How People Learn II

Project Submission

Analysis of debugging strategies in MOOCs

Szabolcs Balási[a], Robin Demesmaeker[a], Johannes Wüthrich[a]

[a]*Ecole Polytechnique Fédérale de Lausanne (EPFL), Route Cantonale, Lausanne 1015, Switzerland*

Abstract

Massive Open Online Courses (MOOCs) are a huge innovation in education, bringing knowledge to a wide variety of people. A lot of studies explored student behavior in MOOCs, but some aspects still remain untouched. In this paper, we use clickstream data to examine learning behavior during exercise solving. Using a clustering algorithm, we detect six "debugging" strategies. We show among others that people do not stick to one strategy and suggest that the context in which you register for a MOOC, influences your behavior. The presented results can help MOOC makers to create an adaptive, individually centered, learning environment.

Keywords: Massive Open Online Courses; clickstream analysis; learning styles; clustering; behavioral patterns;

1. Introduction

Massive Open Online Courses (MOOCs) define a new, Internet based, educational model that has become point of an ever-increasing interest over the last few years. They deliver high quality courses from world leading universities to millions of students each year without demographic, economic or geographical constraints. Moreover, they open up new horizons in educational research by providing unprecedented amounts of data about students' interaction with the course material.

In this paper, we extend the existing literature about student behavior in MOOCs by focusing on what people do in between the first and last submission of an assignment. Since one of the special characteristics of MOOCs is that people have the possibility to submit multiple times an assignment while getting immediate feedback, it is interesting to study their learning or "debugging" strategies when solving exercises. MOOCs provide a vast amount of resources like videos and forums. But do students take advantage of these features? Do they really try to fill the gaps in their knowledge based on the feedback they receive? What strategies do they use for it? Those are the questions we address in this paper by analyzing the clickstream data collected by the MOOC platform. Identifying the differences in students' learning styles would allow to provide personalized guidance for learners in order to further improve the efficiency of MOOC learning.

After starting off with a comprehensive review of the existing literature, we show our methodology for analyzing the data. We present and discuss the results of our research along with directions of continued work.

2. Literature review

In this paper we focus on the analysis of xMOOCs, which can be seen as a scaled-up version of the traditional university educational model (Siemens, 2013). Such MOOCs are offered by several providers, such as Coursera, edX or Udacity and count for the vast majority of the available online curriculum.

Most MOOCs are composed of short video lectures and exercises. Often, additional lecture notes and discussion forums are also at students' disposal. However, the way learners use the available resources can greatly depend on their personal preferences, often referred to as learning styles (Chang et al., 2015). Nevertheless, these preferences are dynamic and learners might adopt new strategies depending on the context (Kennedy, 2002).

Although the content of the courses is generally presented in chronological order on a weekly basis, students are free to go back to previous content at any time, even during assessments. In fact, most students tend to take advantage of this freedom and define their own learning paths (Guo & Reinecke, 2014) instead of following the linear structure of the course. As navigation through the course content is a fundamental component of the overall learning experience, understanding students' behaviour and preferences would allow the design of a learning-style adaptive MOOC system (Onah & Sinclair, 2015) which could improve the learning gain in web-based educational settings.

Recent MOOC research showed that different learning behavior patterns can be identified both from large-scale aggregated statistics (Anderson et al., 2014) as well as from fine grained clickstream data (Wen & Rosé, 2014). Some of these behaviours can be intuitively paired with the elements of different learning style models, such as heavy forum usage with social, repeated exercise solving with experimental and video watching with passive learning styles (Kolb, 1976). Other studies in this field focused more on some particular aspects of students' interaction with the course content, such as dropout prediction based on the navigation in video lectures (Sinha et al., 2014) or the connection between forum interaction behaviour and the final grades obtained (Wang et al. 2015). However, no previous research examined in finer detail these behavioral patterns together during exercise solving.

Beside video lectures, weekly assessments and final examinations are also essential parts of most MOOCs. When students submit an exercise, they usually receive immediate feedback, which can have an important impact on their learning gain in general (Hattie & Timperley, 2007). Moreover, in contrast with traditional university settings, students are allowed to attempt a formative assessment multiple times, letting them build a solid basis before moving on to further topics and to learn for mastery (Do et al., 2013).

The aim of this paper is to extend the available knowledge of behavioral pattern analysis in MOOCs with a focus on assessments. More precisely, students' activity between the first and last submission of a given problem will be examined in detail in order to identify and compare different debugging and learning strategies applied by students in MOOCs.

