This is a pre-copyedited version of a contribution published in 2020 21th International Conference on Interactive Collaborative Learning published by Springer. The definitive authenticated version is available online via https://doi.org/10.1007/978-3-030-11932-4_22.

Study effort and student success: a MOOC case study

Jeanette Samuelsen[1(✉)] and Mohammad Khalil[1]

[1] University of Bergen, Norway
{jeanette.samuelsen, mohammad.khalil}@uib.no

Abstract. Learning was once defined as the function of efforts spent in relation to efforts needed [3]. Provided that effort is closely linked to time, previous research has found a positive relationship between student effort over time and student success, both in university education and Massive Open Online Courses (MOOCs). With the complex environment of tracing and identifying relevant data of student learning processes in MOOCs, this study employs learning analytics to examine this relationship for MITx 6.00x, an introductory programming and computer science MOOC hosted on the edX MOOC platform. A population sample from the MOOC (*N* = 32,621) was examined using logistic regression, controlling for variables that may also influence the outcome. Conversely, the outcome of this research study suggests that there is a curvilinear relationship between effort over time and student success, meaning those who exert effort for the longest amount of time in the MOOC actually have a lower probability of obtaining a certificate than others who exert effort over somewhat less time. Finally, research implications are discussed.

Keywords: Massive Open Online Course (MOOC); efforts; learning analytics; logistic regression; total time; study success

1 Introduction

Academic success through course completion and/or degree attainment is a requirement for many types of jobs. In addition, obtaining a degree is linked to different long-term benefits [15]. Effort is about the "exertion put forth during a task" [20, p. 13]. It may seem intuitive that effort toward academic activities over time will influence student success. However, it is also common for people to view success more as a result of innate abilities, whereby high effort is viewed as a sign of low ability [6].

The relationship between student effort over time and student success in higher education has been examined in the literature. Bowman et al. [2] conducted multiple regression analyses to examine the relationship between perseverance of effort and grade point average (GPA) among undergraduates at Bowling Green State University and the University of Wisconsin at La Crosse. They found a significant positive relationship between the two variables. Strayhorn [23] examined the role of grit on college grades among a subpopulation at a research university, using hierarchical regression techniques. The results indicated that grit was positively related to college grades among the subpopulation. Grit is actually a composite measure encompassing both the behavioral part perseverance and the cognitive aspect passion. However, evidence suggests that behavioral measures are more important than cognitive regarding academic outcomes [23]. Cross [5] examined the role of grit on current student GPA for a group of non-traditional doctoral students in a private university. The results showed a small, but significant relationship between grit and GPA.

A recent trend has been for institutions to host their own Massive Open Online Courses (MOOCs). "A MOOC is an online course with the option of free and open registration, a publicly-shared curriculum, and open-ended outcomes" [17]. Thus, MOOCs allow people with varying levels of time commitment and education levels to participate. Learning analytics studies in MOOCs have found that MOOC participation rate varies to a larger degree than for traditional higher education [4]. The dropout rate is also much higher in MOOCs than in traditional university courses [12, 13].

There has been some research on the relationship between effort over time and success in MOOCs. Researchers at Google examined the "Mapping with Google" MOOC [25], where participants could earn a certificate only through completing a final project. This study found that completing other course activities was positively correlated with earning a certificate. Likewise, researchers analyzing data from the HarvardX course "CB22x: The Ancient Greek Hero", found that taking many actions in the MOOC was positively related to earning a certificate [21]. While these studies offer valuable insights, individual MOOCs often have large differences in instructional conditions, student characteristics and collected data [9]. Thereby, statistical models for MOOCs, and their implications, may not be generalized to MOOCs which occur in different contexts.

This study expands upon previous research by examining the relationship between student effort over time and student success in an introductory level MOOC for programming and computer science, "MITx 6.00x". The insights derived from studies such as this one could benefit not only the research community, but also individuals in pursuit of academic success and its potential positive outcomes, and course instructors/providers who want to help and motivate students to realize their potential.

