characteristics.

Concerning **Maximum** and **Minimum**, the concepts themselves are straight forward, however, we did not provide the student with our limited scale prior to their filling out of the questionnaire. Thus, We interpret precise position of **1** vaguely, that is with respect to the actual illustration provided by the student. Consider the examples provided in Figure 5. While for (A2) we can guess that the student started with high understanding, we my ask is it **.75** high, or **1** high? Our position on this matter is that the student did not mean to make such a distinction. They merely wanted to tell us that as the course progress there understanding dropped. Thus, from our interpretation, the only correct choice would be a maximum of **1**. A similar situation is faced for assigning the **Minimum**. Once again concerning (A2), we my ask is it **.25** low, or **0** low? Again, the student most likely did not reflect on this issue, and in all likelihood does not want to say they understood nothing at the end of the course. If they really wanted to state this, they would have made it more explicit, as some of the students did. Thus, we will interpret such an illustration as have a **Minimum** of **.25**.

One more interesting point concerning **Maximum** and **Minimum** is illustrated in (C2) of Figure 5. While it is apparent that the start point is lower than the end point, hence the **Slope** of **1**, it is not entirely clear what the minimum value ought to be. It depends on our value assignment to the start and end positions. By inspection, One may set the start position to **.5** and end position to **.75**, however the clear choice of **.25** to **Minimum** does not capture how significant of a drop occurred. Thus, in such cases we will assign the start position **.75** and end position **1** as the assignment of **.25** to **Minimum** can now be distinguished from a subtle drop to a significant one. In this particular case we where left with an opportunity to capture the students intentions without additional characteristics. Unfortunately, it is not always possible to denote significant changes