3. Methods

3.1. Clickstream data

In this paper we analyzed the clickstream data of the course 'Introduction to Object Oriented Programming in JAVA' which was given as part of a course at EPFL at the end of 2015. The MOOC was available on the Coursera platform, and was composed of seven weeks of lectures and 11 programming assignments out of which 9 were graded.

We started by cleaning the dataset, removing obvious artefacts introduced by the logging system. It appeared from the data that people quite often went from a good grade to a zero grade and back to a good grade again in a short time. We argued that these events were most probably due to typos in the code and the resulting compilation errors. In our opinion this does not represent a learning strategy and in order not to bias the results.

As the focus of our research was on examining learning behavior during exercise solving, we only kept the data of students who solved at least one programming assignment (653 out of 995 registered students). For these students we then took all the clickstream data in-between the respective first and last submission of an assignment, i.e. 30942 problem submissions, 58278 forum events and 53595 video events were taken into account for all 11 programming assignments together. The clickstream events were then further categorized into the following groups:

- Problem submissions: single and repeated
- Video events: download, load, play, pause and seek
- Forum: launch, load, view, post and comment

3.2. Strategy clustering

Now that we had all the clickstream events of interest for our research question, we grouped them for each unique (problem, user) tuple into chronological event strings. These sequences were then transformed into Markov transition matrices with each element representing the probability of going from one kind of event to another. In total 3634 such matrices were constructed.

Since the goal was to discover different strategies, these Markov matrices were fed into a K-means clustering algorithm. A crucial parameter for this method is the number of clusters. We observed that going higher than 6 clusters gave characteristic Markov matrices that were barely distinguishable. Therefore, 6 was chosen as the number of clusters in the following.

4. Results and discussion

Several analyses connected to the "debugging" strategies were carried out and are presented here. First we start by elaborating the results of the clustering itself and by taking a closer look at the identified behaviors. We then go on to explore how students' behavior changes along the course. We look at the obtained grade in function of the used strategy and last we analyze how students taking the course as part of their EPFL curriculum behave compared to others. This gives us some insight on the influence of the learning setting on students' behavior.

Due to the size of our dataset, the result of any chi-square test of independence tends to signal significant results. This is a well known fact, since the chi-square value linearly depends on the sample size. In order to counter this, we always also calculate the Cramér's V.

4.1. Six different strategies

The resulting typical behavior for each cluster can be seen in Fig. 1. Each sub-figure represents the average transition matrix for each cluster and the number of occurrences is also shown. We identify the following behaviors:

- Single Submit: The student made only one submission for the problem, without trying to improve his score or fix any errors.
- Multi Submit: The student made multiple submissions to solve the problem without consulting any of the available course supports.
- Light Forum: The student solved the problem by mainly resubmitting while consulting the forum from time to time.
- Heavy Forum: The student spent a lot of time in the forum and made relatively fewer submissions.
- Passive Video: The student watched videos while solving an exercise.
- Active Video: The student watched videos while solving an exercise and actively sought within those videos.

An intuitive expectation would be that a high percentage of people use the resources available on the MOOC to solve the programming assignments. Therefore, the distribution of strategies used is quite surprising. Over 64.5% of all problems were solved without consulting the course supports. Around 28.5% of the problems were solved mainly consulting the forum and only around 7% of the problems were solved using videos as the primary support. One would also expect that a higher percentage of people re-watch videos in order to overcome issues during the problem solving. This seems clearly not to be the case, the preferred support appears to be visiting the forum, probably to look for other students who already encountered the same problem. Even students who re-watch the videos during problem solving tend to visit the forum. Finally, it should be noted that only a very small fraction of students (around 3% of the problems solved) actively searches for solutions in the videos via seeking. So we can see that indeed there seem to be different learning styles present. One could attribute the Multi Submit strategy to an experimental learning style. On the other hand, forum usage links to a social learning style while video watching to a passive one.