1.1 Research Question

Based on previous findings in other contexts from traditional university courses and MOOCs [2, 25], this research study will examine the relationship between student effort over time and student success in a MOOC case study. The assumption being that more effort over time will continually lead to higher likelihoods of success (i.e. increasing effort will result in better outcomes) [8]. Following, the research of the current study will tackle the following research question:

What is the type of relationship between students invested efforts over time and their success in MOOCs?

2 Method

2.1 MITx 6.00x

MITx 6.00x was an introductory course to computer science and programming offered by MIT from 2012 to 2013, hosted on the edX MOOC platform (http://edx.org). The course included content such as video lessons, homework questions, assignments, three exams, and a forum. Among resources used were 148 videos, 209 problems, and 31 web pages. So-called chapters gave an overarching structure for a majority of the MOOC content (forums were organized outside of this structure). 14 chapters were released over 15 weeks. Earning a certificate, at no cost, was based on getting a final grade of at least 55%. The final grade depended on the performance of exercises, homework, and exams [22].

2.2 Dataset and participants

This study uses the HarvardX-MITx Person-Course dataset [19], which is a freely available dataset. The dataset contains de-identified, aggregated information for each individual that participated in MOOCs from Harvard and MIT on the edX MOOC platform; excluding individuals that could not be reliably de-identified. The dataset was loaded into R from a comma-separated values (CSV) file. Data from 16 MOOC offerings were included in the dataset, but through filtering on the course ID of our studied MOOC, the data used in this study contained only observations from the MITx 6.00x MOOC. Information for the variables in the dataset was either derived from the usage of the MOOC (through log data), or self-reported by the participants in an online questionnaire provided upon registration.

In total, 84,511 students originally registered for the course, but due to removal of observations, records from 32,621 participants were analyzed using logistic regression.

2.3 Data pre-processing

Filtering and corrections were carried out before the dataset was analyzed. Outliers of extreme amount of interactions were removed. Some other observations were filtered out due to inconsistencies. Unrealistically high values and observations with blank values on some of the control variables (indicating participant unwillingness to answer a specific question) were also removed. Some variables were transformed before being used: The dataset variable named year of birth was transformed to age. The dataset variable start time was transformed from date format to day format to ease the analysis process. Avoiding discrimination was also considered, for instance individual countries and regions were recategorized into continents.

2.4 Measures

This study employed logistic regression, a method that has been widely used in the learning analytics and educational data mining fields [1]. The dependent variable for the regression, representing student success, was certified (0: no; 1: yes). As stated before, obtaining a certificate implied getting a final grade of 55% or more. The independent variable, representing effort over time, was number of days active (at least one click in a given day). Both variables were derived from log data. Some control variables were included in the regression: age, gender, level of education, continent and start day. Continuous variables were the following: number of days active, age and start day. The other variables were categorical: gender, level of education and continent.

Measure for effort over time

Deciding to use the number of days active as the most efficient measure metric for effort over time from the log files is based on two reasonings: (a) supporting literature such as the research studies by Khalil and Ebner [13] and Kloft et al [14]. And (b) exploratory examination of correlations, knowledge of the problem domain, and examining descriptive statistics for the variables as the following.

There were also four other nominees that could represent effort over time, in addition to the number of days active: number of events (interactions with the MOOC), number of video play events, number of chapters accessed, and number of forum posts. Executing Pearson correlation for the five variables (n = 35,115 due to removal of observations missing values) made it clear that the number of forum posts had very little correlation with the rest of the variables. Its largest correlation was with number of days active (r = .26), and its smallest correlation was with the number of video play events (r = .16) (see Table 1). Based on the exploratory examination of the correlations, it did not seem that the number of forum posts was a good indicator of effort over time. In retrospect, it seemed that the number of forum posts might be a measure of social behavior [16].

