Final version published in LNCS 8095, EC-TEL 2013, pp. 410-423

Teaching Logic Through Web-based and Gamified Quizzing of Formal Arguments

Peter Øhrstrøm[1], Ulrik Sandborg-Petersen[1], Steinar Thorvaldsen[2], and Thomas Ploug[1]

[1] Department of Communication and Psychology, Aalborg University, 9000 Aalborg, Denmark E-mails: {poe,ulrikp,ploug}@hum.aau.dk
[2] Department of Education, University of Tromsø, 9037 Tromsø, Norway E-mail: steinar.thorvaldsen@uit.no

Abstract. In this paper the focus is on the construction, use, pedagogical potential, and long-term sustainability of certain web-based tools designed for teaching logic. A series of web-based tools have been implemented as a two-part system, and the tools have been tested and evaluated in a number of practical experiments. The first part of the system is a student-facing Java-Applet running in the student's browser, implemented using the Prolog programming language as embodied in a Java implementation called Prolog+CG. The second part is a teacher-oriented, server-based backend for logging the progress of students. Section 2 includes a presentation of the pedagogical and technical ideas of construction that underpin the tools which have been made so far. It is explained how the Prolog+CG tools designed for teaching logic can be developed as web applications using gamified quizzing. Section 3 of the paper contains an evaluation of the potential of log data as learning analytics offered by these tools in the context of university courses introducing basic logic and formal aspects of argumentation. In section 4 of the paper the Prolog+CG tools designed for teaching logic are evaluated quantitatively and qualitatively. The evaluation also includes a discussion of the ethical aspects concerning the logging of student data. Section 5 investigates how the development of these kinds of tools for teaching logic can be carried out in a sustainable way.

1 Introduction

Since the rise of the medieval university in Europe the study of basic logic and argumentation has been seen as an essential component of basic academia. At many modern universities this is still the view. In addition to this traditional interest in logic, the various studies of computer science, information science and communication problems have made it important to focus on logic and human reasoning. In consequence, it is commonplace at modern universities to offer basic courses in logic and argumentation. Such courses normally include the study of Aristotelian syllogistics and basic propositional arguments such as Modus Ponens and Modus Tollens (see [1], [2], and [3]). On this background it seems obvious to look for tools which can support such courses and also tools which may support distance education in logic and argumentation.

A series of web-based tools for teaching logic have been implemented as two-part systems, and the tools have been tested and evaluated in a number of practical experiments. The first part of the system architecture which has been applied is a student-facing Java-Applet running in the student's browser, implemented using the Prolog programming language as embodied in a Java implementation called Prolog+CG (see [2], [4] and [5]). The second part of the system architecture is a teacher-oriented, server-based backend for logging the progress of students and for locating the difficulties in the learning process. In [7] and [8] we have discussed these tools and their potential in logic teaching. The present paper should be seen as a continuation of these earlier studies.

In section 2 of the present paper, a discussion can be found of the paedagogical and technical ideas of construction that underpin the tools which have been made so far, as well as a brief circumscription of the problem domain of teaching logic. In section 3 of the paper it is discussed how the tools designed for teaching logic can be developed as web applications using gamified quizzing. Section 4 of the paper contains a treatment of the teaching potential of these tools in the context of university courses introducing basic logic and formal aspects of argumentation, as well as a discussion of the ethical aspects concerning the logging of student data. This evaluation is based upon practical teaching experiences and various tests of the tools which have been carried out at Aalborg University, Denmark. The last section of the paper first investigates how the development of these kinds of tools for teaching logic can be carried out in a sustainable way. Section 6 offers some concluding remarks.