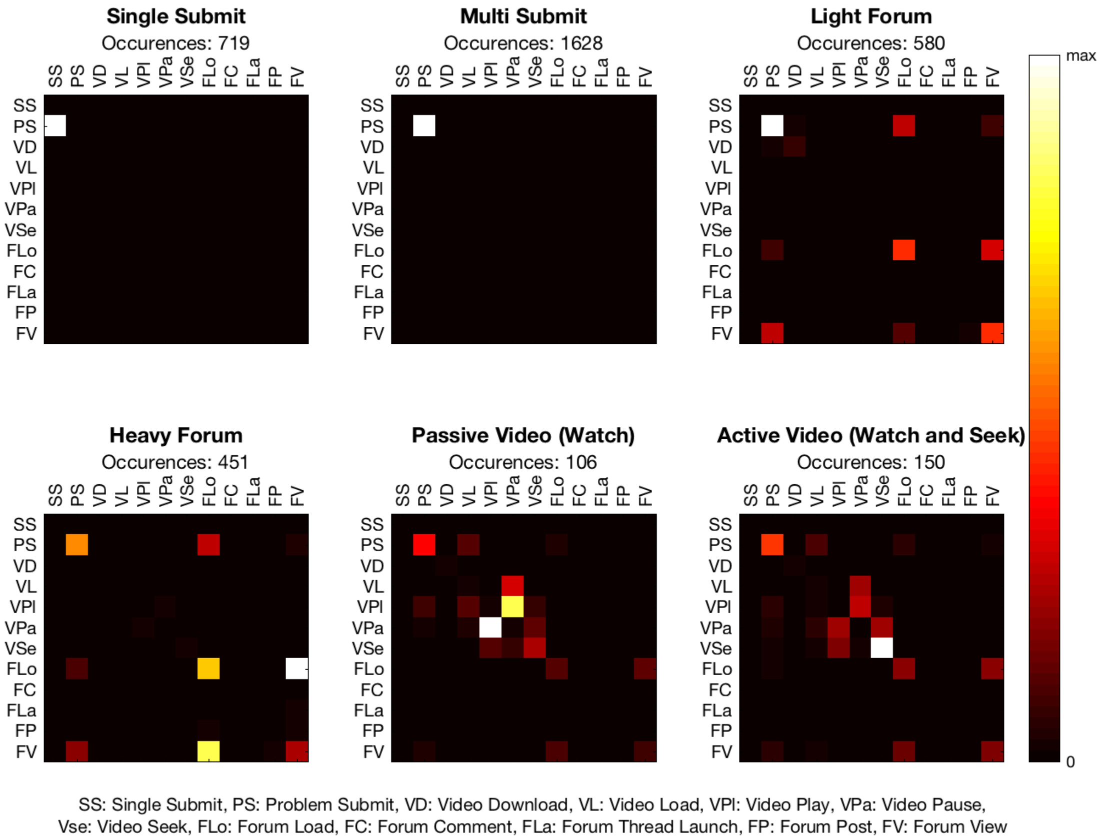

Fig. 1. discovered “debugging” strategies.

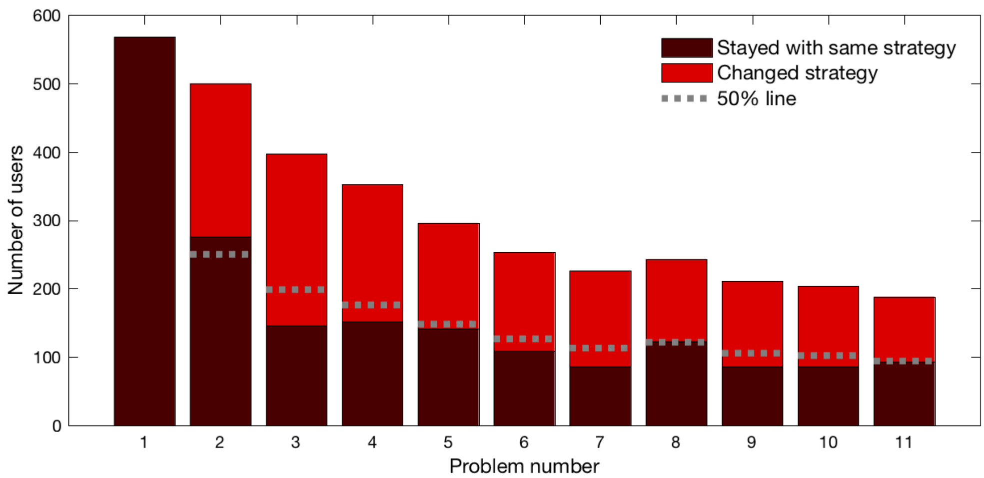

Fig. 2. people using the same strategy versus people switching strategy for 2 subsequent problems.