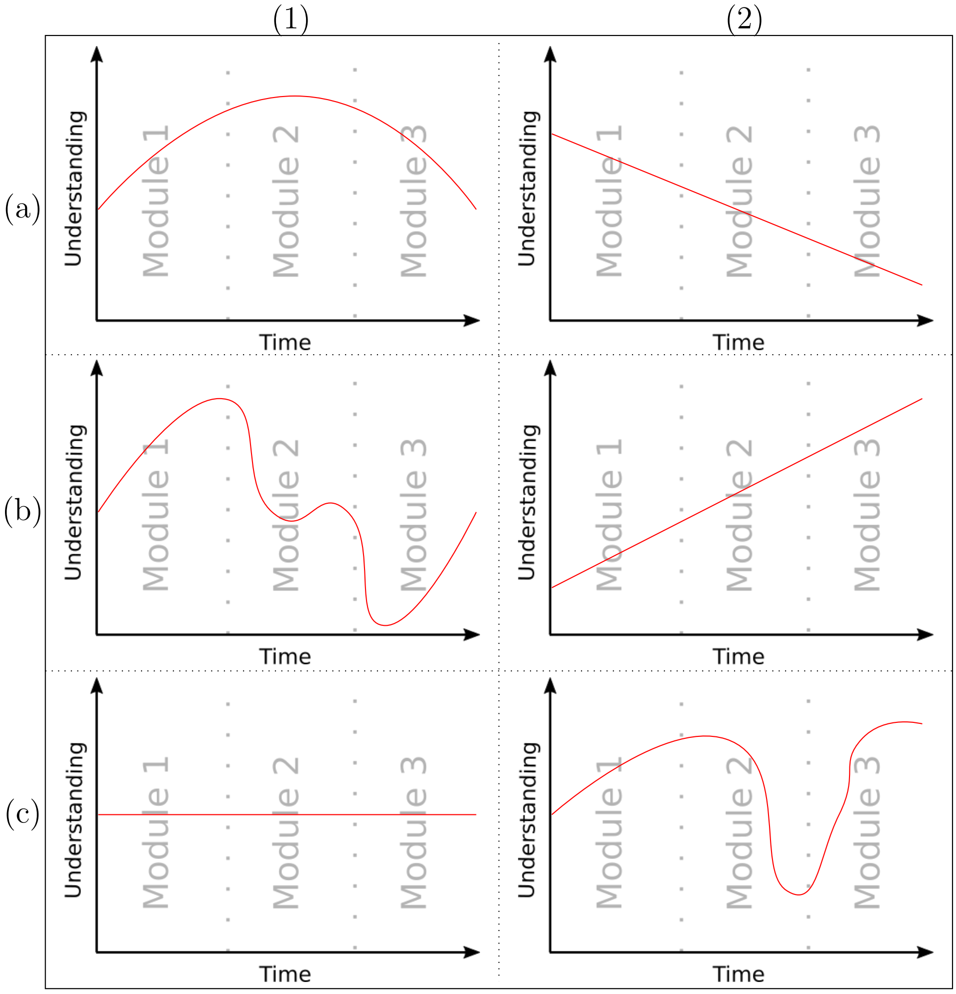


Figure 5: Illustrations demonstrating possible variations and there associated **Slope**, Namely, **Slope** (A1) = 0, **Slope**(A2) = -1, **Slope**(b1) = 0, **Slope**(b2) = 1, **Slope**(C1) = 0, and **Slope**(C2) = 1.

using just **Maximum** and **Minimum**. This justifies the introduction of our last two characteristics, **Jump** and **Drop**.

The intention behind **Jump** and **Drop** is to capture dramatic changes within the curve illustrations. while in some cases, both of these characteristics may be capture by the previously mentioned characteristics, there are some cases where there is too much information within the illustration to interpret it without **Jump** and **Drop**. For example, the righthand side of Figure 4 has an extreme shape, but only in the second Module only. To make matters worse, let us assume it has **Slope -1** in the second Module and its **Maximum** and **Minimum** are **1** and **0** respectively. This would make it hard to tell apart from the third Module unless we state that both **Jump** and **Drop** are **yes** in the second Module, but only **Drop** is **yes** in the third.

This covers our interpretation of student curve illustrations. We will leave analysis of the data to a later section. In the next section we discuss the interpretation of there answers to questions 2a) and 2b) as well as 4a) and 4b) (see Figure 2) which while being free-form ended up having more uniformity than the curve illustrations.

Solutions	10
Easy Material	1
Difficult Material	1
Particular Lecturer	7
Examples	13
Exercises	39
Software	13
Moodle Quizzes	26
Box Game	8
Bonuses	12
Particular Topic	2
Minitest	4
Weird	5

Table 1: Answers to question 2a) of questionnaire.

4 What Influenced Student Understanding and Interest

As mentioned earlier, an important consideration we took when designing this questionnaire was "how to make students fill it in?". Part of the problem with such surveys is the lack of expressive power one has when there is a fixed rubric one must follow. Problematically, free form questions, i.e. "write what you want", can also be daunting task as they require describing difficult concepts. Of course, by removing a strict rubric, as we have seen in the last section, the result is increased complexity on the side of the interpreter. Interestingly, while we left significant room for self-expression in 2a), 2b), 4a) and 4b) (see Figure 2), there was a noted lack of enthusiasm in the student desire to answer these question compared to the illustrating activity associated with 1a) and 3a) (see Figure 2). As mentioned earlier, out of the 134 student who filled out the questionnaire, only 3 did not draw a curve. However, the number of student who did not answer 2a) was 54 (41%) and the number of students who did not answer 4a) was 76, over 50%!

Given that the total number of answers a student could have provided is 3 for each question, There are a total 402 answers which may have been provided. Concerning question 2a) where a total of 80 student provided answers, each student provided an average of 1.76 items which influenced there understanding. Concerning question 4a) where a total of 58 student provided answers, each student provided an average of 1.5 items which influenced there understanding. Apparently, describing what particularly stirred your interest in the course is more difficult than we thought it would be.

While with such low numbers of participation it is hard to gauge what precisely was beneficial to understanding and interest, one can at least gather patterns within the answers. We will cover these patterns in the two subsections below.

4.1 Understanding: Necessity of a Feedback Loop

In Table 1 we provide a list of the common answers of students to the question "what influenced your understanding of the course material?" The answers are not precisely what the students wrote, but are abstractions capturing the idea of the students answer.

From example “Particular Topic” would replace the topic the student named if we where to be completely faithful to the students response. Note that these categories where not compiled before we analyzed the questionnaire, but after we compiled all the data. Our goal was to find a set of answers which capture all student responses. Of course this was not entirely possible, and thus we have the weird category which cover answers such as “friends” and “Wikipedia”, neither of which was provided as course material. This is true for all answers in this category and thus it can be safely ignored.