Table 1. Correlation matrix for candidate variables effort over time variables

	No. events	**No. days active**	**No. video play events**	**No. chap. accessed**	**No. forum p.**
No. events	1				
No. days active	.87***	1			
No. video play events	.74***	.66***	1		
No. chapters accessed	.81***	.88***	.61***	1	
No. forum posts	.25***	.26***	.16***	.22***	1

* $p < .05$; ** $p < .01$; *** $p < .001$

With regards to the number of chapters accessed, it is reasonable to assume that this variable would have a high value for individuals who exerted much effort over time in the MOOC. However, because chapters served as an overarching structure for materials such as exercises, homework and exams, it would also be high for students who earned a certificate despite exerting low effort in the MOOC (for instance students with prerequisite knowledge who obtained a certificate only through graded exercises, without learning anything new).

The next variable to consider as a measure for effort over time was the number of video play events. Descriptive statistics for this variable suggested video play habits of individuals varied greatly. The maximum number of play events was 8,632, and 583 records contained more than 1,000 video play events (n = 36,289 after deletion of observations missing values). A thousand plays of the 148 available videos would imply that each video had been seen almost seven times. Here, a more plausible explanation is that these numbers are a result of pressing pause and play, rewinding the videos, etc. It seemed that this measurement encompassed different types of interactions; hence, it was excluded.

The number of events variable did not seem like a good choice either, as it included video play events and forum posts. In the end, we decided that the most reasonable measurement for effort over time was the number of days active. In fact, Kloft et al. [14] has also identified the number of active days as the most important metric to predict dropout. Although this surrogate measure admittedly is more of a quantitative measure than a qualitative (it is impossible to assess the exact level of effort exerted over the number of days active), it does indicate students' commitment to the MOOC.

Control variables

As previously mentioned, this study controls for differences in start day, age, continent, and level of education ("Less than Secondary", "Secondary", "Bachelor's", "Master's", "Doctorate"). Start day is an integer signifying which day a student registered for the MOOC, relative to days since registration was made possible (start day 1 would mean that the student registered on the first possible day). This variable could influence certification, as MITx 6.00x was a highly structured course [22]. Considerable research has also suggested that demographic factors may influence student success [23]. In addition, it seems reasonable to assume that level of education could have an impact on individual students' likelihood of obtaining a certificate. Start day, age and continent variables were either transformed or recategorized from original dataset variable to measure used in this study (see Section 2.3). Of the control variables, start day and parts of the continent data were based on information derived from the log files. Data for the other measures were self-reported.

2.5 Statistical analysis

Correlations between candidate measures for student effort over time were assessed initially, to help find one or more appropriate measures to include in the study. Descriptive statistics were utilized to better understand the characteristics of the participants, with regards to the independent, dependent and control variables. Logistic regression was run to examine if there was an association between the independent variable student effort over time (represented by number of days the student was active) and the dependent variable student success (obtaining a certificate for successfully completing the MOOC). We also controlled for additional variables by adding them as additional independent variables to the logistic regression.

Categorical variables were dummy coded for use in the logistic regression. The reference category for gender was female, the reference category for level of education was less than secondary, and the reference for continent was North America.

Measures were taken to assess how well the data met the assumptions of logistic regression, and to make necessary corrections where possible. To check if the continuous predictors were related to the log of the outcome variable, interaction terms for the continuous predictors were tentatively added to the logistic regression and assessed for their significance score after running the regression. This was based on the recommendation by Field et al. [7, p. 344-345]. Actually, it was found that the interaction term for the number of days active was significant, indicating that the assumption had been violated for this variable. To address this violation, the squared term for the number of days active was added to the logistic regression. Running the logistic regression with the square of the number of days active yielded a significantly better model fit than the model that did not include the squared term ($X^2 = 226.43$, $p < .001$, df = 1), suggesting a curvilinear relationship between the predictor and outcome variable. Curvilinear relationships are a quite common occurrence within the social sciences [18].

To check the assumption of absence of multicollinearity for logistic regression, variance inflation factor (VIF) and tolerance values were assessed for the independent variables entered into the logistic regression (number of days active, number of days active squared, and the control variables). Because the number of days active squared had been entered into the regression, it was natural to assume that this term would be highly correlated with the number of days active variable (i.e., itself not squared). However, in the instance of curvilinear relationships between predictor and outcome, multicollinearity can still be accepted [18]. Both number of days active and number of days active squared had a VIF of 9. The VIF value is quite high, nearing the value of 10 which is often especially problematic. The mean VIF was 3.6, which may indicate that multicollinearity can lead to some bias in the model [7].

