Learning Analytics Based on Streamed Log Data from a Course in Logic

Peter Øhrstrøm[1], Steinar Thorvaldsen[2,3] and David Jakobsen [1]

1: Department of Communication and Psychology, Aalborg University, 9000 Aalborg, Denmark
{poe,davker}@ikp.aau.dk
2: Department of Education, NLA University College, 5812 Bergen, Norway
3: Department of Education, UiT- the Arctic University of Norway, 9037 Tromsø, Norway
steinar.thorvaldsen@uit.no

Abstract. This paper describes the analysis of streamed log data generated from the use of two specific e-learning tools in the context of a basic logic course taught and evaluated at Aalborg University, Denmark, during the period from September 2022 to January 2023. The students' work with the two kinds of basic logic introduced in the course, classical syllogistics and elementary propositional logic is analysed using the stored log data. Furthermore, using stored log data it is also discussed whether anything is gained if we change the order in which the two topics are presented during the course.

Keywords: Streamed log data, learning analytics, learning tools, syllogistics, propositional logic, validity of arguments.

1 Introduction

As many other universities around the world Aalborg University, Denmark, offers basic courses in elementary and classical logic and argumentation theory. In this paper we study some important aspects of the learning analytics obtained during various versions of a logic course offered to students studying communication and digital media at Aalborg University in Aalborg and in Copenhagen. The focus of this course is the study of logical validity of arguments formulated in terms of elementary propositional logic and classical syllogistics.

Two learning tools, *Syllog* and *Proplog*, were employed during the course, to generate logic exercises automatically. In this way, it is possible to make the learning experience game-like and enjoyable (see [1], [2], [3]).

The interface of the present version of the *Proplog* tool is shown in Fig. 1. The *Proplog* user can click on 'New argument' to get a new argument presented on the screen. The user should state whether the argument presented is valid or invalid (i.e., whether the conclusion follows from the premises in any possible/thinkable scenario). *Syllog* works in a similar manner and with a similar interface (Fig. 2).

One important learning goal in this context is that students should obtain the needed skills to analyse an arbitrary elementary propositional or syllogistic argument that is formulated in natural language. The students should make use of symbolic logic to evaluate the logical validity of the argument presented. For this purpose, the students may use truth tables and semantic trees to analyse propositional arguments and they may use Venn diagrams and basic inference rules to analyse syllogistic arguments.

The use of the learning tools, *Syllog* and *Proplog*, is continuous logged along with an evaluation of each response to the exercises presented by the system. *Syllog* is designed to generate exercises in classical syllogistics (see [4] and [5]), whereas *Proplog* automatically generates exercises in elementary propositional logic as it is presented in the textbook used at the course ([5]).

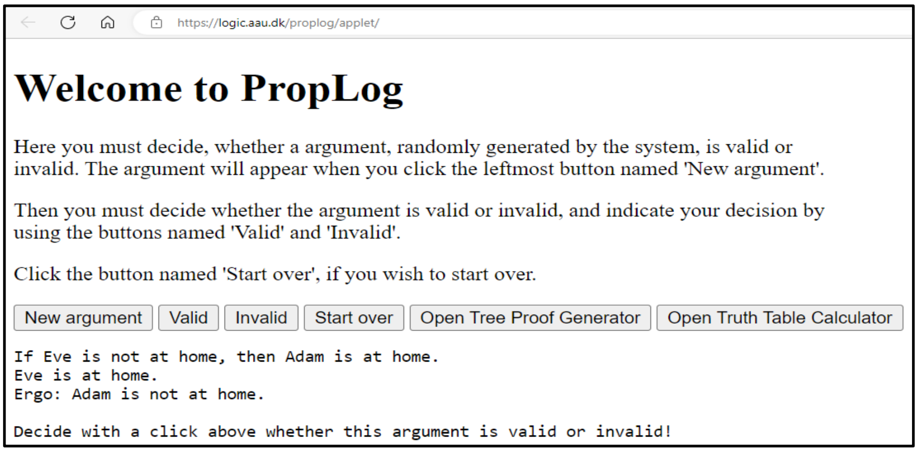

Fig. 1. Interface of the *PropLog* tool. The user is invited to decide whether the argument presented is valid or invalid.