The most interesting observation concerning Table 1 is Exercises holding the top spot at 39 responses matching this description. This is roughly 50% of the students who provided a response. Second place is held by Moodle quizzes, that is non-graded practice test designed to aid the students through topics they would see in the exercises as well as the minitest. Only Module one had Moodle quizzes being that they were an experimental concept of one of the participating lecturers. The somewhat significant position they hold concerning their influence of student understanding ought to be taken into consideration for future iterations of the course.

A bit more surprising is the position held by weekly challenges, i.e. bonus exercises. Unlike the Moodle quizzes, weekly challenges were graded and each Module contained its own set of weekly challenges. Most weekly challenges also had associated software, which if one refers to Table 1 received roughly the same recognition. We will get back to this issue latter in this report when discussing question 5, but the short answer is that weekly challenges provided points and not understanding, one can even say they hampered understanding by forcing the students to do extra for points even when they failed to understanding the initial material.

Returning to exercises, the holder of the top spot, we need clarify what is meant by exercises to provide an accurate picture of the influence this particular activity had on the students understanding. By exercises, the students are not referring to home assignments, but rather lab assignment which are performed with feed back from knowledgeable persons. Putting this eloquently, this is a form of formative assessment. Moodle quizzes partially fall into this category as well because they give quick assessment as if a knowledgeable person were present. Of course, this is a limited form of assessment, but given that the activity holds second place points to the veracity of this observation. Notice that if we continue this line of reasoning a little further solutions and examples both fall into this category of formative instruction.

Overall, what we are told by question 2a) is that some sort of feedback system seems to strongly influence the understanding of the students. Beyond all the above examples, a repeatable and self constructible feedback loop exist for Module one in the form of the box game. Though, this game has not been implement, i.e. it has so far been played physically, nor has it be extended to the other Modules, the results of this section strongly justify such investigations and developments covering other topics in the course.

4.2 Interest: Hands-on Aids the Feedback Loop

As we already mentioned earlier in this section, there where far fewer answers from the students concerning question 4a) when compared to question 2a). Thus, there isn't much we can draw from the following data other than a few weak correlations and the fact that students are not too interested in thinking about what interest them, at least with respect to this course. We do have to consider the fact that level of understanding is an intrinsic part of any course while interest in the may be completely lacking when the course is

Solutions	1
More Applied	4
Particular Lecturer	4
Examples	11
Exercises	17
Software	18
Moodle Quizzes	9
Box Game	7
Bonuses	8
Particular Topic	7
Weird	3

Table 2: Answers to question 4a) of questionnaire.

taken as a mandatory part of a curriculum. We might want to look ways to catch the interest of the disinterested.

First point to note concerning Table 2 is that Exercises holds a high position as it did in Table 1, though it does not hold the highest position which is held by Software. What one could gather from this is that while the understanding provided by the feedback loop present in the exercise labs does translate into increased interest in the course, software, actually implementing the course material provides just as much increase in student interest. However, the current software does not provide much in the way of understanding compared to the feedback loop present during exercise sessions. This most likely concerns the nature of the software currently introduced during the course. For example, SAT solvers and SMT solvers do not provide much benefits to the users in terms of understanding until the user actually understands the processes behind SAT and SMT. Concerning Theorema, one again we have a tool making the process of formal proof easier, but if one does not understand what a formal proof is, then one is lost. And once again with RISCAL, a understanding of the fundamental concepts is needed to fully appreciate the software. Thus, the software introduced is great at capturing the attention of those who already understand, but will not provide the missing links to those who are currently at loss. We will get back to this problem in the next section when we address question 5 which concerned the weekly challenges.

5 Weekly Challenge: Worth the Effort?

Weekly challenges where added to the course as a way to introduce material, i,e, software, which may negatively impact the efficacy of the course. The problem with the introduction of the particular software added to the Logic course is that much of it is external to the course material and its introduction is an experiment which may fail. For example consider Limboole, while the satisfiability problem is a key component of the course as well as methods of solving it, knowing how to use an existing SAT solver is not directly necessary and is more of a learning aid. If the students find installing, setting up, and using Limboole difficult to the point that it takes up time better spend on learning the actual course material the net effect of adding the software would be negative. While this only concerns this course at first, given that the course is part of the foundation for

Attempted Weekly Challenges		Why?	
yes	108	Easy	6
		Points	51
		Interest	8
		Understanding	8
		Interest and Understanding	1
		Points and Interest	10
		Points and Understanding	10
		Time	1
		Why not?	
no	26	No time	9
		Not worth effort	4
		Do not need points	9
		software difficulties	1