2 The Construction of Prolog+CG Tools for Logic Teaching

From a modern point of view classical syllogistics may be seen as a fragment of first order predicate calculus. A classical syllogism corresponds to an implication of the following kind:

$$(p \wedge q) \supset r$$

where each of the propositions p, q, and r matches one of the following four forms:

a(X, Y)	(read: "All X are Y")
e(X, Y)	(read: "No X are Y")
i(X, Y)	(read: "Some X are Y")
o(X, Y)	(read: "Some X are not Y")

These four functors were suggested by the medieval logicians referring to the vowels in the words "affirmo" (latin for "I confirm") and "nego" (latin for "I deny"), respectively. – The classical syllogisms occur in four different figures:

$$
\begin{array}{ll}
(x(M,P) \wedge y(S,M)) \supset z(S,P) & (1^{\text{st}} \text{ figure}) \\
(x(P,M) \wedge y(S,M)) \supset z(S,P) & (2^{\text{nd}} \text{ figure}) \\
(x(M,P) \wedge y(M,S)) \supset z(S,P) & (3^{\text{rd}} \text{ figure}) \\
(x(P,M) \wedge y(M,S)) \supset z(S,P) & (4^{\text{th}} \text{ figure})
\end{array}
$$

where x, y, $z \in \{$a, e, i, o$\}$ and where M, S, P are variables corresponding to "the middle term", "the subject" and "the predicate" (of the conclusion). In this way, 256 different syllogisms can be constructed. According to classical (Aristotelian) syllogistics, however, only 24 of them are valid. The medieval logicians named the valid syllogisms according to the vowels, $\{$a, e, i, o$\}$, involved. In this way the following artificial names were constructed (see [1]):

1$^{\text{st}}$ **figure:** barbara, celarent, darii, ferio, barbarix, feraxo
2$^{\text{nd}}$ **figure:** cesare, camestres, festino, baroco, camestrop, cesarox
3$^{\text{rd}}$ **figure:** darapti, disamis, datisi, felapton, bocardo, ferison
4$^{\text{th}}$ **figure:** bramantip, camenes, dimaris, fesapo, fresison, camenop

In these names the consonants signify the logical relations between the valid syllogisms, and they also indicate which rules of inference should be used in order to obtain the syllogism in question from syllogisms which were considered to be fundamental: barbara, celarent, darii, ferio. – In fact, the system of syllogisms may in this way be seen as the first axiomatic system ever (see [1] and [3]).

The Prolog+CG system dealing with syllogisms picks some terms from a given selection, an arbitrary figure number (1–4), and an arbitrary syllogism number (1–12). The syllogism numbers 1–6 point to the valid syllogisms listed above, whereas the numbers 7–12 stand for invalid syllogisms assumed to be somewhat "tempting":

1$^{\text{st}}$ **figure:** aia, oae, iai, ieo, iii, oao
2$^{\text{nd}}$ **figure:** oae, aoe, oio, ioo, ieo, oao
3$^{\text{rd}}$ **figure:** aaa, iaa, aia, eae, oae, eie
4$^{\text{th}}$ **figure:** aaa, aie, iaa, eae, eie, aee

In this way the system operates with a database of 48 syllogistic schemes among which 24 are invalid and 24 are valid according to Aristotelian syllogistics.

The Prolog+CG tools designed for teaching syllogistics make use of this database of syllogistic schemes. The steps in the program are the following:

1. A syllogistic scheme is picked at random.
2. The argument is presented to the user in terms of natural language with clear indication of the premises and the conclusion.
3. The user gives his evaluation of the validity of the argument.
4. The system states whether or not the answer given by the user is correct – if possible also providing a short explanation.

5. Unless certain criteria have been met, such as getting 5 or 10 correct answers in a row, the user is given the opportunity to go back to step 1 and thereby get another question to answer.

In addition to this tool dealing with syllogisms, we have implemented a tool dealing with basic propositional arguments. The valid arguments included in this tool can be proved using Modus Ponens, Modus Tollens, or usual disjunctive argumentation. This means that the following arguments are tested:

$$p \supset q, p \vdash q$$
$$p \supset q, \neg q \vdash \neg p$$
$$p \vee q, \neg p \vdash q$$
$$\neg(p \wedge q), p \vdash \neg q$$

Again the propositional tool is an implementation of the steps 1–5 mentioned above using variations of the propositional arguments mentioned. In this way a database of 36 schemes of propositional arguments has been constructed. 18 of these schemes are invalid, whereas the other 18 are valid.