4.2. People do not stick to one strategy

A first important question to be answered is whether people have a preferred strategy or not? The answer appears to be no. Fig. 2 shows for each assignment the number of people that submitted the exercise, split into two groups: those who stayed within the same strategy compared to the previous assignment and those who switched. It is clear that on average more than 50% of the students change their strategy. Furthermore, it is worth mentioning that almost 100% of the students switched strategy at least once during the course.

4.3. Strategies versus grades

One of the main questions in the context of learning styles and learning behaviors is if there is an influence of the applied learning strategy on the outcome - the results obtained by the student. In order to evaluate this question, the following two analyses were done. First we looked at the final grade of each problem for all students in function of the applied strategy. Second, we explored the average grade improvement per submission of each student in function of the strategy. In both cases we divided the nominal variable (the final grade resp. the average grade improvement) into five equal-sized ordinal buckets on the interval [0, 1].

The distribution of the final grade of a problem in function of the employed strategy can be seen in Fig. 3(a) (the tabular data can be found in the Annex). A chi-square test of independence gives us the following values: (chi-square = 183.3, $p < 0.001$). The Cramér's V value is 0.1, indicating low association.

Visually inspecting the grade distributions in Fig. 3(a), one can notice that the main difference stems from the distribution of students who employ a Single Submit strategy. Excluding Single Submit data points we get the following value for the chi-square test: (chi-square = 46.4, $p < 0.001$) and for Cramér's V: 0.05 indicating very low to no association.

The distribution of the grade improvement per submission in function of the strategy employed can be found in Fig. 3(b) (the numerical data is included in the Annex). A chi-square test of independence gives us the following values: (chi-square = 2756.4, $p < 0.001$) and Cramér's V is 0.39 indicating medium to strong association.

Visually we observe that the Single Submit distribution stands out and so we again test for independence ignoring the Single Submit data points. This results in the following values for the chi-square test: (chi-square = 419.8, $p < 0.001$) and for Cramér's V: 0.17 indicating low to medium association.

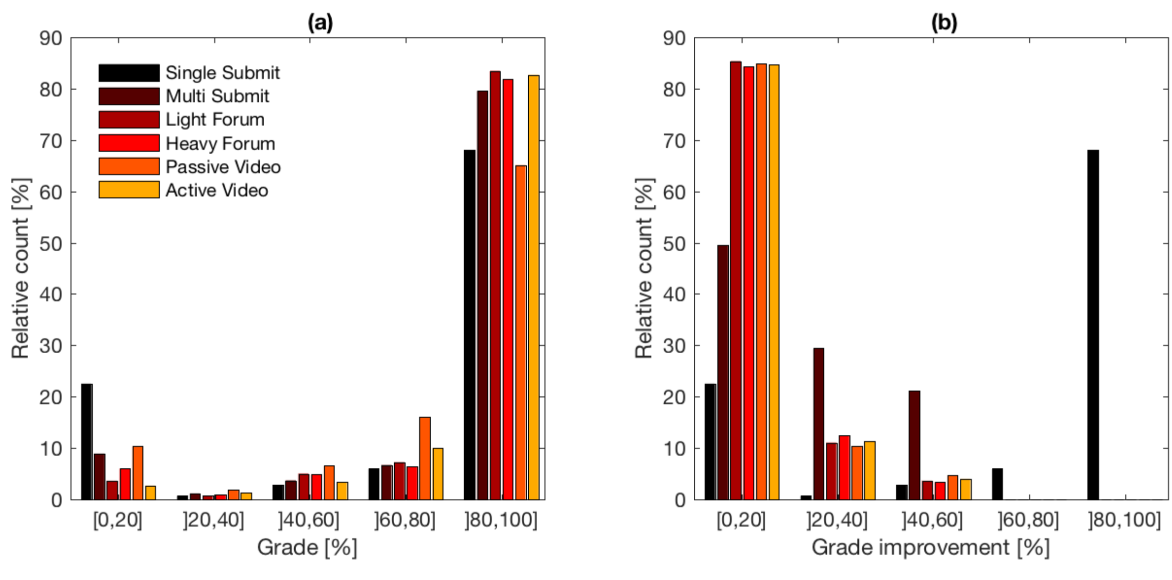

Fig. 3. (a) grades versus strategy; (b) grade improvement versus strategy.

