The Effect of Personality and Course Attributes on Academic Performance in MOOCs

Mahdi Rahmani Hanzaki[1] and Carrie Demmans Epp[1,2](✉)

[1] Department of Computing Science, University of Alberta, Edmonton, Canada
{rahmanih,cdemmansepp}@ualberta.ca
[2] EdTeKLA Research Group, Edmonton, Canada

Abstract. Predicting academic performance has been a topic of research for years, with different factors having been used to predict student grades. One of those factors is personality, with little work having focused on the effect of personality on academic performance in Massive Open Online Courses (MOOCs). Contributing to our lack of understanding of how personality is linked to academic performance in MOOCs, studies that predict academic performance by combining personality with attributes of online course design to have yet to be reported. In this paper, we try to tackle this problem by using personality and level of collaboration (a course attribute) to predict academic performance. We chose level of collaboration as one of the course attributes in our research because social factors, such as the amount of student interaction, can impact learner attrition in MOOCs. We apply machine learning algorithms to two different feature sets. The first feature set only uses personality as a predictor and the second feature set uses personality and level of collaboration in a course as predictors of academic performance. A comparison of these predictive models revealed that adding level of collaboration can increase their performance significantly. These results provide further evidence of the importance of validating classroom-based research in online settings. Moreover, the results of this work can be useful in several ways. For example, we may be able to give better recommendations to users based on their personality and the attributes of courses. We may also be able to adapt course attributes to match the personality characteristics of each student.

Keywords: Personality · Course attributes · MOOC · Educational data mining
Learning analytics · Grade prediction

1 Introduction

Attempts to show that a student's personality can be used to predict his or her academic performance have shown a relationship between these two factors [7, 9]. However, personality is only one potential predictor of how students behave and perform in online courses. Other factors that can affect learners' performance are the attributes of an online course's design [19], such as its length [18] or the facilitation method employed to support student discussion [20]. For instance, students with certain personality traits might feel more comfortable in courses with certain attributes and therefore do better in those courses. As explained in [17], different personalities prefer different ways of

V. Pammer-Schindler et al. (Eds.): EC-TEL 2018, LNCS 11082, pp. 497–509, 2018.
https://doi.org/10.1007/978-3-319-98572-5_38

learning. For example, introverts like to study in quiet environments while extroverts prefer situations that are more dynamic, like classroom discussion.

A relationship between academic success and students' personality traits was found in traditional, face-to-face learning environments [7, 9]. Since the learning environment influences student behaviors, this means that we do not know the extent to which this relationship holds in online learning environments such as Massive Open Online Courses (MOOC). Even though this newer form of online learning environment shares many characteristics with more established online learning settings, it differs in one fundamental way: the course size, diversity in the student body, and design may be biased towards the learning preferences of some students. Furthermore, there is a clear bias in the demographic backgrounds (e.g., sex and country of origin) of those who withdraw from MOOCs [21], and we do not yet have a strong understanding of how learner personality may be tied to student retention or performance in this setting. Consequently, there is a need to study the effect of learner personality and course attributes on the academic performance of MOOC learners (i.e., personality traits together with the style of the course may be a good predictor of academic success). We address these gaps by employing machine learning algorithms to predict academic performance using personality traits and level of collaboration in a course. We chose level of collaboration in a course as one of the course attributes in our research because previous work (e.g., [5]) has shown that social factors, such as the amount of student interaction, can impact learner attrition in MOOCs. To provide a basis for the current study, we will first discuss a theory of personality to support the later discussion of prior work related to predicting academic performance using personality, the relationship between personality and system usage, and the importance of social factors in MOOC attrition.

2 Personality, Behavior, and Predicting Academic Performance

2.1 The Big Five Personality Traits

The Big Five personality traits is a model detailing individuals' personality. It describes an individual's personality using five traits: (1) Neuroticism (tendency to be prone to psychological stress), (2) Extraversion (tendency to seek stimulation in the company of others), (3) Openness (appreciation for art, emotion, adventure, unusual ideas, curiosity, and variety of experience), (4) Agreeableness (tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others), and (5) Conscientiousness (tendency to be organized and dependable). These traits have been studied extensively in areas such as learning and job performance, and they have been related to academic achievement [9], academic motivation [7], and job performance [12].