The Prolog+CG tools concerning syllogisms and propositional arguments have been developed as web applications using gamified quizzing. The gamified versions of the tools count the number of correct answers in a row. When the user has got a certain fixed number of correct answers in a row, he or she has won the game. In the syllogistic tool and the propositional tool the winning criteria have been 5 and 10 correct answers in a row, respectively.

The Prolog+CG programs of the tools are made available on the internet and all the student activity using the tools are logged in the teacher-oriented backend.

3 The Potential of the Prolog+CG Tools in a Teaching Context

As shown by [8] the teacher may obtain very useful information from the log-data of the student activities with the tools. In this way the teacher may also find out which types of arguments the students find easy and which they find difficult. This kind of log data seen as learning analytics can obviously be very useful for the teacher in his or her preparation and design of the lectures and work in the classroom. Two interesting examples of results of this kind will be given here.

In continuation of the results obtained in [8] a study has been carried out with the syllogism-tool. During February 2012, a number of students at Aalborg University, Denmark, have been asked in small groups (2–3 students in each) to work with a version of the syllogism tool which took them through all 48 schemes in a random order. They have been asked to do this within half an hour. 103

Table 1. Scores for syllogisms in figures 1, 2, and 3. In figure 1, no. 1-6 are valid, while no. 7-12 are invalid. In figure 2, no. 13-18 are valid, while no. 19-24 are invalid. In figure 3, no. 25-30 are valid, while no. 31-36 are invalid. In figure 4, no. 37-42 are valid, while no. 43-48 are invalid.

Syllogistic figure 1

Validity	No.	Score	No.	Score	No.	Score
Valid	1	0.82	2	0.64	3	0.92
Valid	4	0.77	5	0.49	6	0.51
Invalid	7	0.77	8	0.81	9	0.13
Invalid	10	0.39	11	0.14	12	0.19

Syllogistic figure 2

Validity	No.	Score	No.	Score	No.	Score
Valid	13	0.69	14	0.73	15	0.76
Valid	16	0.82	17	0.66	18	0.56
Invalid	19	0.61	20	0.80	21	0.28
Invalid	22	0.23	23	0.29	24	0.32

Syllogistic figure 3

Validity	No.	Score	No.	Score	No.	Score
Valid	25	0.67	26	0.85	27	0.88
Valid	28	0.61	29	0.82	30	0.81
Invalid	31	0.28	32	0.74	33	0.76
Invalid	34	0.35	35	0.78	36	0.66

Syllogistic figure 4

Validity	No.	Score	No.	Score	No.	Score
Valid	37	0.60	38	0.73	39	0.87
Valid	40	0.61	41	0.73	42	0.57
Invalid	43	0.28	44	0.87	45	0.62
Invalid	46	0.40	47	0.52	48	0.74

Table 2. The 24 Aristotelian syllogisms, divided into two groups based on whether they are drawn into question in modern logic or not. The reason to draw the 9 on the left into question is that they are invalid if terms denoting the empty set are allowed, but valid if empty set denotations are not allowed.

(a) 9 questioned syllogisms

Figure	Syllogisms
1st	barbarix, feraxo
2nd	camestrop, cesarox
3rd	darapti, felapton
4th	bramantip, fesapo, camenop

(b) 15 unquestioned syllogisms

Figure	Syllogisms
1st	barbara, celarent, darii, ferio
2nd	cesare, camestres, festino, baroco
3rd	disamis, datisi, bocardo, ferison
4th	camenes, dimaris, fresison

groups participated, and this measurement was made before the students were introduced to syllogistics. The results found are depicted in Table 1.

Obviously, this measurement is very useful for teachers who want to prepare a course in syllogistic reasoning. It would clearly be relevant for the teacher to note which syllogisms the students find particularly difficult (i.e. the syllogisms with low scores). Having made this observation, the teacher would know on which challenges it would be especially important to focus.