Due to focusing on a sole MOOC, it seemed reasonable to assume that the data were not related (i.e. errors are independently distributed). Observations were assessed for their DFBETA value to identify influential cases for the logistic regression. No observations were identified as having a substantial influence (DFBETA value above 1). Observations missing values for variables used in the logistic regression were removed before analysis.

2.6 Limitations

Some limitations apply to this study. The full population participating in the MOOC could not be analyzed, for a variety of reasons. When downloading the dataset some observations had been removed, for anonymity reasons. After filtering out data and removal of observations with missing values, we analyzed a sample of only 32,621 students with logistic regression. This may have biased estimates and inflated standard errors since data were not missing completely at random. The use of a squared term in the logistic regression, for number of days active, resulted in a better model fit but also introduced a degree of multicollinearity, which may have somewhat biased the model [18]. Number of days active was admittedly a quite coarse-grained measure for student effort over time, and if available we might have found that one or more other measures (e.g. combined through factor analysis) were better options. Using certification as a measure of student success is limited by the fact that some individuals may not see test scores and certification as a necessity. One more concern is that some information for the control variables were self-reported, which may have introduced bias [2].

3 Results

3.1 Descriptive statistics

Table 2 presents the means and standard deviations for the continuous variables included in the study, and the percentages for each level of the categorical variables.

From the table, we see that only five percent of the participants earned a certificate (mean 0.05, SD 0.22). This percentage amounted to 1601 of the 32,621 participants. The dropout ratio is as high as reported in many studies like [12, 13]. For certificate learners, the mean number of days active were 66.31.

Table 2. Descriptive statistics for independent (including control) variables and the dependent variable in the study

n = 32,621	**M/SD**	**Percent(%)**
Number of days active	9.15/16.43	
Start day	65.29/34.69	
Certification	0.05/0.22	
Age (years)	26.14/7.51	
	Gender?	
Male		86
Female		14
	Level of education?	
Less than secondary education		3
Secondary education		34
Bachelor's degree		43
Master's degree		19
Doctoral degree		1
	Continent?	
Africa		10
Asia		28
Europe		24
North America		29
Oceania		1
South America		7

3.2 Logistic regression analysis

A logistic regression was used to predict the relationship between student effort over time and student success, controlling for some demographic variables, level of education, and start time. A likelihood ratio test of the full model against a null model was statistically significant, indicating that predictors can reliably separate between students who obtain a certification and students who do not ($X^2 = 9607.44$, $p < .001$, df = 14) (see Table 3). The model correctly classified 97,3% of the observations.

To interpret how the number of days active was related to obtaining a certificate, the total logit for the different possible values for number of days active (1-138) was calculated, holding the other continuous variables at mean (start day=65.29, age=26.14), and the categorical variables at reference group (gender, female; level of education, less than secondary; continent, North America). The individual total logits were then transformed into probabilities, and the calculated probabilities for obtaining a certificate were plotted for the individual number of days (all these operations were coded manually in R, due to limitations in the margins library for R). As seen in Fig. 1, the results suggest there is initially almost a linear positive relationship between the number of days active and the probability of obtaining a certificate, but around day hundred the previously almost linear relationship seems to hit a plateau, and somewhat later the positive relationship actually weakens (the relationship is curvilinear). Here, the model implies that the participants with the most number of days active were actually less likely to obtain a certificate than participants who were active for somewhat less number of days (since the probability of earning a certificate is based on total logits it is dependent on the values of the control variables).