It has been suggested that personality can affect how people behave in an online learning environment as well as the learning approaches they prefer [17]. For instance, while extroverts tend to prefer more dynamic environments like classroom discussions, introverts prefer to work alone or in small groups and in a quiet environment [17]. These tendencies and the above findings partially motivated our use of a measure of the Big Five personality traits as one of the predictors of academic performance in MOOCs.

To study this phenomenon, we must also understand (1) the relationship between personality and academic performance, (2) the relationship between personality and behavior in online settings, and (3) how students are known to behave in MOOCs. The following sections summarize our current understanding of these contributing factors.

2.2 Personality and Academic Performance in Traditional Learning Environments

Most of the work exploring the relationship between personality and academic achievement has been done in traditional learning environments. In [9], they tried to find the relationship between a student's personality traits and his or her academic motivation and achievement. For this purpose, they performed a survey to measure students' Big Five personality traits, their academic motivation, and their Grade Point Average (GPA). This survey also collected socio-demographic information. In the end, they found that the Big Five personality traits (especially conscientiousness, openness, agreeableness, and neuroticism) are significant predictors of GPA. In another project [7], they tried to find the relationship between academic performance, personality and learning styles. They used the Big Five framework for personality traits and they used the four learning styles introduced by [13] which consist of (1) synthesis-analysis (processing information, forming categories, and organizing them into hierarchies), (2) elaborative processing (connecting and applying new ideas to existing knowledge and to the learner's personal experiences), (3) methodological study (what is traditionally emphasized in most academic environments, such as being careful and methodical while completing all assignments on time), and (4) fact retention (processing information so that the main ideas are memorized with the goal of doing well on tests rather than understanding the meaning of what is being learned). After doing regression analysis, correlation analysis, and mediation analysis, they found that personality traits (neuroticism, openness, agreeableness, and conscientiousness) can predict GPA, with the synthetic analysis and elaborative processing learning styles mediating the relationship between openness and GPA.

These findings demonstrate a relationship between personality and performance, but they rely on self-reported grades rather than students' actual performance, and they do not include course features, such as level of collaboration. Our study includes both, and it considers specific student behaviors within the online course environment. To better understand this potential relationship between personality and learner behavior in online course settings we first detail what we know about personality and behavior in everyday online settings.

2.3 Personality and Internet Usage: Implications for Online Learning

The studies detailed above enable us to understand how the Big Five has been applied in a traditional learning setting to explain the effects that personality has on learning performance. Considering that online settings are different from the traditional classroom setting, the above findings may not hold for online learning as different personalities might have different behaviors when they are in online settings than when they are in a physical location with their classmates and instructor. This behavioral difference is likely if we consider the differences in students' behavior between these settings in their everyday contexts.

Landers et al. [1] investigated the relationship between personality traits and self-reported Internet usage and found that total Internet usage was negatively related to three of the Big Five traits - Agreeableness, Conscientiousness, and Extraversion. The relationship between Internet usage and these personality traits suggests that the personality factors that are predictive of higher performance in face to face classrooms (i.e., agreeableness and conscientiousness [7]) are tied to inaction in online settings. Lander's et al.'s results imply that those with personality types who use the Internet less than others might fare better in traditional learning environments when this environment supports their learning activities. These students may also choose to interact differently with online course materials, which indicates that personality traits may have a relationship with learning in online environments.

2.4 Student Behaviors in MOOCs: Social and Personality Factors

Only a few studies have explored the relationship between personality and MOOC usage. For instance, Chen and colleagues [16] tried to find whether personality influences learner behavior and learner success. In order to do so, they sent questionnaires to collect information about students' Big Five personality traits and combined this data with features describing students' activities in the course (e.g., number of forum posts and time watching videos). Their analysis revealed that various features describing system use were correlated with openness and conscientiousness for learners with low prior knowledge. For those with higher prior knowledge, only conscientiousness was related to multiple system usage features (i.e., the amount of time spent watching video lectures and number of quiz questions learners attempted). Another project aimed to predict student success based on their MOOC usage data from the first week of the course [8]. The course they analyzed had two study tracks (basic and scholar track), and the corresponding certificate was given to students based on their activities in the course. They used logistic regression to predict which certificate the learner received and whether the learner would drop the course or receive a normal certificate. Their models revealed that the students who were more connected in the forum in the first week were more likely to receive a certificate with distinction than a normal certificate.