The results listed above give important evidence for the difference between the syllogisms whose validity is drawn into question from a modern point of view (see Table 2(a)), and those which are not questioned from a modern perspective (see Table 2(b)). It is well known that the first group of 9 syllogisms are invalid in general if terms corresponding to empty sets are accepted, but valid if no empty-set terms are allowed in syllogisms.

Table 3. This table summarizes how often the 103 student groups evaluated syllogisms in the two subgroups as valid. The difference between the evaluation of the 15 unquestioned syllogisms and the 9 questioned is strongly significant by the two sided chi square test.

	Correct reply? (Pre-test N=103 groups)	
Syllogism	Yes	No
Valid in modern syllogistics (15)	**1023**	**270**
Questioned modern syllogistics (9)	**454**	**319**
p-value	< 0.0001	

Table 4. This table summarizes counts of how often students (N=8) replied correctly to questions regarding the validity of propositional arguments. The difference between modus ponens and modus tollens is strongly significant by Fisher's exact test. (The chi-square test may not be used because some of the cell counts are low.)

	Correct reply? (N=8 students)	
	Yes	No
Modus ponens	**63**	**0**
Modus tollens	**21**	**55**
p-value	< 0.0001	

On the basis of our investigation we may discuss how the students handle the 15 syllogisms that are not drawn into question within modern logic, versus the 9 Aristotelian syllogisms which have been questioned in modern logic. We may compare the group of these 15 versus the 9 questioned syllogisms and the results are shown in Table 3. The table shows strong statistical evidence against the presumption that students will handle the two subgroups of syllogism equally well, and thus support our previous findings.

This study shows that although the students accept the Aristotelian notion of validity, seeing all 24 syllogisms as valid, they are much more certain of the validity when they are dealing with the 15 valid syllogisms than when they are dealing with the 9 syllogisms which are drawn into question by modern logicians who want to allow for the use of terms corresponding to the empty set in the syllogisms.

Another interesting result based on log-data has to do with the propositional tool. Although only a few students (N=8) participated in the test of this tool it does in fact give strong evidence in support of the expectation that students handle modus ponens significantly better than modus tollens. The results can be summarized as in Table 4.

It seems obvious that studies of this kind based on the log-data from student work with the tools can provide very useful information for teachers who have to prepare courses in logic.

Table 5. The table below summarizes the Pre-test versus Post-test results of the students score. The first table line is for the Pre-test versus Post-test of the student groups with score > 0, while the second line shows the same data for the 24 best groups. The first line supports statistical evidence against the presumption that students will handle syllogism equally well before and after the lecture. The p-values are by the two sample t-test (one sided not assuming equal variance).

	Mean Score	**SD Score**	**p-value vs. Post**
Pre (N=33)	0.455	0.261	0.040
Pre (N=24)	0.566	0.213	0.351
Post (N=24)	0.596	0.316	-

4 The Use of Gamified Quizzing in Teaching Logic

As mentioned in Section 2, the gamified version of the syllogistic tool will count the number of correct answers in a row. When the user has got 5 correct answers in a row, he or she has won the game. This gamified system was tested on March 7, 2013, at Aalborg University before and after a lecture on Aristotelian syllogistics.

The immediate experience during the test was that the students enjoyed the game and that the use of the game served as motivation for working with the theoretical question of validity which was introduced during the following lecture. It was also clear that the students who stayed to work with the game after the lecture now felt better equipped to deal with the questions of validity, and that they now had much more to talk about when facing the problems in the game.

The students were asked to run the syllogistic tool in groups of 2–4 both before and after their logic lecture (i.e. Pre- and Post-test). All these test results were logged by the system. The score was computed as:

$$\text{Score} = \frac{\text{last correct in a row}}{\text{answer count}}$$

The statistical analyses of the scoring data were performed using standard methods from descriptive statistics and statistical testing. The two sample t-test has been applied to detect increased score, and to look for significant differences between results from the Pre-test and the Post-test. The Pearson correlation coefficients were computed to look for association between time used at each task and scoring outcome. The chi-square test has been applied to detect differences in how students handle syllogisms with true or false conclusions.