Table 3: Answers to question 5 of questionnaire.

an entire curriculum, this negative effect may propagate through the curriculum, or even worse, impact their desire to continue the studies. While this was luckily not the case with Limboole, software such as Theorema seems to have at a net negative impact on students who had problems installing, setting up, and using the software with over 50% of those who used it complaining about some issue with it. Thus, experiments in such an environment may only be perform through something external to the graded course material, i.e. extra credit. But of course, when given a choice student ask, is it worth the effort? This question seems to have defined the students approach to the weekly challenges.

As one would expect, the main motivation for the students to use the software was not out of interest or understanding, but either to protect themselves from future poor marks, or current poor marks. Out of the 108 who stated that they handed in at least one weekly challenge, 71 stated that they did it for the bonus points. Only 19 stated that they did weekly challenges for understanding, and 11 out of interest. See Table 3. Interestingly, the average number of weekly challenges a student completed was 50%. We do not have any clear reason why students stopped after completing 50% of the weekly challenges, though we noted a bias towards completion of earlier challenges vs. later ones. But most likely, given the numbers in Table 3, they stopped when they had enough points. This hypothesis is further evidenced by the "no" section of Table 3 which clearly states that once points where no longer needed or not enough time was available the students decided to skip on the weekly challenges. Given that weaker students are precisely the ones who take longer to perform the assigned work, they were most likely the ones skipping out on the weekly challenges, i.e. they were challenged enough already.

Note, that the questions "why?" and "why not?" where meant to be associated with the respective answers "yes" and "no". However, many students took a different interpretation and instead answered both questions even if they did complete some of the weekly challenges, thus providing us with extra and unexpected data. We will take this into account when designing future Questionnaires (See Table 4). While some students refrained from answering both questions, what we see is a rise in the number of students who did not do weekly challenges for reasons concerning time and their current grade.

		Why not?	
no	26	No time	21
		Not worth effort	17
		Do not need points	11
		software difficulties	6

Table 4: Student answers to "why not" including students who attempted weekly challenges.

5.1 Students Reactions to Weekly Challenges

Of the 108 students who attempted weekly challenges, the average completion was roughly 50% of the total number of challenges. This shows that they were both able to comprehend the task, this includes installing software, and completing it. Thus, the software did not impede on their ability to solve the task. While this holds for a majority of the provided software, Theorema lags behind in that most of the students complaints where targeted at it. The issues that students raised fall into two categories, namely, too hard to use and Mathematica is not free. On a rare occasion RISCAL was included into this category of too hard to use, but mostly concerning software problems such as installing the necessary virtual machine and not the interface itself, nor documentation.

In particular concerning RISCAL, which is offered to the student at the crux of the course, The weekly challenges did not really benefit those who were unable to comprehend the material presented. A mere three students who where failing at the point when RISCAL weekly challenges where presented passed the second Module by completing them. Essentially, the software did not introduce dramatic change in students comprehension, nor did it hamper it. Similar, statistics may be observed with respect to the other modules. What was evident is a shift in the number of middle range grades (grades of 2 or 3). This seems to imply that the majority of students completing weekly challenges used the challenges to raise their grade in the course. While this does not rule out an increase in understanding as well, there are no data to corroborate this conclusion. Thus, all we may conclude is that average students seem to have a positive reaction to the weekly challenges and software as long as the amount of effort needed is low. This is not the most surprising result, but it does entail that integrating software use into the course ought not harm the students comprehension of the material.

6 Software: A Theorema Problem

The student were introduced to three particular pieces of software, Limboole, RISCAL, and Theorema, and for the third Module were allowed to use any of the existing SMT solvers. As the semester progressed there was a drop in the number of students who used the introduced software. Limboole, being the first introduced was used by 121 students (out of the 134) followed by RISCAL with 82 students. Both Theorema and a SMT solver of choice had roughly the same number of users at 51 and 54, respectively. These numbers where taken from the answers to question 6 of the questionnaire. Taking into account some of the points mentioned earlier, a decrease in the number of student using a particular software might have external explanations, that is external to the software. For example Limboole is the first software introduced. At the beginning of the semester students may be open to learning a piece of software and have more free time. While SMT solvers are introduced at the end of the semester when students suffer from a lack of time as well as decide not to participate in the course. Also, one should note that in the previous section we realized that students used the software for points, thus if they reached their goal grade they no longer have much incentive to complete more weekly challenges. This pattern shows up in the final grade more than one would expect.