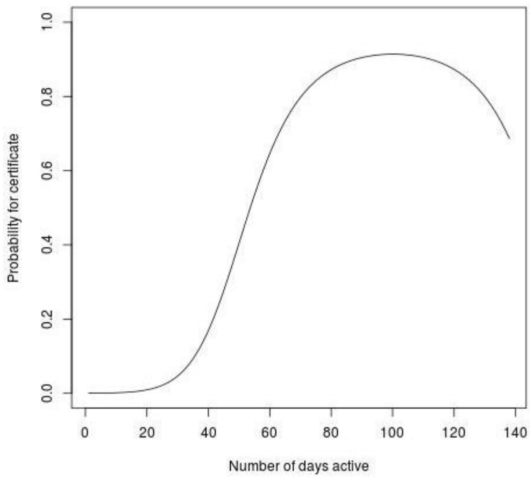

Fig. 1. Calculated probability for obtaining a certificate, related to the number of days active (continuous variables set at means, categorical variables set at reference group)

For the control variables, we see in Table 3 that age is significant, at 0.95 odds ratio, implying that as age increases by one unit, the odds of obtaining a certificate decrease by five percent, when holding the other variables constant. Start day is significant with the odds ratio of 1.01, implying that the odds of obtaining a certificate increases by one percent for each successive day of a participant registering for the MOOC. For gender, being male has a less positive relationship with earning a certificate than being female (odds ratio 0.62). For instance, if we examine the gender differences in the total probability for certification with 100 days active (thus generally a high probability for earning a certificate, as seen in Fig. 1), holding the other continuous variables at their mean, and the other categorical variables at their reference, the model suggests there is a 4.5% higher probability for earning a certificate for females than for males (91.4% versus 86.9%). For level of education, we see that having less than secondary education (the reference group) is associated with much lower odds for obtaining a certificate than the other education levels. For continents, Africa is significant (odds ratio 0.42), implying lower odds of obtaining a certificate for people from this continent than for those from North America (the reference group). On the other hand, being from Asia is associated with higher odds for obtaining a certificate than being from North America (odds ratio 1.58). Results from the other continents suggest that being from those respective continents are associated with higher odds for obtaining a certificate than being from North America; however, these results are not significant.

Table 3. Logistic regression analysis of the relationship between student effort over time and student success in the studied MOOC

		95% CI for odds ratio		
	B (SE)	Lower	Odds Ratio	Upper
No. days active	0.22*** (0.01)	1.23	1.25	1.26
(No. days active)^2	0.00*** (0.00)	1.00	1.00	1.00
Start day	0.01*** (0.00)	1.00	1.01	1.01
Age	-0.05*** (0.01)	0.93	0.95	0.97
Male	-0.47*** (0.14)	0.48	0.62	0.82
Secondary ed.	1.18*** (0.31)	1.80	3.27	5.99
Bachelor's deg.	1.12*** (0.31)	1.67	3.07	5.68
Master's degree	1.27*** (0.33)	1.88	3.56	6.78
Doctoral degree	1.58** (0.49)	1.85	4.84	12.64
Africa	-0.86** (0.28)	0.24	0.42	0.72
Asia	0.45** (0.15)	1.18	1.58	2.11
Europe	0.12 (0.13)	0.88	1.13	1.44
Oceania	0.47 (0.47)	0.62	1.60	3.96
South America	0.10 (0.23)	0.70	1.10	1.72
Constant	-7.81*** (0.40)	0.00	0.00	0.00

Note. R^2 = 0.75 (Hosmer-Lemeshow), 0.26 (Cox-Snell), .79 (Nagelkerke).
Model X^2 (14) = 9607.44, $p < .001$.
* $p < .05$; ** $p < .01$; *** $p < .001$

4 Discussion and conclusion

To answer our research question "*What is the type of relationship between students invested efforts over time and their success in MOOCs?",* this research study findings suggest that there is initially almost a linear positive relationship between student effort over time and student success in a MOOC, as was expected. However, interestingly enough, the study indicates that the previously almost linear relationship plateaus over time, and eventually the positive relationship actually weakens. This suggests that those who exerted effort over the longest amount of time actually had a lower probability of obtaining a certificate than others who exerted effort over somewhat less time (as exemplified in Fig. 1). Thus, the study suggests the relationship between effort over time and success is actually curvilinear. The curvilinear relationship was quite surprising given the initial assumption presented in Section 1.1 that more effort over time would continually lead to higher likelihoods of success.