4.1 Results

The Pre-test was performed with 53 groups of students. But only 33 of the groups seems to have taken the test seriously and achieved a score > 0, and hence the

Table 6. The table summarizes the Pre-test versus Post-test results of the time which students spent on each task of the exercise. The first table line is for the Pre-test versus Post-test of the student groups with score > 0, while the second line shows the same data for the 24 best groups. Both support strong statistical evidence against the presumption that student will handle syllogism equally fast before and after the lecture. The p-values are by the two sample t-test (two sided not assuming equal variance).

	Median time (s)	SD time (s)	p-value vs. Post
Pre (N=33)	20.0	16.2	0.002
Pre (N=24)	20.6	15.1	0.005
Post (N=24)	34.3	27.0	-

Table 7. The Pearson correlation coefficients between score and time spent on each task of the syllogism exercise.

	Correlation coefficient	p-value
Pre (N=33)	0.26	0.14
Pre (N=24)	0.03	0.88
Post (N=24)	0.31	0.14

20 groups with score=0 were not included in the analysis. The Post-test was performed with 24 groups of students, so apparently 29 groups went home after the lecture without completing the Post-test. This big dropout creates a bias for our analyses, and we had to compare the Post-test result versus both the 24 best results from the Pre-test as well as the 36 Pre-test results with better score than 0. Both results are shown in Table 5.

We also compared how much time the student groups spent thinking on each task of the syllogism exercise. Here we used the median time for each group, since such student groups may spend time talking about other things while they work. The results are shown in Table 6, and shows that the students spent significantly more time to think about each task after the lecture has been given.

Finally we looked at correlations between scoring and time spent on each task of the exercise. Table 7 presents the results, but shows no significant correlations, although they are positive.

4.2 Discussion of the Results

Results from Tables 5–8 shows mixed results. By learning classical syllogisms the students seem to learn to think, but not necessarily to achieve logically better results. However, Table 8 shows that the best groups base their estimation on validity of the syllogisms independent of the conclusion of the syllogism itself, while the Pre-test group of N=33 are initially more influenced in their logic by the closing conclusion of the syllogistic argument. In the last case they mainly base their evaluation of the validity of a syllogism on the final statement of the

Table 8. The table below summarizes counts of how often students in their evaluation of syllogisms agreed with a False or True statement in the conclusion of the syllogism. The first 2x2 subtable is for the Pre-test of the student groups with scores > 0, while the second 2x2 subtable shows the same data for the 24 best groups, and the third 2x2 subtable shows the results from the Post-test. The p-values are by the two sided chi-square test.

	Pre-test (N= 33 groups)		Pre-test (N= 24 best groups)		Post-test (N= 24 groups)	
	False	True	False	True	False	True
Statement in conclusion:	242	342	99	131	94	138
Student evaluation:	226	358	62	168	73	159
p-value:	0.37		0.0004		0.057	

syllogism alone, and how it is related to reality as they know it. The premises of the argument are to a large extent ignored.

4.3 Post-test interviews: Results and Discussion

Following the test, two semi-structured focus group interviews with a total of 11 students were conducted. The interviews were structured around the following questions:

1. To what extent – if any – can the relevant system be of assistance in the acquisition of competence in classical logic?
2. What could or should be changed in the system in order to further the acquisition of competence in classical logic?
3. To what extent – if any – does the gaming element influence one's use of the system, i.e. one's play?
4. To what extent – if any – does the logging of results influence one's use of the system, i.e. one's play?

The answers to these four questions may be summarized as follows.