Thus, prior to external factors it would seem that both Theorema and SMT solvers cause the most problems for students based on the number of users. However, contrary to this belief, very few complaints where levied at SMT solvers concerning students problems with the software, while most complaints were levied at Theorem in terms of the complexity of use. The most common response to question 6b) of the questionnaire was that Theorema is hard to use with 19 responses, followed by installation of Theorema, Theorema Documentation, and Theorema general issues, with 6 responses each. Thought, 74 of the 134 students who completed questionnaires left this question blank, we did phrase it as "answer if you had particular problems". Furthermore, only 51 of the students stated that they used Theorema in the first place. Thus, 73% of the students who used Theorema had issues with the software. RISCAL had the second most complaints at 12% of users. What this shows is that if Theorema will be integrated into the course, i.e. not as part of a weekly challenge, it will require addressing the issues students had with it during this iteration of the course.

6.1 Software: Don't really want more of it

While students did use the software, and for the most part, sans Theorema, things went smoothly, the majority of students do not want to see the software used in other courses. Only 41 (31%) would actually like to use the software in other courses. Given the problems for Theorema to tackle mentioned in the previous section we would like to make a note that 11 students did say they would like to use Theorema again in other courses. Only Limboole had more interest with 23 of the 41 stating they would like to see it again, while RISCAL came in third with 10 students. While many hated Theorema it seems that those who liked it, really liked it.

However, overall it is quite unexpected that so few students would want to use the software again during their studies. This seems to imply that they did not find it helpful, which, paradoxically, is why software was introduced, that is to help the students. It seems more has to be done to make the students understand the uses of the introduced software because at the moment it seems they take it more as a necessary evil rather than a study aid.

7 Gamification: The Box Game Sparks Interest

The only point which students seem to unanimously agree on is the presence of the box game within the course. Out of the 134 students who filled in the questionnaire 93 (69%) would like to see more box game like activities in the course, with the majority, 51 out of the 93, wanting a box game like activity for first-order logic. We will go over our analysis of the understanding and Interest curves in the next section, but as a short teaser, a majority of the students showed a drop in understanding precisely at the point when first-order logic is introduced. This is the most difficult topic covered during the semester and is most likely a topic the students have no acquaintance with. This suggest a line of investigation into the gamification of other topics introduced during the course. For example, unification, derivation tree construction, and proof tactics.

8 Analysis of the Curves: Understanding

In Figure 6 we illustrate the number of students who drew each type of curve discussed in Section 3. Notice that the majority of students drew either a parabolic or a saw tooth curve thus implying that certain topics caused problems for those students. In Table 5 we provide the actual numbers for each curve category. Concerning student illustrations of the linear type, the majority of the 32 curves categorized as such, i.e. 26, had an overall negative slope and showed minimal understanding in the third Module. There is nothing particularly off-putting with these results as one expects the course material to become more difficult as the semester progresses. However, what is strange, as we shall see soon, is that a decent number of student saw the third Module difficult. This is counterintuitive being that many comments where levied at the second module. This result may imply that a particular concept from the second module was difficult for these students, namely unification and term manipulation, concepts needed for SMT.

Concerning the 46 students who drew parabolic curves, the majority of students draw curves whose minimum was located in the second module. Notice that in Figure 7 there are more minimal values than there are students who drew parabolic curves. This strange occurrence is present being that some students drew a minimal value at the border between two modules. This, more often than not happened between the second and third module, thus, to some extent covering the issues noted for linear curves. This tells us that roughly 86% of the students who drew parabolic curves had understanding issues in the second module. Together with the negative slope of the linear curves this shows that understanding seems to drop once terms, proof construction, and quantifiers are introduced.

Before moving on the saw tooth curves we would like to address the one point missing from this analysis, the concavity of the parabolic curves. While it is obvious from the current analysis that the curves where most likely concave down, it would be interesting to note how many of the students had concave down parabolic drawings and where this concavity evened out, the slope inverted.