Although the previous explanation clarifies the correlation between certification ratio and student efforts over time, there are other variables that affect this correlation. Among the included control variables, we saw that increasing age somewhat negatively influences the odds for student success, increasing start day slightly positively influences the odds for student success, that females have a higher probability of earning a certificate than males, and that those with less than secondary education had much lower odds for obtaining a certificate than their counterparts with more education. Being from Asia is associated with higher odds for obtaining a certificate than being from North America, while the opposite is true for Africa.

Returning to the finding that the relationship between student effort over time and student success is suggested to be curvilinear, one possible reason for this may be related to the concept of achievement goals. Achievement goals are about why someone shows achievement [20, p. 255]. The two types of achievement goals are presentation goals and mastery goals. People who set presentation goals are generally more concerned with proving to others that they are competent and have high ability. We can envision that earning a certificate is one such way of proving competence. On the other hand, people who set mastery goals are more concerned with self-improvement, developing competence, and overcoming challenges through effort. Thus, it may be that some of the people who exert the most effort over the longest periods of time are mastery oriented, i.e., they work hard and master challenging tasks, but may not even be interested in earning a certificate (proving their ability). It has been found that people who set mastery goals are often more internally than externally motivated [20]. Another possible explanation for the finding of the curvilinear relationship between student effort over time and student success could be that some students may just need more time to learn and develop competence in introductory programming and computer science than others, for instance, based on their prerequisite knowledge.

While researchers have found a positive relationship between effort over time and student success, both for university education and MOOCs [2, 25], a curvilinear relationship between student effort over time and student success is to best of our knowledge a unique conclusion in this research study. Given that MOOCs have a higher dropout rate than more traditional university education, and that there are more pressures (e.g. economic pressures) related to completing a university education than a MOOC, it does not necessarily follow that we would have the same findings when researching more traditional university courses. Since it is difficult to generalize statistical models and the implications of this study to other MOOCs, it is unclear if this finding would apply to MOOCs occurring in other contexts as well. However, since this study accounts for student characteristics, we may expect that the results could, at least to a larger extent, be generalized to similar types of MOOCs (introductory programming and computer science MOOCs), provided that instructional conditions and data collection procedures are closely matched.

Control variables also had a significant impact on the outcome variable. The finding that increasing age had a negative relationship with odds for obtaining a certificate could be due to younger people generally being more used to information technology than the older adults. The finding that increasing start day has a slightly positive effect on the odds of obtaining a certificate could actually be because the start day is set to the first day when it was possible to register, instead of for instance the day of the MOOC launch. The finding that gender influences the probability of success is consistent with other findings from more traditional education [24]. In the studied MOOC, we saw that there was a much higher amount of men than women enrolled. This could be due to the subject matters of computer science and programming, which tend to have a higher amount of males than females, both in education and in the workforce [11]. It is perhaps unsurprising that having less than secondary education could influence the odds for success in a MOOC, compared to having more education. The finding that being from the continent of Africa suggests lower odds for student success in comparison to being from North America may for instance be influenced by difficulties with the English language. However, the continent measure is an aggregate, meaning that there may be large differences among individual countries. For Asian learners, the finding that they have higher odds for student success than learners from North America could for instance be influenced by the fact that some of the Asian countries are among the best on all three facets of the PISA performance rankings [10], suggesting quite excellent education systems for some of the countries. Still, it should again be stressed that continent is an aggregate measure.