The system is generally considered to be useful for the attempt to acquire competence in classical logic (cf. 1 above). Several of the students summarize their experience in statements like "One gets a very fundamental sense of what logical validity is", "You get this 'aha experience' about what logical validity is that you do not get in the lectures", "You develop a sense of what it means for a statement to follow from another", and "You catch a whiff of some underlying patterns and of logical validity". Interestingly, in these reports the students refer to the acquisition of skills that seem intuitive and difficult to articulate. They speak about developing "a sense" having "an experience", and catching "a whiff". Several students contrast this intuitive grasp of logical validity with the theoretical understanding acquired in the lectures: "The system would be a fine supplement to the lectures because the lectures were theoretical in the sense that they provided a number of concepts to be used for theoretical analysis while the system provides a sense of logical validity in practice.” Another student remarks “I needed my notes from the lectures to aid me in creating some sort of overview of the syllogisms. It is easy to identify the figure ‘barbara’; from there on it gets harder.”

These reports point towards an important distinction in the skills and competences that may be acquired, on the one hand through the lectures, and on the other hand through the use of the system. In lectures, the students are taught concepts and models for the systematic analysis of the logical validity of arguments. In apprehending these concepts, the students first and foremost acquire the ability to apply the concepts to the systematic analysis of the logical validity of an argument. They do not first and foremost develop their ability to determine the validity of arguments without explicit, systematic analysis. By contrast the system builds the ability to “see” or “sense” the validity/invalidity without explicit, conceptual analysis.

It seems, however, that the ability to “see” or “sense” the validity/invalidity of an argument is something one may have independently of one’s ability to provide an account of what logical validity is and how it may be systematically determined. A great many people will thus be able to determine that the argument “If it rains, the road is wet. The road is wet. Therefore it rains,” is invalid on the grounds that the road may be wet for other reasons that it raining, without being able to define logical validity. Thus one might ask whether the system is, in fact, building competence in the conceptual analysis of logical validity, or whether it only reinforces pre-theoretical ability to do so.

It is, however, clear from the reports of the students that they do in fact apply the theoretical concepts from the lectures. As mentioned above, one student stated: “I needed my notes from the lectures to aid me in creating some sort of overview of the syllogisms. It is easy to identify the figure ‘barbara’ . . . ”. The student thus explicitly identified a type of argument by means of the theoretical vocabulary, and would have wanted to analyze the arguments by means of the theoretical concepts, had they been available during the test. Two other students back this by stating that they to some extent “took notes on a piece of paper in order to determine the logical structure of the arguments.” Thus, in the mind of the students, there seems to be no clear separation of the pre-theoretical and the theoretical ability to determine the validity of arguments. By exercising one, one is exercising the other. At any rate, the analytical skills seem to be sharpened by doing rather than merely listening to a lecture.

In response to the second question of the interviews (cf. 2 above) on possible improvements of the system, the students suggest that “One should receive a notification if the validity of arguments of the same form are evaluated wrongly on several occasions,” and that “The syllogisms should also be shown in formalised and/or graphic form.” Both suggestions underscore the points made above. It is clear that the students to some extent perceive of the test as a test of the skills and competences acquired in the theoretical lectures, but also that the testing of these skills and competences could be more explicit. In terms of the specific suggestions for improvement of the system, they would clearly bridge this gap.

The students are in general reacting positively towards the gaming element in the test (cf. 3 above). Not surprisingly, the gaming element is noted by several to be motivating. It is recognised, however, that in this case – where the gaming element consists in the goal of getting 5 or 10 consecutive "right answers" – the gaming element may be especially appealing for the "stubborn, persistent and very competitive" students. The important lesson to be learned is straightforward. Different gaming elements attract different people. Future revisions and testing of the system should address this issue by more systematically incorporating and testing different gaming elements and their appeal to different types of "players". This would obviously be primarily with the end-goal of improving the learning effect of the system on the students, though other ends could also be identified.

In response to the effects of the gaming element of the system, several of the students have observed an interesting pattern in their behaviour during the test. They report that "The gaming element made me inclined to take a chance at the point of having 4 consecutive right answers. After all, it was a game . . . ." The introduction of gaming elements thus seems to play a rather significant role not only for the students' perception of the system, but also for their behaviour in using the system. A behavioural pattern of this kind clearly runs counter to the purpose of the system to motivate students to form some kind of reasoned judgement (pre-theoretical or theoretical) on the logical validity of arguments. It thus becomes of vital importance for future developments of the system to determine the extent to which the gaming elements trigger such behaviour.