Consistent with student success being tied to their connectedness within a course, social barriers can contribute to course attrition because our personality and course attributes (some courses are more collaborative than others) influence our behaviors. In [3], surveys were used to collect information about student experiences. Student responses revealed that a lack of social interaction was the main barrier to online learning. Similarly, [5] aimed to understand why people dropped out of MOOCs so they sent a questionnaire to the students who had dropped the course. Students said that "having little interaction with others" [5], which can be related to the style of the course as well as the personality of the individual, was a main reason behind their decision to drop the course. This may mean that level of collaboration in a course might have an impact on students' performance in that course.

Further exploring this idea of student interaction and community in MOOCs, Rosé and colleagues [4] tried to find the social factors that contribute to attrition in MOOCs by focusing only on student participation in the class discussion forums. Using a mixed membership stochastic blockmodel, students' transitions between subcommunities were tracked which showed that membership in one of the subcommunities significantly predicted the dropout rate. This finding suggests that being a member of certain subcommunities may increase a student's probability of dropping a course. However, belonging to a certain subcommunity could be related to personality (e.g. people with similar types of personality might enjoy each other's company). Thus, this work partially motivates our study of the effect of personality on academic performance in MOOCs.

Using a broader set of student behaviors across time, Kloft et al. [2] tried to predict MOOC dropout: they applied machine learning techniques (i.e., principal component analysis [PCA] and support vector machines [SVMs]) to students' clickstream and forum data. They computed attributes like number of requests, number of video views, and number of homework page views to predict whether a student would drop the course in any given week. Continuing in this vein, Zheng et al. [6] studied the effects of small group size on students' drop-out rate and learning performance in a MOOC. In their study, the people who responded to their initial questionnaire (asking them about their demographic and personality information) were automatically divided into groups of 10 by applying k-means clustering to this data. Other students were randomly divided into groups of 10. At the end of the course, a second email was sent to students asking them about their satisfaction. Based on these data, Zheng and colleagues were able to see that the students who had been grouped using k-means clustering had lower dropout rates, confirming that the community a student is a member of can influence their dropout, with personality playing a role in community membership. Since community membership and personality influence drop out, it is reasonable to think that they will also influence course scores. However, not all MOOCs have obvious subcommunities and aspects of how the course is designed can influence the development of these subcommunities, which are based on some form of mutual knowledge sharing and support. This type of knowledge sharing and support is fundamental to community development [25], but as the above examples demonstrate, has not been studied at the whole course level. Consequently, we do not yet know how course-wide collaboration and community within MOOCs as well as student personality predict student performance.

To better study how student performance in online courses relates to their personality and factors of the course design, we use the students' automatically logged interactions, a brief measure of their personality traits, information about how collaborative the course is, and a measure of classroom community to predict student grades as recorded by the system.

3 Method

Using a learning analytics approach, we aimed to predict student performance by augmenting automatically captured information about their grades and collaborative behavior with perceptual information that was collected from students via questionnaires. For this research, we focused on the collaboration level in a course. Our hypothesis is that some personality types might do better in a course with certain attributes. For instance, introverts might feel uncomfortable in a course that requires a lot of collaboration with other learners and therefore they might not do very well. To measure this, we introduced two feature sets to predict student grades. Then, we applied machine learning algorithms to both feature sets to see whether adding level of collaboration in a course would increase the accuracy of our models.

Our hypotheses are:

[H1] Students' MOOC grades can be predicted by the Big Five.
[H2] The joint use of the Big Five and the level of collaboration within a course can predict student MOOC grades.
[H3] Prediction of students' MOOC grades will be more accurate when both the Big Five and level of collaboration are used as predictor variables.

3.1 Dataset

To test our hypotheses, we performed secondary analyses on data from two MOOCs (Epidemics, Pandemics, and Outbreaks; Disaster Preparedness) that had been offered through the coursera platform. This data consists of a pre-course questionnaire which collected information about learner demographics and personality (using Gosling et al.'s short form of the Big Five personality traits [10]), students' grades in those courses, data about forum posts in each course, and a post-course questionnaire which included student responses to Rovai's classroom community scale (CCS) [22].

The questionnaires and system logs recorded learner data using a common identifier. This allowed us to link students' responses to their activities and performance within the MOOCs. We only predict the performance outcomes of those who responded to the pre-course questionnaire: In total, 323 students responded to this questionnaire. Of these students, 85 were taking the Epidemics, Pandemics and Outbreaks MOOC and 238 were taking the Disaster Preparedness MOOC. All students agreed to the use of their data for research purposes.