Only one student showed concavity in the first module which evened out prior to the second module. This most likely implies that the student had never seen a logical formula before and once they got use to it all was fine. In Module II however, we have 23 students showing concavity of some sort of which 5 students pass over the border to the third module before the concavity evens out and 18 of which even out within the second Module. No one has positive concavity at this point being that there is no curvature present from the first module to the second. The one data within the third module concerns the positive concavity passed from the second to the third. This overall implies that the majority of the student drew parabolic curves which concaved up, that is reach a single minimum and have two maximum.

The students who drew saw tooth curves had a much more even distribution of the minimum attained value most likely because they reached it in multiple places. However, there is still a bias towards the second and third module. See Figure 8. This bias towards difficulties in the second and third module can also be seen in how the curves bend, once again, their concavity. Notice that in Figure 8 that most minimum values are found within Module two. Not only is the part of the curve where concavity is most likely to be present

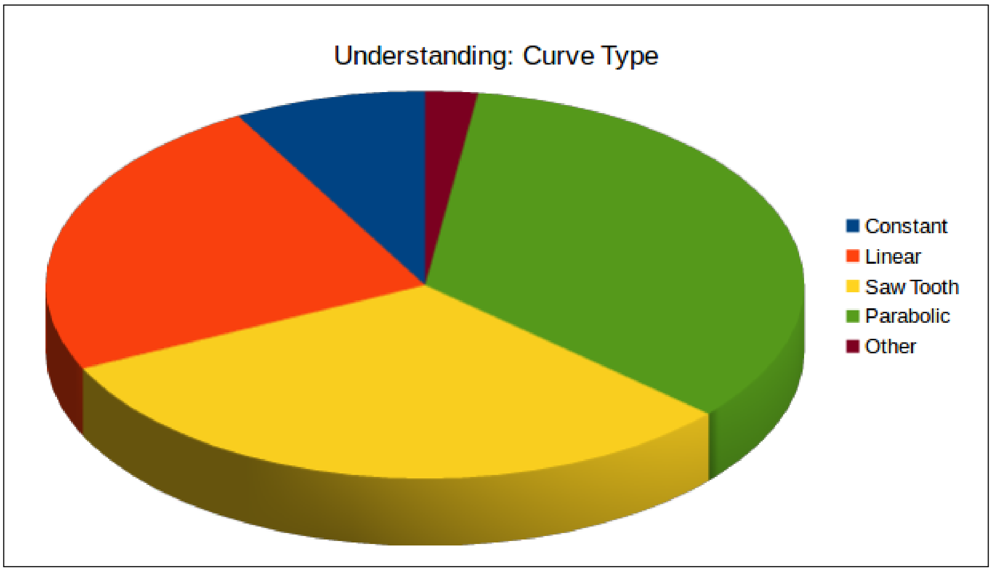


Figure 6: Types of curves drawn by students representing their understanding of the course material.

Constant	11
Linear	32
Saw Tooth	42
Parabolic	46
Other	3

Table 5: Student Illustrations divided into curve types.

is in Module two, it is also the Module where an overall concavity up is present. This implies that upon reaching Module three the curve is still bending downwards. Unlike for parabolic curves this downwards curvature even leaves the bounds of the graph with 17 students showing concavity in the third Module and 10 of those students showing concavity up. While this seems hard to interpret, being that the data is not as uniform as it was for the other curve types the implication is quite obvious, the end of each Module was difficult for students who drew saw tooth curves while the beginning of each module provided relief.

In some sense this is good in that each of the modules is a separate entity, but at the same time, at least for some of the students, there is not much knowledge carried over from the earlier modules. At least, not much knowledge which they know how to apply to the new material. It might be worth making the course a bit more cohesive. We conjecture that doing so would reduce the number of students with saw tooth curves.

9 Analysis of the Curves: Interest

Unlike the student illustrations for understanding, the illustrations for interest where most likely to be linear or parabolic making up about 66% of all the illustrations. See Figure 9. A very interesting observation concerning student interest can be found in Table 6. The minimal value for students who draw linear curves is most likely to show up in the first