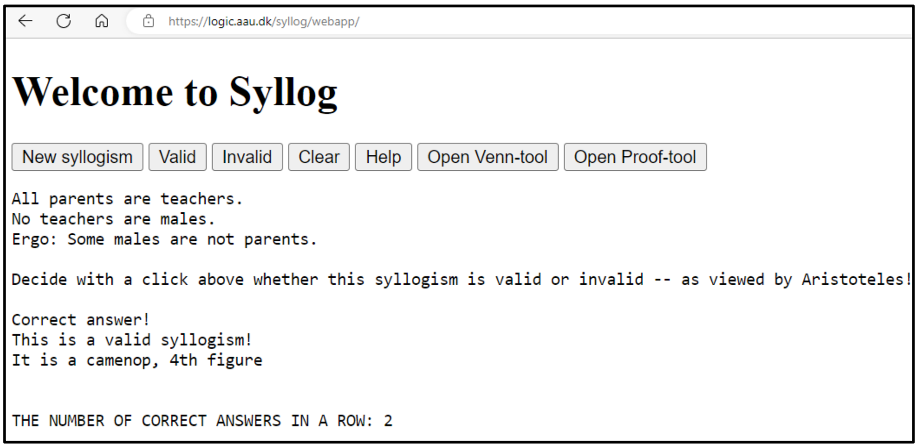

Fig. 2. Interface of the *Syllog* tool. The user is invited to decide whether the argument presented is valid or invalid. In the case of a valid syllogism, the system will give the classical medieval name of the syllogism (see [4] and [5]).

The score can be computed for any subset of exercises answered and logged in the system. The definition is straight forward:

Score = correctanswers/answercount

If a period is specified, the data from the system will include calculations of the average score during the period in question. In both *Syllog* and *Proplog*, valid and invalid arguments occur with the same frequency. This means that the score should be 50% if all users of the system answer based on random guessing.

The life-streamed log data from the student activity including the calculated score can be made immediately and continuously available to the teacher. This means that the teacher can decide to let the students know how well they are doing at any time during their interaction with the systems, *Syllog* and *Proplog.* The teacher may decide to adjust the focus and the emphasis of the teaching based on this information.

The statistical analyses of the data were performed using standard methods from descriptive statistics and statistical testing. The chi-square test is applied to detect group differences using frequency (count) data, and to look for significant differences between results. Effect sizes were estimated by Cramer's V coefficient, with conventions: 0.1=small effect, 0.3=medium effect, and 0.5=large effect [6, p.746].

A specific module has been made to support the teacher's analysis of the log data (Fig. 3). The teacher should specify which calculation should be made, i.e., when were the log data added to the database, how long time did the user use to answer, what kind of data should go into the calculation etc.

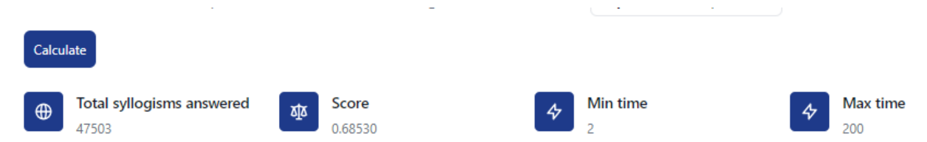

Fig. 3. A part of the interface of the teacher's module to support the analysis of the log data. Here the system is asked to calculate the score on the basis of all log data in the database from the use of the *Syllog* tool where the answer is given after more than 2 sec. and less than 200 sec. (The teacher might alternatively have chosen to concentrate on data from selected periods within 2014-2023.)