3.2 Data Pre-processing

Data Cleaning: Removing Invalid Data. We removed the students who: (a) had only partly completed the questionnaire since their partial responses mean that the Big Five result was inaccurate, (b) did not have a user ID associated with them as the result of a temporary bug in the survey software, (c) did not yet have a grade and filled the questionnaire recently (less than 6 months ago) since they may be still taking the course.

This cleaning reduced our dataset to include that from 306 students.

The Big Five. We followed Gosling's instructions [10] to calculate each student's scores for the Big Five personality traits: extraversion, neuroticism, openness, conscientiousness, and agreeableness. Although in [7] only openness, conscientiousness, agreeableness and neuroticism were found to be related to performance, we included extraversion in our feature set because it might influence student performance when combined with the level of collaboration of the course.

Grades. Since grade is a continuous variable and many machine learning algorithms produce better models when the attributes are discrete rather than continuous [29], we needed to address this mismatch between data format and computational approaches. One of the ways to deal with this problem is by binning or categorizing the variables. For this purpose, we coded grade ranges to make the dependent variable discrete. We labeled grades 90% or higher as A, grades from 80% to 90% as B, grades from 70% to 80% as C, grades from 60% to 70% as D, grades from 50% to 60% as E, grades below 50% as F, and dropout as N. The dropout (N) group consisted of those who had completed the questionnaire more than 6 months ago, had not received a grade, and who had no activity in the last six months. This is possible in MOOCs that are offered through the coursera platform because people can transfer from one offering of a MOOC to the next without losing their progress.

Course Collaboration. We defined a measure of course collaboration using a combination of behavioral and perceptual data. We used the forum posts data for each course and divided the total number of questions and answers by the number of active users (an active user is a user who has posted at least one question or answer). This statistic was used as a proxy to for the level of collaboration within each course. The question and answer per active user was 2.7 for the Disaster Preparedness course and 4.6 for Epidemics, Pandemics, and Outbreaks.

We used the Classroom Community Scale (CCS) as a second proxy for the level of collaboration in a course [22]. For this purpose, we calculated CCS for each user in each course (using the post-course questionnaire) and then calculated the CCS average and standard deviation for each course. Then for each student, we added the CCS average and standard deviation according to the course they were enrolled in. The CCS average and standard deviation were 21.42 and 5.94 for the Disaster Preparedness course and 21.87 and 5.91 for Epidemics, Pandemics, and Outbreaks.

3.3 Data Analysis

Model Building. The independent variables derived from the pre-course questionnaire consisted of students' scores for each of the Big Five personality traits: extraversion, openness, conscientiousness, agreeableness, and neuroticism. The course question and answer per user statistic was the independent variable that was obtained using system logs. The post-course questionnaire provided the average and standard deviation of the CCS score for each course; these statistics provided the final independent variables for our models. Students' coded final grade (described above) served as the dependent variable.

We divided the dataset into training and test data. The training data was used to train our machine learning models, and the test data was used to see how good our trained model was. The training data was 70% of the dataset and the test data was 30% of the dataset. The assignment of data to the training and test data sets was random and it was performed at the student level: all of the data from a single student was randomly put into either the test or the train set. After that, we ran different machine learning algorithms on the data with regularization (which helps to avoid overfitting and acts as feature selection) and k-fold cross-validation with $k = 3$ (which avoids overfitting).

We ran different machine learning algorithms on the dataset. Our chosen algorithms were: Support Vector Machines (SVM) [15] and Logistic Regression [14]. SVM builds a model using the training data and uses that model to predict the test data. It is also one of the most common methods used in classification tasks since it is fairly robust and accurate. Logistic Regression uses the sigmoid function (logistic function) to predict the probability of a data point being in each class and assigns the class with highest probability to that data. Logistic regression is useful in this case because it is a simple model that performs well on relatively small amounts of data. Due to the small amount of data, more complex classifiers (e.g., Neural Networks) were not used as they can overfit the model to the data (i.e., create a classifier with high accuracy that generalizes poorly) and thus produce poorer results. Therefore, we chose classifiers that are simpler and are known to perform better on small data sets.

Each classifier was run with different parameters and the parameter with the best accuracy was found. The model was then tested using the test data. As is common in machine learning [28], we chose a set of predefined values for our hyperparameter the inverse of regularization weight, C, and tested the models with those values. Since each model takes a long time to run, it was only practical to test the models on a subset of values. A wide range of values was selected so that different magnitudes for the parameter could be tested. For SVM this parameter set was {C: [1, 2, 10, 15, 20, 100, 1000], kernel: linear}, and for Logistic Regression the parameter set was {C: [1, 2, 10, 100, 1000]}.