Looking at the final grade obtained for each problem, we can see that there is a big trend of people getting more than 80% correct. This fact can be attributed to the nature of MOOC problems, as it is possible to resubmit multiple times. Therefore, the obtained grade is not necessarily a measure of understanding but also of persistence and effort put into solving the problem.

It is therefore difficult to evaluate the efficacy of different strategies using only the final grade as a measure. The results obtained support this assumption. We can see that even though the chi-square test rejects the zero hypothesis with a high margin, the association between strategy used and grade obtained is small. The only real difference in the grade distribution comes from the group of students using the Single Submit strategy. Looking closer at the grade distribution of this group we also see that the main difference comes from the fact that many people get a grade close to zero. This is also to be expected as this represents the people who try an exercise once and then give up immediately. So again the grade is less a result of the strategy used, but more of the persistence shown by the student.

In order to see if there is some influence of the strategy used on the efficiency of problem solving we analyzed the average grade improvement per submission. The thought behind this, given that all strategies (excluding Single Submit) lead to a similar grade, is that some lead to students making less submissions to get the same result.

The results of the chi-square tests and the Cramér's V values seem to indicate such a dependency of the average grade improvement on the strategy used, both including and excluding the Single Submit strategy. The Single Submit strategy can hardly be seen as a learning behavior and is more a result of border cases (directly getting 100% or getting close to 0% and abandoning). So when excluding the Single Submit Data we still get a low to medium association according to the Cramér's V value. Looking at the distributions in Fig. 3(b) we see a quite surprising reason for this result: while the strategies using videos and the forum yield nearly identical results, there is a clear difference compared to the Multi Submit strategy. This seems to indicate that not using the course material while solving problems leads to better results. One explanation is that people not using the course material to solve the problems, are likely to be users having some previous knowledge. Especially in the domain of programming, previous knowledge (for example of a different programming language) is a huge advantage when faced with a new language. Additionally, those students might still consult some help, for example external web sites (e.g. *stackoverflow.com*). Especially in the context of programming assignments in a popular language like Java there is a vast amount of information already available on the Internet.

So again we face a situation where the grade (more precisely the grade improvement) is not a product of the strategy employed but of secondary effects (e.g. previous knowledge, motivation to solve the problems). In general, we can say that we did not find conclusive prove that the employment of a certain learning / debugging strategy yields better or worse results in the context of the MOOC analyzed.

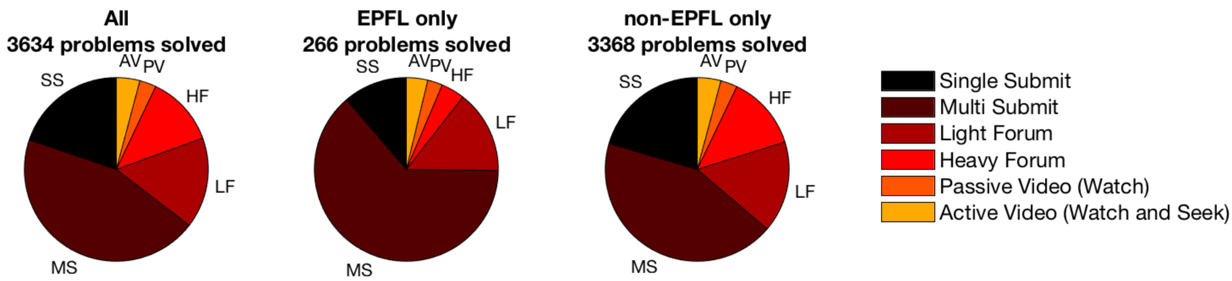

Fig. 4. distribution of strategies used by EPFL and non-EPFL students.

4.4. Link between learning setting and strategies

In order to have a basic impression of the influence of the learning setting on the behavior of students' problem solving, we analyzed the difference of strategies used by students taking the MOOC as part of their university curriculum and the strategies used by all the other students.

Fig. 4 presents the data in a visual form. It seems to be clear that the general behavior of EPFL and non-EPFL students is different. EPFL students tend to favor a Multi Submit strategy. The significance as calculated with a chi-square test seems to be given: (chi-square = 51.7, $p < 0.001$). Calculating Cramér's V we get a value of 0.11, indicating a low association.