2 The temporal aspect of streamed log data

The log data include the time periods related to the exercises – both when the problems are presented on the screen and how long time the students use to answer. Fig. 4 and 5 show the time series for the submitted answers to *Syllog* and *Proplog*. All user data over the years 2014-2023 are stored in the system.

It is evident that some students answer very quickly – in fact within the first two seconds after the presentation of the exercise on the screen. It seems very likely that many of them are just guessing, although some of these very early answers may also be given on the basis of a very quickly established intuition. Other students answer much later, and it may be assumed that they have considered the problem more carefully before they give their answers.

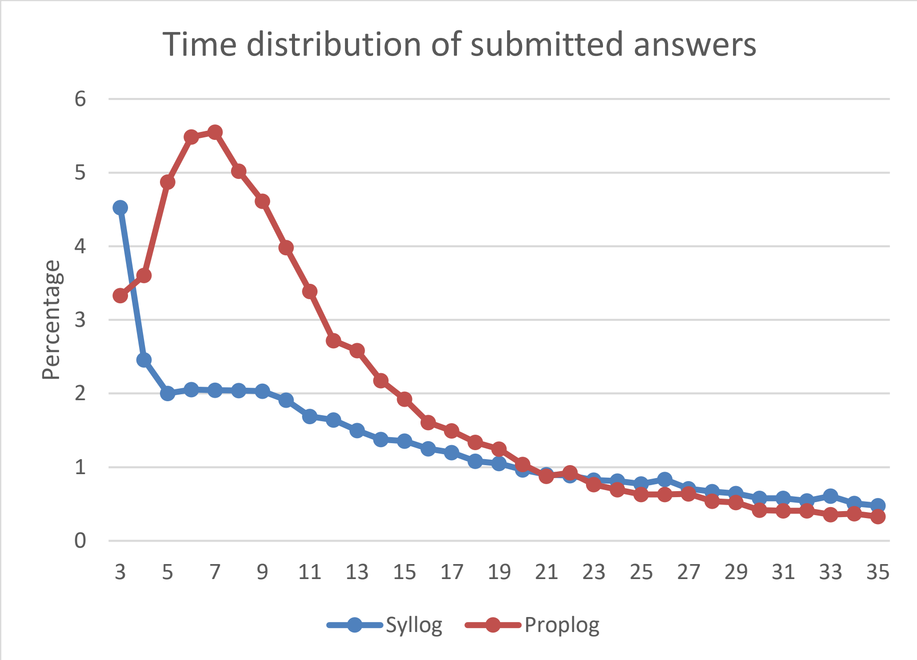

Fig. 4. The distribution from 3 to 35 seconds of the answers given by the students to *Syllog* and *Proplog*. All data during 2014-2023 are counted.

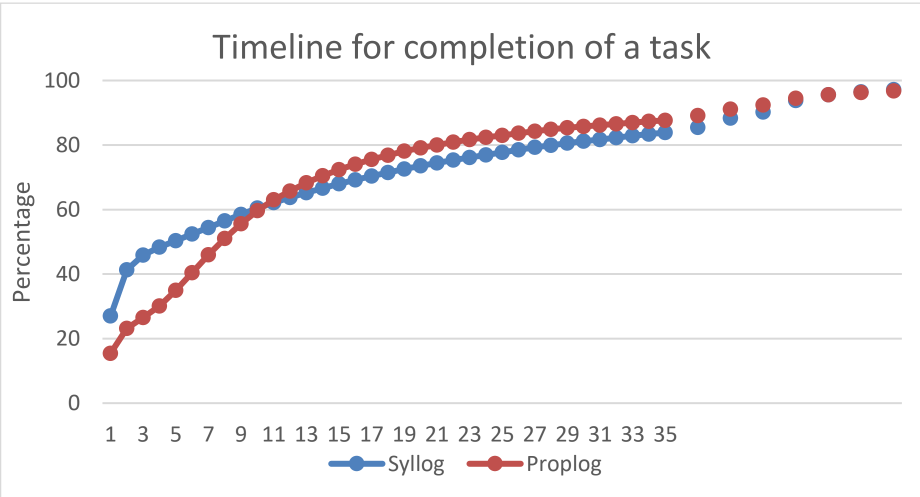

Fig. 5. Cumulative timeline for completion of a task generated by *Syllog*/*Proplog*. The results are presented for the first 35 seconds of reasoning, plus subsequent intervals up to180 seconds.