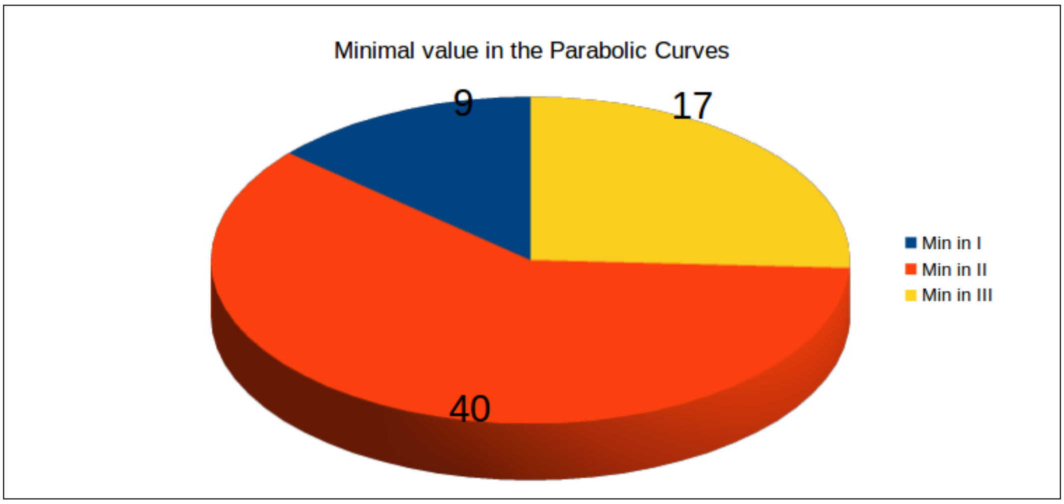


Figure 7: Number of minimal values per module for parabolic curves.

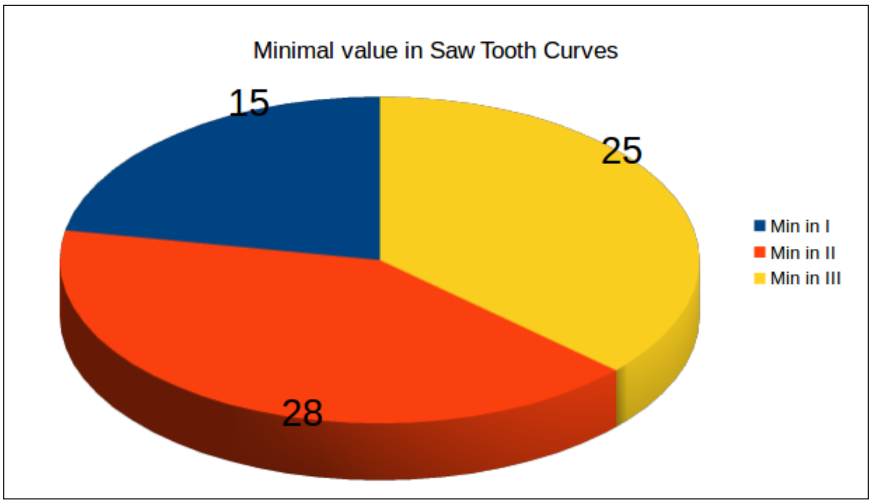


Figure 8: Number of minimal values per module for saw tooth curves.

Type	Number	Min in Module 1	Min in Module 2	Min in Module 3
Constant	19	19	19	19
Linear	43	20	13	27
Saw Tooth	28	17	20	15
Parabolic	41	7	34	13
Other	3	0	0	0

Table 6: Student Illustrations divided into curve type and minimal value location .

or third module while the minimal value of student who drew parabolic curves is most likely to show up in the second module. Saw tooth curves follow a similar pattern as the parabolic curves but less significant while constant curves can be grouped with linear ones. This shows a strict separation within the course between students who loss interest because of Module two and those who either gained, loss, or where indifferent, irrespective Module two. Given that there are slightly more students that had a minimum value in the Third Module and drew linear curves, we can state that interest was slowly loss over the course, but it is a difference which is not significant enough to go into further detail. Note that the number of students who loss interest in the course because of module two, the saw tooth and parabolic drawings, make up 40 percent of the students, a significant number.

Concerning the concavity of the students who drew both parabolic and saw tooth curves, the majority drew curves whose bend occurred during the second Module, as expected. In most cases this bend evened out before the third module for parabolic curves, 20 concave drawings of which the total number of curves with concavity within Module two and evened out in Module two was 18. For saw tooth curves there were 19 concave drawings in the second Module of which half evened out in the second Module and half evened out in the third Module.

Overall the picture is quite clear, not only did understanding decrease in the second Module, interest decrease as well, mostly likely because the students had less understanding.