Hypothesis Testing. We ran our machine learning algorithms on the data with two different feature sets:

- personality_test - contains student scores for each of the Big Five personality traits. This feature set was used to test H1.
- personality_collaboration_test - contains student scores for each of the Big Five personality traits and the level of collaboration in a course (CCS average, CCS standard deviation, and the number of questions or answers per active user). This feature set was used to test H2.

We used the Zero Rule classifier as a baseline and compared the performance of our algorithms to it. We chose the Zero Rule classifier because it will perform better on our dataset since one of the values of performance ("N" which corresponds to dropout) is more frequent than the other grade classifications: 41.5% of our data was "N". The Zero Rule classifier predicts each entry as "N". It, therefore, has an accuracy of 0.415 for our data (127 students out of 306 were labeled as N).

To test H3, we used paired t-tests to determine whether the model based on personality alone (personality_test) was outperformed by the model that also relies on information about student collaboration (personality_collaboration_test).

4 Results

4.1 Individual Models

For personality_test the classifiers were run in 10 distinct runs with only personality traits as a predictor of the grade. The average and standard deviation of the models' accuracy (percentage of correct predictions) are shown in Table 1. With this feature set, we see little if any improvement over a majority class prediction as represented by the Zero Rule classification model, thus H1 is not supported.

Table 1. Classifier accuracy using the Big Five and Collaboration as predictors of student grade

Classifier	personality_test	personality_collaboration_test
	M (SD)	M (SD)
SVM	0.410 (0.0323)	0.518 (0.0587)
Logistic Regression	0.411 (0.0374)	0.503 (0.0542)
Zero Rule	0.415	0.415

Next, we ran our classifiers with the personality_collaboration_test feature set. We again used 10 distinct runs. For this feature set, we had personality and collaboration in a course as the predictors of the grade. Collaboration was represented via proxies. These proxies were calculated at the course level and included CCS average, CCS standard deviation, and number of questions or answers per active user. Table 1 reports the average and standard deviation of model accuracy for both models. With this feature set, we see improved classification accuracy over the Zero Rule classifier. Therefore, H2 is supported.

4.2 Model Comparison

While the comparison of the Zero Rule classifier to the models based on each feature set suggested H3 would hold because these comparisons supported H2 and failed to support H1, we formally tested H3. To do this, we ran paired t-tests with a 95% confidence interval to determine whether adding the level of collaboration in a course as a predictor had an effect on classifier performance.

For SVM, the personality_collaboration_test feature set was more accurate than the personality_test feature set: $t(9) = -5.3713$, $p < 0.001$, $d = -0.75$. This difference is large. For the personality_test feature set, 125 out of 306 student grades were predicted correctly; and for personality_collaboration_test, 158 students were assigned the correct grade label.

For Logistic Regression, similar results were observed. The model using the personality_collaboration_test feature set was more accurate than the one using the personality_test feature set: $t(9) = -0.3841$, $p < .001$, $d = -0.70$. For the personality_test feature set, 126 out of 306 students' grades were predicted correctly; and for personality_collaboration_test, 154 students' grades were labeled correctly.

In both cases, models were more accurate when they used the combined feature set which supports H3. So, learner personality and the collaborative features of a course can be used together to predict learner grades. Furthermore, the similarity in results between the logistic regression and SVM model suggest that the data is linearly separable so other types of algorithms that group data by dividing that data using a line (logistic regression) or hyperplane (SVMs) may also outperform the Zero Rule classifier.

5 Discussion

In a MOOC setting, personality is not enough to predict student grades, even though personality has successfully predicted student scores in face to face learning environments [7, 9]. The inability of our models to outperform the Zero Rule classifier may be due to the fact that MOOC dropout rates are much higher than those of traditional classrooms. This difference in attrition means that one grade value (i.e., N - withdrawal) occurs far more frequently than others, making the accuracy of a Zero Rule classifier relatively high, which suggests that excluding all those who withdrew from the course may be appropriate in this context. However, this choice would fail to account for a sizable portion of the potential student population and their outcomes.