It should be noted that the number of answers given within the first two seconds is relatively higher (40,9 %) for *Syllog* than for *Proplog* (24,9 %). Apparently, the students want to use more time for thinking before they answer when they are working with *Proplog* than when they work problems provided by *Syllog*. This is consistent with the subjective experience reported from the teachers at the course (Peter Øhrstrøm and David Jakobsen) that the students in general find the exercises given by *Proplog* more difficult and demanding than the exercises generated by *Syllog*. It appears that about 70 pct. of the students have completed the task within 15 seconds - slightly more for *Proplog* (72%) than for *Syllog* (68%).

The log data also make it possible to calculate the scores for submitted answers of each interval of one second after the presentation of the problem on the screen, for both *Syllog* and *Proplog*. Thereby it may be studied how the scorings evolved as a function of time of reply by the whole group of students in 2014-2023.

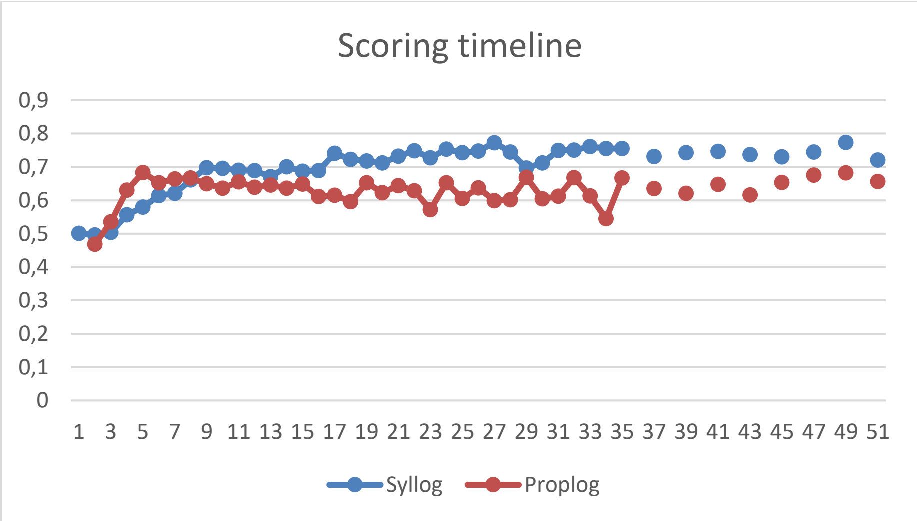

Fig. 6. The scoring timeline shows the calculations of the scores for each period of 1 second after the presentation of the problem on the screen for both *Syllog* and *Proplog*. The results are presented for the first 35 seconds of reasoning, plus some occasional subsequent intervals.

The timeline of scoring is shown in Fig. 6. We discovered an error in the program that computed the scoring for *Proplog* during the *first* second, and hence these data are omitted in the figure (the error has now been corrected in the program). Note that *Syllog* has its peak score (0.773) after 27 seconds of reasoning, while *Proplog* reach top score (0.683) after only 5 seconds. Both timelines of scores reach a new maximum after about 3 minutes, with *Syllog* at 0.774 and *Proplog* at 0.682.

When calculating the scores, it may be reasonable to ignore answers given after less than 2 seconds after the presentation of the problem on the screen in order only to exclude answers not based a rational process but rather on random guessing or some kind of intuition.