References

1. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Learning Analytics: From Research to Practice, pp. 61–75. Springer, New York (2014)
2. Bowman, N.A., Hill, P.L., Denson, N., Bronkema, R.: Keep on truckin'or stay the course? Exploring grit dimensions as differential predictors of educational achievement, satisfaction, and intentions. Social Psychological and Personality Science **6**(6), 639-645 (2015)
3. Carroll, J.: A model of school learning. Teachers College Record **64**(8), 723-733 (1963)
4. Clow, D.: MOOCs and the funnel of participation. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge, pp. 185-189 (2013)
5. Cross, T.M.: The Gritty: Grit and Non-Traditional Doctoral Student Success. Journal of Educators Online **11**(3), 1-30 (2014)
6. Dweck, C.S.: The secret to raising smart kids. Scientific American Mind **18**(6), 36-43 (2007)

7. Field, A.P., Miles, J., Field, Z.: Discovering statistics using R. London, Sage (2012)
8. Firmin, R., Schiorring, E., Whitmer, J., Willett, T., Collins, E.D., Sujitparapitaya, S.: Case study: Using MOOCs for conventional college coursework. Distance Education **35**(2), 178-201 (2014)
9. Gašević, D., Dawson, S., Rogers, T., Gasevic, D.: Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education **28**, 68-84 (2016)
10. Gurria, A.: PISA 2015 results in focus (2018). http://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf
11. Hill, C., Corbett, C., St. Rose, A., American Association of University Women: Why So Few? Women in Science, Technology, Engineering, and Mathematics. AAUW (2010)
12. Khalil, H., Ebner, M.: MOOCs completion rates and possible methods to improve retention-a literature review. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 1305–1313 (2014)
13. Khalil, M., Ebner, M.: What massive open online course (MOOC) stakeholders can learn from learning analytics? In: Spector, M., Lockee, B., Childress, M. (eds.) Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy, pp. 1–30. Springer, Heildelberg (2016). http://dx.doi.org/10.1007/978-3-319-17727-4_3-1
14. Kloft, M., Stiehler, F., Zheng, Z., Pinkwart, N.: Predicting MOOC dropout over weeks using machine learning methods. In: Proceedings of the EMNLP 2014 Workshop on Modeling Large Scale Social Interaction in MOOCs, pp. 60-65 (2014)
15. Kuh, G.D., Cruce, T.M., Shoup, R., Kinzie, J., Gonyea, R.M.: Unmasking the effects of student engagement on first-year college grades and persistence. The journal of higher education **79**(5), 540-563 (2008)
16. Lackner, E., Khalil, M., Ebner, M.: How to foster forum discussions within MOOCs: A case study. International Journal of Academic Research in Education **2**(2) (2016)
17. McAuley, A., Stewart, B., Siemens, G., Cormier, D.: The MOOC model for digital practice (2010)
18. Mehmetoglu, M., Jakobsen, T.G.: Applied Statistics Using Stata: A Guide for the Social Sciences. Sage (2016)
19. MITx, HarvardX: HarvardX-MITx Person-Course Academic Year 2013 De-Identified dataset, version 2.0. Harvard Dataverse (2014). https://doi.org/10.7910/DVN/26147
20. Reeve, J.: Understanding motivation and emotion, 6th edn. John Wiley & Sons (2014)
21. Reich, J., Emanuel, J., Nesterko, S., Seaton, D.T., Mullaney, T., Waldo, J., ..., Ho, A.D.: HeroesX: The Ancient Greek Hero: Spring 2013 Course Report (2014)
22. Seaton, D.T., Reich, J., Nesterko, S.O., Mullaney, T., Waldo, J., Ho, A.D., Chuang, I.: 6.00 x Introduction to Computer Science and Programming MITx on edX Course Report - 2012 Fall (2014)
23. Strayhorn, T.L.: What role does grit play in the academic success of black male collegians at predominantly white institutions? Journal of African American Studies **18**(1), 1-10 (2014)
24. Zhang, G., Anderson, T.J., Ohland, M.W., Thorndyke, B.R.: Identifying Factors Influencing Engineering Student Graduation: A Longitudinal and Cross-Institutional Study. Journal of Engineering education **93**(4), 313-320 (2004)
25. Wilkowski, J., Deutsch, A., Russell, D.M.: Student skill and goal achievement in the mapping with google MOOC. In: Proceedings of the first ACM conference on Learning@ scale conference, pp. 3-10 (2014)