Even though the association between the strategy used and the fact whether or not a student is from EPFL seems to be low, there is still a significant difference between the behavior of the two groups. The most obvious difference seen, the percentage of people using a Multi Submit strategy, is also the most expected. People using this strategy do not use the course support while solving problems, but just resubmit multiple times without any other apparent action in-between. This behavior clearly coincides with the learning setting the students at EPFL are exposed to. In contrast to other participants, EPFL students have this MOOC as a part of their curriculum, integrated with ex-cathedra lectures and exercise sessions. During the latter students have teaching assistants (TAs) at their assistance to help them with the programming assignments. The results suggest that students prefer to rather ask a question to a TA than to look for a solution using the material available online.

5. Conclusion

In this paper, we showed how different "debugging" strategies in MOOCs can be observed by clustering students' clickstreams between first and last assignment submission. We discussed the six discovered strategies and explored some important questions related to learning behavior. In summary, the results showed that people do not always use the same strategy and that there is only a very week correlation between the used strategy and the grade. Finally, we explored how the setting in which you sign up for a MOOC might influence your behavior while solving problems.

As the results obtained in this paper are based on the data of only one MOOC, it would certainly be interesting to run analysis proposed in this work on a wider dataset. We think that in general analyzing the behavior of students during problem solving might help to improve the learning environment of MOOCs.

Acknowledgements

We would like to thank Roland Tormey (CAPE, EPFL), Patrick Jermann (CEDE, EPFL) and Francisco Pinto (CEDE, EPFL) whose support all along this research project has been highly appreciated.

Appendix A. Data Tables

Table 1. Grade distribution in function of strategy

	[0, 0.20]	]0.20, 0.40]	]0.40 – 0.60]	]0.60, 0.80]	]0.80, 1.00]	Total
Single Submit	162	5	20	43	489	719
Multi Submit	145	19	60	109	1295	1628
Light Forum	21	4	29	42	484	580
Heavy Forum	27	4	22	29	369	451
Passive Video	11	2	7	17	69	106
Active Video	4	2	5	15	124	150
Total	370	36	143	255	2830	3634

Table 2. Grade improvement distribution in function of strategy

	[0, 0.20]	]0.20, 0.40]	]0.40 – 0.60]	]0.60, 0.80]	]0.80, 1.00]	Total
Single Submit	162	5	20	43	489	719
Multi Submit	805	479	344	0	0	1628
Light Forum	495	64	21	0	0	580
Heavy Forum	380	56	15	0	0	451
Passive Video	90	11	5	0	0	106
Active Video	127	17	6	0	0	150
Total	2059	632	411	43	489	3634

Table 3. Strategies used by EPFL and non-EPFL students

	EPFL	non-EPFL	Total
Single Submit	30	689	719
Multi Submit	169	1459	1628
Light Forum	39	541	580
Heavy Forum	11	440	451
Passive Video	7	99	106
Active Video	10	140	150
Total	266	3368	3634

References

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2014). Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*. ACM, pp. 687-698.

Chang, R.I., Hung, Y.H. and Lin, C.F. (2015). Survey of learning experiences and influence of learning style preferences on user intentions regarding MOOCs. *British Journal of Educational Technology*, *46*(3), pp. 528-541.

Do C.B., Chen Z., Brandman R. and Koller D. (2013). Self-driven mastery in massive open online courses. *MOOCs Forum 1,* pp. 14-16.

Guo, P. J., and Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM, pp. 21-30.

Hattie, J., and Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77(1),* pp. 81-112.

Kennedy, P. (2002). Learning cultures and learning styles: Myth-understandings about adult (Hong Kong) Chinese learners. *International journal of lifelong education*, *21*(5), pp. 430-445.

Kolb, D. (1976). *Learning style inventory.* Boston: McBer and Company.

Onah, D. F. O., and Sinclair, J. (2015). Massive open online courses: an adaptive learning framework. *INTED2015 Proceedings*, pp. 1258-1266.

Siemens, G. (2013). Massive open online courses: Innovation in education. *Open educational resources: Innovation, research and practice*, p. 5.

Sinha, T., Jermann, P., Li, N., and Dillenbourg, P. (2014). Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions. In *2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*.

Wang, X., Yang, D., Wen, M., Koedinger, K., and Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *Proceedings of the 8th International Conference on Educational Data Mining.*

Wen, M., and Rosé, C. P. (2014). Identifying latent study habits by mining learner behavior patterns in massive open online courses. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, pp. 1983-1986.

WORDCOUNT: 3238 without references and annex