10 Conclusion

There are a few take aways from this questionnaire which we would like to address in order of importance. The first point concerns Theorema which seems to have given the students the most problems out of all the provided software. One issue which is easily handled is the complaint that mathematic cost money. A significant number of students stated that they did not want to pay and thus refrained from using the software. Providing an installation of mathematic which the students can access would easily deal with this issue. Secondly, students found the task of actually using the software daunting, many complained that the documentation was not good enough. We are not sure what this actually means, but a good guess would be that there is a gap between their ability to read documentation and the provided documentation for Theorema. We conclude that some intermediary documentation ought to be provided which is more suitable for first-year students. The most problematic comments by students concerning Theorema where “It worked but I don’t know why?”. They essentially treated it as a combinatorial problem which is to be solved brut-force. This completely eliminates its benefits as a teaching aid.

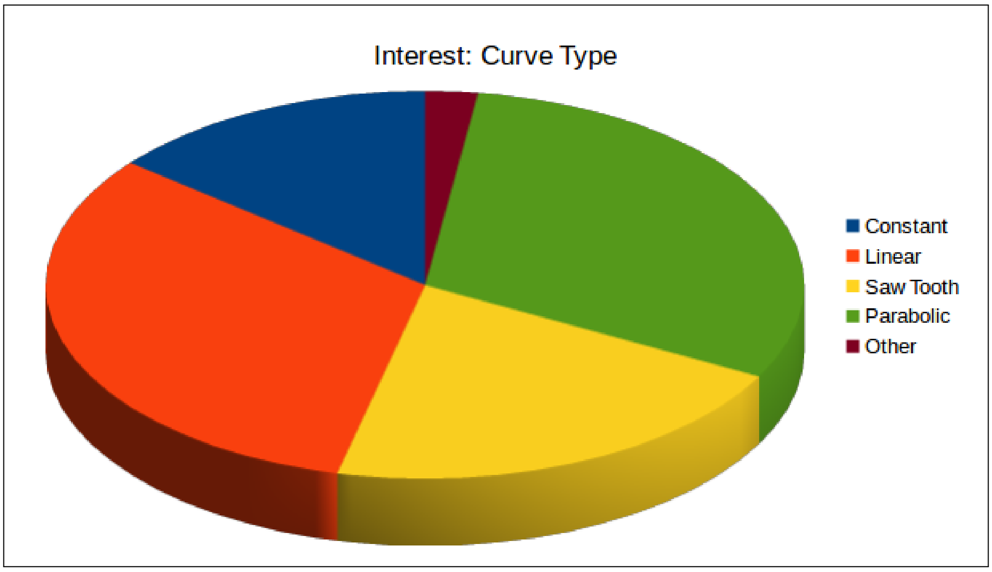


Figure 9: Types of curves drawn by students representing their interest in the course material.

Eliminating the possibility of passive use of the software ought to be a goal for the next iteration of the course. We assume that this passive use exists given that Theorema is currently introduced in its entirety and thus exercises need to be simplified. Maybe an educational fragment of the software needs to be discussed and developed.

Concerning the Second Module problem, there seems to be too much of an introduction of complex and difficult topics without a break for the students to catch up. Comparing the first Module, where there is the box game, and second Module one sees that both interest and understanding are overall higher. Though, we do have to note that the material of the first Module is technically less difficult and easier to introduce in a accommodating manor, it is not impossible to do the same for the second Module. Moreover, it is not necessarily the entire Module which is problematic, rather it is only the border between the first and the second Module which sharply divides the course. Finding a way to slowly introduce the students to the more complex first-order concepts may increase their understanding throughout the Module. The lack of a graceful transition between the two parts of the course could account for the problems noted in the above analysis.

Finally, the most interesting point for future work and iterations of the course is the importance of the box game to the students. As we noted nearly 70% of the student are in favor of it and would like to see more of it in the course. The key aspects of the game are its ability to provide students with an interactive environment where they can hone their skills. This is completely lacking in the other parts of the course. This is not because it is impossible to make such games for first-order logic or SMT but simply because it has not been done yet. RISCAL gets close to providing such an environment, but it is still more of a tool than an interactive learning device. Taking some of the aspects of RISCAL and turning them into a game would, by their own account, benefit the students. Thus, it is advised by the next iteration of the course, such gamifications of elements of first-order logic ought to be developed and introduced.