In addition to the difference in student attrition, aspects of student background are more variable in MOOCs than they are in most classroom settings. In traditional learning environments students usually have similar background preparation, live in the same cultural milieu, and typically speak a common language with facility. Whereas in MOOCs, students are from all around the world, have diverse cultural and language backgrounds, and have received widely different background preparation [26]. This may increase the number of parameters that affect academic performance. Thus, personality might not be enough to predict the academic performance of students in MOOC settings.

The results of our research show that we can predict grades in MOOCs when using both students' personality and the level of collaboration in a course. This supports findings from other contexts suggesting that the social elements of online courses are tied to academic performance [23, 24]. We build upon these findings that collaboration and community predict students' learning activities [23] and their perceived learning [24], by predicting student grades using two proxies for student collaboration: their sense of community and the amount of interaction within their MOOC discussion forum. Our findings confirm some qualitative work [23] that indicated individual student preferences, such as a desire to learn as part of a community and a desire to avoid interaction - which are both indicative of personality traits, influenced student engagement within their online courses. In our case, a lack of course engagement was represented through student attrition, whereas reduced activity patterns represented disengagement in earlier work that investigated online graduate-level courses [20, 23].

Building on this work, our results show that adding level of collaboration as an attribute of the course design helps improve the prediction of students' academic performance. This finding is aligned with those showing how the discussion forum facilitation method that instructors chose to encourage in their online courses influences student collaboration and learning experiences [18]. Our findings suggest, that while personality can predict academic performance in classroom settings [7], additional information is needed if we want to accurately predict student performance in online courses, especially MOOCs. This may mean that a person with the same personality will perform differently based on the design of the online course, especially if one course promotes learning activities that align with that learner's personality while another promotes activities that conflict with that learner's personality.

These results hold potential for informing the creation of adaptive features within MOOCs and MOOC design. Given the relationship between features of courses (such as collaborative activities and forum discussion) and student personality, we may be able to design courses so that student activities support multiple personality traits. Essentially, the same learning objectives may need to be supported through multiple activities so that students who may preferentially select certain learning opportunities are not precluded from acquiring the knowledge or skills that the course is meant to develop. For instance, if one learner is more introverted and dislikes working with others, we may have him or her interact with an agent instead of completing work in a group or this learner may be asked to perform additional assignments or quizzes. In contrast, if a learner is more extroverted and is comfortable in larger groups and working with strangers then we can encourage that student to complete course work as part of a group.

5.1 Limitations

While our dataset was sufficiently large, collecting more data from a broader set of courses would be beneficial. It would also allow us to test model generalizability since student behaviors and elements of student background or personality are known to differ from one MOOC to the next [26].

Our measure of collaboration, while predictive and based on the work of others (e.g., [18, 22, 23]) could be improved through deeper analyses of the interactions in which students engaged. Moreover, this measure only captures one element of course design. As such, it demonstrates the importance of capturing the role of these type of features. Going forward, it would be beneficial to consider other elements of course design, including how the course was facilitated, its length, and the instructional domain.

6 Conclusion

Since little work has explored whether personality and features of course design predict student performance in MOOCs, we built predictive models of student scores based on individual students' Big Five personality traits and the level of collaboration that the student body experienced within a MOOC. These models were tested using two feature sets. The first feature set only consisted of individual students' personality traits and the second set consisted of their personality traits and the level of collaboration in their course. In the end, a students' personality was insufficient for predicting their MOOC score on its own (H1). The discrepancy between this finding and those from classroom-based studies reinforces the importance of validating prior findings from traditional educational settings in MOOC settings. Adding information about course design features, specifically collaboration, improved the models to a point where they could predict student scores (H2 & H3). This reinforces the idea that MOOC design may be biased towards certain types of students [26, 27]. Further study is needed to see which other features of course design can be used to predict student scores and the extent to which these features of online courses interact to influence both student learning and their learning experiences.