3 Measuring the learning

One straightforward use of the log data is to evaluate the numerical learning effect of the course simply by measuring the score when the course starts and when it ends. It should, of course, be remembered that there may be very important effects of the course that cannot be measured in this way. On the other hand, it is interesting and relevant to investigate how well the students are doing when they try to evaluate syllogistic and propositional arguments. To do so, we should only use data from the dates when the course begins and when it ends, i.e., the 3 days exam.

Table 1. Syllog data at the beginning and at the end of the courses in Aalborg and Copenhagen. Answers given after less than 2 seconds are ignored in the calculations of the scores. The course in Aalborg started 8-10 September 2022 and ended 8-11 November 2022. The course in Copenhagen started 16-18 October 2022 and ended 3-6 January 2023. There were 81 students in Aalborg, and 57 students in Copenhagen.

	At start		**At exam**	
	Total number	Score	Total number	
Aalborg	1369	0.59898	1950	0.62923
Copenhagen	1472	0.60394	438	0.65525

We notice that the student group in Copenhagen had a higher start level in syllogistic logics, and they also had the largest progress of score (0.051 in Copenhagen versus 0.030 in Aalborg). However, some of the numbers in Table 1 are relatively small, and it is not certain how the students used the *Syllog* tool during the exam. In order to obtain more reliable results, the conditions should be more controlled. On the other hand, in principle this shows how the learning effect may be measured.

4 The effect of the order of the presentations during the course

An interesting question is whether the order of the presentations during the course has consequences for the outcome of the learning during the course. In other words, will the scores in propositional logic raise if the topic is presented as the first and not as the second, as it has normally been done when this course has been given? To answer this question, the order in which two types of logic were presented was changed in the spring of 2021 (F21), i.e., in the period February 2021– June 2021. The results from (F21) are compared with the results from (E22) i.e., September 2022 – January 2023). Table 2 shows the results.

We observe that by presenting propositional logic first, the results increase significantly. However, the effect size determined by the Cramer’s coefficient is very small (0.024). Both *Proplog* and *Syllog* favour somewhat higher scores when being presented as the first topic in the lectures, with an increase of 0.024 and 0.009, respectively.

Table 2. The average scores concerning syllogistics and propositional logic and during the logic courses in Aalborg and Copenhagen in the period February 2021– June 2021 (F21) compared with the results from the period September 2022 – January 2023 (E22). In (F21) Propositional logic was presented as the first type of logic, whereas syllogistics was presented as the first type of logic in (E22). ***: p-value <0.001

	Score	**Correct answered?**		**p-value**	**Effect-size**
		No	Yes		
F21 Proplog	.60198	5919	8952	0.0002***	0.024
E22 Proplog	.57800	4109	5628		
E22 Syllog	.66189	2641	5170	0.29	0.009
F21 Syllog	.65368	2684	5066		

We may also observe that the *Syllog* scores are significantly higher than *Proplog* (p-value < .001 by the chi-square test) both in 2021 and 2022, still with small effect sizes of 0.05 and 0.09, respectively. This difference was in fact contrary to what had been expected by the teachers, who had anticipated that the *Proplog* scores would have been higher than the *Syllog* scores. – An observation of this type certainly demonstrates the usefulness of this this kind of learning analytics.

5 Types of reasoning

One interesting use of streamed log data in the context the logic course in question focusses on some specific types on reasoning which have had particular attention during the history of logic. Using the log data, it may simply measure to which extent humans can evaluate these interesting types of reasoning correctly.

In this section we focus some of the classical syllogistic arguments which traditionally have been understood as interesting in particular. In doing so we refer to the classical four syllogistic figures as they were presented in Medieval logic (see [4], [5], [7]).