The students all recognise that the logging of their answers had an effect on their behaviour even though they were anonymous (cf. 4 above). It is worth noting, however, that no one found the logging to be disturbing or caused him or her to be worried or anxious. On the contrary, all report that the logging initially had a positive effect on their motivation. As one says "It does mean something that you know someone is registering your score. It is positively motivating." Several of the students note, however, that the actual effect of the logging on their results is not straightforward. Thus, in the course of the test the logging also had adverse effects on their motivation to do their best. "At the outset one is focused on evaluating the first five arguments correctly . . . . After having realised that it will not look good, you become more careless and start guessing." The exact effect on the scores generated by the logging is consequently hard to estimate.

5 Sustainability of the Tools

Sustainability of systems and tools within education has become a major issue in recent years, as people and projects may need to transcend the organizational, personnel, and budgetary boundaries of any one institution. One primary example is the Moodle[3] Course Management System, which has grown to become a world-wide phenomenon within education, with many contributors from several institutions and organizations in disparate locations around the world. The success of Moodle may arguably be attributed in part to the decision of its orginator, Martin Dougiamas, to license the source code of the system as Open Source. The effects of Open Sourcing software has been explored in various contexts, including [9] and [10].

Licensing the source code as Open Source by itself does not guarantee sustainability. Making the code generally available is obviously also necesssary. Good documentation is essential to supporting usability and the possibility of user adoption. Finally, building a community around the systems and tools in order to maintain the software beyond the capability, interest, and budgetary limits of the originators is of paramount importance for both sustainability and user adoption.

We have taken the first three steps towards sustainability of our tools listed above, of licensing as Open Source, making the source code available, and documenting it. The tools can be downloaded from the SourceForge.Net Open Source publishing platform at the following URL:

http://syllog.sourceforge.net/

Whether the fourth step, namely building a community around the tools, will become a reality, remains to be seen.

6 Conclusion

We have in this paper presented two systems for teaching logic, namely the "syllog" system for teaching Aristotelian syllogistics, and the "proplog" system for teaching aspects of propositional logic. The two systems build on a number of common ideas, which may be formulated as imperatives. Though these imperatives are by no means normative for the broader range of educational software, they may illustrate some important aspects of the integration of software systems into traditional University-level courses on logic.

Support learning by doing: As [11] points out, higher learning outcomes can often be achieved through the student's doing an activity related to the theoretical material provided in lectures, than through lecturing along as a teaching activity. In Section 4.3, we have seen how the students reported obtaining a deeper understanding of the theoretical material through using the system than they had acquired through the lectures. Thus supporting "learning by doing" is an essential educational feature of our tools.

Support learning of analytical inquiry skills: The process of analytically finding an answer to a question of validity in the field of logic may itself be founded upon the ability to ask relevant questions, especially since proofs may be seen

[3] http://www.moodle.org/

as a sequence of propositions, the link between which consist of inference rules, validity arguments, application of axioms and theorems, etc. Thus the ability to think analytically about logical arguments involves the ability to inquire, or ask analytical questions of the problem at hand. The nature of the questions posed by our systems does not by itself ensure that the students will acquire skills in analytically oriented inquiry, but the repeated prompt to apply analytical skills, including methods of inquiry, may help instil in the student the basic potential for analytical inquiry. That is indeed a higher learning goal of many courses in logic, and our systems seem to support this goal, as indicated in Sections 3 and 4.3, as well as [7, 8].

Provide learning incentives through gamification: Not all students will be equally inclined to take all learning activities seriously enough for their own good in meeting the learning outcome objectives of the courses which they attend. The element of gamification, as we have seen in 4.3, although eliciting mixed responses from the students, does seem to indicate that at least with the student focus group in question, the element of gamification had some positive effects on their motivation to learn. Gamification, of course, has many other ramifications and potentially positive outcomes than student motivation, including a strengthening of the element in the learning experience of the student of the application of the principle of repetition, which in turn may lead to higher learning outcomes. The proverbial abundant harvest founded upon diligent sowing applies equally well in farming as in many other kinds of work, including the work of learning. With our systems, the student is encouraged, through gamification, to sow more than they otherwise might have had the stamina or interest to do.