References

1. Landers, R.N., Lounsbury, J.W.: An investigation of Big Five and narrow personality traits in relation to internet usage. Comput. Hum. Behav. **22**, 283–293 (2006)
2. 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 Analysis of Large Scale Social Interaction in MOOCs (2014)
3. Muilenburg, L.Y., Berge, Z.L.: Student barriers to online learning: a factor analytic study. Distance Educ. **26**, 29–48 (2005)
4. Rosé, C.P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., Sherer, J.: Social factors that contribute to attrition in MOOCs. In: Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S 2014 (2014)
5. Gütl, C., Rizzardini, R.H., Chang, V., Morales, M.: Attrition in MOOC: lessons learned from drop-out students. In: Uden, L., Sinclair, J., Tao, Y.-H., Liberona, D. (eds.) LTEC 2014. CCIS, vol. 446, pp. 37–48. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10671-7_4
6. Zheng, Z., Vogelsang, T., Pinkwart, N.: The impact of small learning group composition on student engagement and success in a MOOC. In: Proceedings of the 8th International Conference of Educational Data Mining, pp. 500–503 (2015)
7. Komarraju, M., Karau, S.J., Schmeck, R.R., Avdic, A.: The Big Five personality traits, learning styles, and academic achievement. Pers. Individ. Differ. **51**, 472–477 (2011)
8. Jiang, S., Williams, A., Schenke, K., Warschauer, M., O'dowd, D.: Predicting MOOC performance with week 1 behavior. In: Educational Data Mining 2014 (2014)
9. Komarraju, M., Karau, S.J., Schmeck, R.R.: Role of the Big Five personality traits in predicting college students academic motivation and achievement. Learn. Individ. Differ. **19**, 47–52 (2009)
10. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the Big-Five personality domains. J. Res. Pers. **37**, 504–528 (2003)
11. Barrick, M.R., Mount, M.K.: The Big Five personality dimensions and job performance: a meta-analysis. Pers. Psychol. **44**, 1–26 (1991)
12. Costa, P.T., McCrae, R.: Neo personality inventory-revised: (NEO PI-R). Psychological Assessment Resources, Florida (1992)
13. Schmeck, R.R., Ribich, F., Ramanaiah, N.: Development of a self-report inventory for assessing individual differences in learning processes. Appl. Psychol. Measur. **1**, 413–431 (1977)
14. Agresti, A.: Categorical Data Analysis. Wiley-Interscience, New York (2002)
15. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
16. Chen, G., Davis, D., Hauff, C., Houben, G.J.: On the impact of personality in massive open online learning. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, pp. 121–130. ACM (2016)

17. Rose, D.: Personality as it relates to learning styles in online courses. In: Society for Information Technology & Teacher Education International Conference, pp. 827–831. Association for the Advancement of Computing in Education (AACE) (2012)
18. Demmans Epp, C., Phirangee, K., Hewitt, J.: Student actions and community in online courses: the roles played by course length and facilitation method. Online Learn. **21**, 53–77 (2017)
19. Scardamalia, M., Bereiter, C.: Pedagogical biases in educational technologies. Educ. Technol. **48**, 3–11 (2008)
20. Demmans Epp, C., Phirangee, K., Hewitt, J.: Talk with Me: student behaviours and pronoun use as indicators of discourse health across facilitation methods. J. Learn. Anal. **4**, 47–75 (2017)
21. Rovai, A.P.: Development of an instrument to measure classroom community. Internet High. Educ. **5**, 197–211 (2002)
22. Phirangee, K., Demmans Epp, C., Hewitt, J.: Exploring the relation-ships between facilitation methods, students' sense of community and their online behaviours. Special Issue Online Learn. Anal. Online Learn. J. **20**, 134–154 (2016)
23. Richardson, J.C., Maeda, Y., Lv, J., Caskurlu, S.: Social presence in relation to students' satisfaction and learning in the online environment: a meta-analysis. Comput. Hum. Behav. **71**, 402–417 (2017)
24. Garrison, D.R., Anderson, T., Archer, W.: Critical inquiry in a text-based environment: computer conferencing in higher education. Internet High. Educ. **2**, 87–105 (1999)
25. Baikadi, A., Schunn, C.D., Long, Y., Demmans Epp, C.: Redefining "What" in analyses of who does what in MOOCs. In: 9th International Conference on Educational Data Mining (EDM 2016), pp. 569–570. International Educational Data Mining Society (IEDMS), Raleigh (2016)
26. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Learning Analytics and Knowledge (LAK), pp. 170–179. ACM, New York (2013)
27. Kizilcec, R.F., Halawa, S.: Attrition and achievement gaps in online learning. In: Proceedings of the Second (2015) ACM Conference on Learning @ Scale, pp. 57–66. ACM, New York (2015)
28. Kuhn, M.: Applied Predictive Modeling. Springer, New York (2016)
29. Kotsiantis, S., Kanellopoulos, D.: Discretization techniques: a recent survey. GESTS Int. Trans. Comput. Sci. Eng. **32**(1), 47–58 (2006)