Immanuel Kant (1724-1804) notes that the first figure of the syllogisms is interesting. It has the form M-P, S-M, where M stands for the middle term, S for subject and P for predicate, is often the way in which we take a rule, note that something is a case of that rule and deduce a result, such as:

Rule: All human beings have intrinsic value
Case: Some merely physical object is a human being
Result: Some merely physical object has intrinsic value

This way of reasoning, as described in Prior [7, p. 112], is sometimes called *dictum de omni et nullum,* or ‘a principle about all or none’. The major premise must be universal (affirmative or negative), and it could be assumed that this way of reasoning is more intuitively familiar to students who has not studied syllogisms before, than those kind of reasonings that enters into disputation with some postulated rule. Charles S. Peirce noticed that reasoning in the second and third figure, could be seen as a continuation of the same kind of reasoning about a rule, some case and then a result.

Thus, in the second figure, which has the form P-M, S-M, we deduce a controversial result from the Rule, by stating some case that does not have that property in question and deduce a result from this. Again, as in figure 1, the major premise must be universal, but the conclusion must be negative.

Rule: All human beings have intrinsic value
Denial of Result: No merely physical object has intrinsic value
Denial of Case: No merely physical object is a human being

This way of reasoning, called *dictum de diverso,* springs a surprising result of the rule upon us, and we might have weaker intuitions about whether or not we are being taken in. There must be something wrong with this way of thinking, a student might reason who has not learned that this syllogism is *Camestres*, of figure 2. What we have in the above argument is an affirmation of the rule, a denial of the result from which we conclude a denial of the case. Finally, in the figure 3, which has the form M-P, M-S, we deduce an exception to the rule, thus denying its universality, hence the name *dictum de excepto.* From the above premises we can make the following valid *Ferison* argument:

Denial of Result: No merely physical object has intrinsic value
Case: Some merely physical object is a human being
Denial of Rule: Some human being has intrinsic value

This means that a *Darii* of figure 1 in an interesting way can be transformed into a *Camestres,* of figure 2 as well as to a *Ferison* of figure 3. Some will find this transition somewhat surprising. The classical observation is at least that we in this case go from the obvious to the less obvious. This may be documented using our log data from the use of *Syllog.*

Looking merely at the course in Aalborg from the 8. September – 10. September 2022, where we have a total of 1103 responses, we have the following results for figure 1:

Figure 1	Total number	Score
Barbara	24	0.83333
Barbarix	22	0.59091
Celarent	18	0.72222
Darii	24	0.83333
Ferio	28	0.78571
Feraxo	32	0.75000

Given that the average score of that period for the fall 2022 in Aalborg where 0.66189 from September 2022 to January 2023, it is only *Barbarix* which falls under. This relative low value is to be theoretically expected as *Barbarix* is one the 9 of the 24 classically valid syllogisms that are sometimes questioned (cf. [4] and [5]). Similarly, the high values of *Barbara* and *Darii* should also be expected on the basis of classical syllogistics.

We have the following results from the second figure:

Figure 2	Total number	Score
Baroco	30	0.73333
Camestres	15	0.40000
Camestrop	16	0.50000
Cesarox	23	0.78261
Cesare	25	0.48000
Festino	23	0.82509

It is rather significant that the students have a higher spread on how they deal with figure 2 where the major premise, like in figure 1 is universal and the conclusion is negative. It is also significant that the students fall under 0.5 when evaluating *Camestres* and *Cesare.* Only *Baroco* and *Festino* get better results than the total average from that period. We have the following results from the third figure:

Figure 3	Total number	Score
Bocardo	17	0.88235
Darapti	22	0.63636
Disamis	27	0.85185
Datisi	27	0.85185
Felapton	25	0.64000
Ferison	25	0.68000

The students score slightly less than average on *Darapti* and *Felapton*, which both belong to the 9 that may be questioned from a modern perspective. Among the others *Ferison* gets the lowest score. Furthermore, it is quite interesting, and call for more research, that the students are more capable of grasping the intuitions behind arguments of the first and the third figure than the second. Whether this is due to subject matter or form is however difficult to deduce from these results. Are students better at deducing a result from a universal rule and some case, or rejecting the universality of a rule, from a case, because a perceived deduction of that rule would be absurd, then they are at using a rule to reject a case, by denying the result? Or are these results merely a question of the used arbitrary terms in the assignments? It could be worthwhile to investigate these matters further with more data and a case meant to stir once ethical or religious intuitions on some matter.