In future research, it would be interesting to:

- Expand the number of strategies for gamification. The current strategies could be enhanced by communicating streaks, percentages, and grade assessments at the end of each game.
- Expand the number of strategies for enhancing the learning experience of the student, for example by diagramming or otherwise showing the structure of the arguments.
- Elaborate on the kind of questions which the systems are able to ask of the students. In both of the current systems, the central question is the validity of the arguments. It might, however, be advantageous for the learning outcome of the student to be able to, say, pick a valid conclusion from a list of possible conclusions (valid or invalid), assess the truth-value of a given proposition within a given model, pick the correct figure and/or syllogism name for a given valid syllogism, and so on. Many different kinds of questions could be devised, and it would be interesting to research the effect of the nature of the questions upon the learning outcomes and learning experiences of the students.

We have thus presented some systems for teaching and learning two kinds of logic, namely syllogistic logic and basic propositional logic. We have described their construction, and have evaluated their potential in teaching logic through both quantitative and qualitative methods. We have also provided some reflections on the sustainability of the tools, and have described how we have worked towards that goal by Open Sourcing the tools.

It is our hope that the systems presented in this paper can form the basis for good learning experiences of many students outside of Aalborg University in the years to come, and our Open Sourcing of the tools is a means to this end.

References

1. Parry, W.T. & Hacker, E.A.: Aristotelian Logic, State University of New York Press (1991).
2. Sandborg-Petersen, U., Schärfe, H. and Øhrstrøm, P.: Online Course in Knowledge Representation using Conceptual Graphs, Aalborg University, http://cg.huminf.aau.dk, (2005).
3. Aristotle: Prior Analytics, Translated by A.J. Jenkinson. The Internet Classics Archive, http://classics.mit.edu/Aristotle/prior.html, (1994-2000).
4. Kabbaj, A., Janta-Polczynski, M.: From PROLOG++ to PROLOG+CG: A CG object-oriented logic programming language. In Ganter, B., Mineau, G.W., (eds.): Proceedings of ICCS 2000. Volume **1867** of Lecture Notes in Artificial Intelligence (LNAI), Berlin, Springer Verlag, p. 540–54, (2000).
5. Kabbaj, A., Moulin, B., Gancet, J., Nadeau, D., Rouleau, O.: Uses, improvements, and extensions of Prolog+CG: Case studies. In Delugach, H., Stumme, G., eds.: Conceptual Structures: 9th International Conference on Conceptual Structures, Volume **2120** of Lecture Notes in Artificial Intelligence (LNAI), Berlin, Springer Verlag, p. 346–59, (2001).
6. Petersen, Ulrik, Prolog+CG: A Maintainer's Perspective. In: de Moor, Aldo, Polovina, Simon and Delugach, Harry (eds.): First Conceptual Structures Interoperability Workshop (CS-TIW 2006). Proceedings. Aalborg University Press, (2006).
7. Øhrstrøm, P., Sandborg-Petersen, U. and Ploug, T.: Syllog — A Tool for Logic Teaching. Proceedings of Artificial Intelligence Workshops 2010 (AIW 2010). Mimos Berhad, p. 42–55, (2010).
8. Øhrstrøm, P., Sandborg-Petersen, U., Thorvaldsen, S. and Ploug, T.: Classical Syllogisms in Logic Teaching. In: Lecture Notes in Artificial Intelligence (LNAI), Vol. **7735**, pp. 31–43 (2013).
9. Raymond, Eric S.: The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media (2001).
10. Dougiamas, Martin & Taylor, Peter C.: Moodle: Using Learning Communities to Create an Open Source Course Management System. Proceedings of EDMEDIA 2003, (2003).
11. Biggs, John: What the Student Does: teaching for enhanced learning. Higher Education Research & Development, Vol. **18**, No. 1., (1999).