6 Conclusion and perspectives

The analytic use of streamed log data in the context of a logic course it very promising. As we have seen in sections 2 and 3, this kind of learning analytic may offer the teachers some very important information regarding how well and how quickly the topics and problems studied during the course are in fact understood by the students.

Furthermore, as it was shown in section 4 learning analytics based on the log data may give rise to an argument in favour of a change in the order of the topics presented during the course. However, the observed effects were considered too small to justify a change of the teaching strategy. In fact, the *genetic method* in education holds that the progression of teaching a subject should follow the same approach as the development of knowledge itself. The mathematician Otto Toeplitz [8] cleverly discussed the idea about 100 years ago. The historical genetic method aims to lead students from basic to more difficult knowledge, in much the same manner as mankind has progressed in the history. A modern use of the genetic method is also described in recent studies [9]. Since syllogistics (at least Aristotle's original version of it) was developed prior to propositional logic, syllogistics should be presented earlier than propositional logic if we follow the genetic method.

As discussed in section 5 it should also emphasized that the availability of the log data makes it possible to compare philosophically and historically important types of reasoning with the concrete understanding of these types of reasoning in a modern context.

Finally, it should be mentioned that there are important questions that should be explored regarding the future use of streamed log data. One of them would be whether it would improve the learning outcome if the average score of the student group (or perhaps his or her own score) over, for instance, the last 10 minutes, is continuously shown on the screen to the student while he or she is working with *Syllog* and *Proplog*. Will it be an interesting and stimulating facility since it is game-like? Or will it create a stressful and negative learning environment? These and other possible applications of live-streamed log data could be studied and evaluated in future research projects. However, it should also be remembered that there are more qualitative aspects of logic learning that should be studied in other ways, e.g., using interviews with users.

References

1. Øhrstrøm, P., Sandborg-Petersen, U., Thorvaldsen, S., Ploug, T. Teaching Logic through Web-Based and Gamified Quizzing of Formal Arguments. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds) Scaling up Learning for Sustained Impact. EC-TEL 2013. LNCS, vol 8095, pp. 410-423. Springer, Heidelberg (2013).
2. Øhrstrøm, P., Sandborg-Petersen, U., Thorvaldsen, S., Ploug, T.: Teaching syllogistics through gamification and interactive proofs. In: Conole, G., Klobučar, T., Rensing, C., Konert, J., Lavoué, E. (eds.) Design for Teaching and Learning in a Networked World. EC-TEL 2015. LNCS, vol. 9307, pp. 609–612, Springer, Heidelberg (2015).
3. Øhrstrøm, P., Thorvaldsen, S., Sandborg-Petersen, U., Ploug, T. Teaching Propositional and Syllogistic Logic Using E-learning Tools. In: Rønningsbakk, L., Wu, TT., Sandnes, F., Huang, YM. (eds) Innovative Technologies and Learning. ICITL. LNCS, vol 11937, pp. 854-864, Springer, Heidelberg (2019).
4. Parry, W., Hacker, E.: Aristotelian Logic. State University of New York Press, (1991)
5. Øhrstrøm, P., Ploug, T., Jakobsen, D.: Logisk set. 2nd ed. Metaphysica, Aalborg, (2019)
6. Cohen, L., Manion, L. & Morrison, K. Research methods in education, 8th edn. Routledge, (2018).
7. Prior, A.N.: Formal logic, Oxford University Press, (1955).
8. Toeplitz, O. The calculus: A genetic approach. Chicago: University of Chicago Press, (1963).
9. Mosvold. R. Genesis Principles in Mathematics Education. Notodden: Telemarksforsking Notodden, (2002).