Evaluating the Robustness of Learning Analytics Results Against Fake Learners

Citation: Alexandron, Giora et al. "Evaluating the Robustness of Learning Analytics Results Against Fake Learners." EC-TEL 2018, Thirteenth European Conference on Technology Enhanced Learning, 3-6 September, 2018, Leeds, United Kingdom, HTTC e.V., 2018.

As Published: http://www.ec-tel.eu/index.php?id=791

Publisher: HTTC e.V.

Persistent URL: http://hdl.handle.net/1721.1/116511

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Evaluating the Robustness of Learning Analytics Results Against Fake Learners

Giora Alexandron[1], José A. Ruipérez-Valiente,[2], Sunbok Lee[3], and David E. Pritchard[2]

[1] Weizmann Institute of Science giora.alexandron@weizmann.ac.il
[2] Massachusetts Institute of Technology, {jruipere, dpritch}@mit.edu
[3] University of Houston slee96@uh.edu

Abstract. Massive Open Online Courses (MOOCs) collect large amounts of rich data. A primary objective of Learning Analytics (LA) research is studying these data in order to improve the pedagogy of interactive learning environments. Most studies make the underlying assumption that the data represent truthful and honest learning activity. However, previous studies showed that MOOCs can have large cohorts of users that break this assumption and achieve high performance through behaviors such as Cheating Using Multiple Accounts or unauthorized collaboration, and we therefore denote them *fake learners*. Because of their aberrant behavior, fake learners can bias the results of Learning Analytics (LA) models. The goal of this study is to evaluate the robustness of LA results when the data contain a considerable number of fake learners. Our methodology follows the rationale of 'replication research'. We challenge the results reported in a well-known, and one of the first LA/Pedagogic-Efficacy MOOC papers, by replicating its results *with* and *without* the fake learners (identified using machine learning algorithms). The results show that fake learners exhibit very different behavior compared to true learners. However, even though they are a significant portion of the student population (∼15%), their effect on the results is not dramatic (does not change trends). We conclude that the LA study that we challenged was robust against fake learners. While these results carry an optimistic message on the trustworthiness of LA research, they rely on data from one MOOC. We believe that this issue should receive more attention within the LA research community, and can explain some 'surprising' research results in MOOCs.

Keywords: Learning Analytics · Educational Data Mining · MOOCs · Fake Learners · Reliability · IRT.

1 Introduction

The high resolution behavioral data that MOOCs collect provide new opportunities to study learners behavior, in order to improve the pedagogy of interactive learning environments, and to develop data-driven tools for personalization and analytics [20,10]. The implicit assumptions behind such research are typically that the data collected represent genuine learning behavior, and that there are hidden causal relationships between learners behavior and their success, which can be discovered using Educational Data Mining (EDM).

Fake Learners. However, several studies revealed that there are a considerable amount of users who use cynical means to succeed in the courses, such as Cheating Using Multiple Accounts [3,18,4,15], or unauthorized collaboration [19]. Such users break the 'genuine learning behavior' assumption, thus we refer to them as *fake learners*. The data in fake learners logs is largely an artifact with respect to explaining their performance. As was pointed out in [4], this can bias LA and EDM results. For example, fake learners typically make minimal interaction with the learning materials, yet show high success; this can lead to false conclusions regarding the effectiveness of different learning paths or the pedagogic efficacy of course materials. However, this issue remains an open question.

Research Questions. The goal of the current research is to address this issue directly by measuring the effect of fake learners on LA research. Specifically, we study the following Research Questions (RQs):

1. (RQ1) What is the difference between the 'fake' and 'true' learners with respect to the amount of use of different course materials (e-text, videos, checkpoint items, homework, and quizzes), and to various performance measures?
2. (RQ2) What is the effect of fake learners' data on the results of a correlation study, such as the relationships between resource use and performance?

To answer these, we challenge the findings reported in one of the first studies of pedagogic efficacy in MOOCs [7], by replicating its results with and without fake learners data. To identify the fake learners, we use the algorithms published in [4,19].

Findings in brief. In the course that we study about ∼15% of the certificate earners are fake learners (of the types that we can detect; we expect that there are more that are still under the radar). With respect to RQ1, they have a very distinguished learning behavior (e.g., a much lower use of course materials). With respect to RQ2, their data effect the correlations that were studied in [7] in a way that we interpret as not very significant (i.e., no 'change of trend').

Our contribution. Due to the large amount of fake learners that were reported in MOOCs, the risk that fake learners' data can bias LA discoveries raises doubts on the trustworthiness of such studies. However, identifying and removing such learners from the data requires sophisticated algorithms that are not available off-the-shelf. We build upon our previous research on both identifying fake learners, and pedagogic efficacy in MOOCs, to make a stride in the direction of evaluating the robustness of LA research against fake learners. To the best of our knowledge, this is the first rigorous attempt to study this issue.

A broader perspective. This research also touches upon two issues that we believe should receive much more attention within the LA and EDM communities. One is *verification* and *validation* of computational models that rely on noisy data that its quality can be affected by malicious or otherwise unusual behavior. Second is *replication research* as a scientific methodology to explore and confirm the generalizability of LA and EDM results to different educational contexts and their stability under various conditions.

2 Methodology

In this section we describe in brief the experimental setup and the EDM procedures that are used. Some of the methodological contents of this section have been reused from previous work [18,4]

2.1 Experimental Setup

The context of this research is MITx MOOC 8.MReVx, offered on edX.org in Summer 2014[4]. The course attracted 13500 registrants, of which 502 earned a certificate. Gender distribution was 83% males, 17% females. Education distribution was 37.7% secondary or less, 34.5% College Degree, and 24.9% Advanced Degree. Geographic distribution includes US (27% of participants), India (18%), UK (3.6%), Brazil (2.8%), and others (total of 152 countries).

The course covers the standard topics of a college introductory mechanics course with an emphasis on problem solving and concept interrelation. It consists of 12 required and 2 optional weekly units. A typical unit contains three sections: Instructional e-text/video pages (with interspersed concept questions, aka Checkpoints), homework, and quiz. Altogether there are 273 e-text pages, 69 videos, and about 1000 problems.

2.2 Data Mining

Identifying Fake Learners We define 'fake learners' as users who use unauthorized methods to improve their grade is a way that does not rely on learning (or pre-knowledge). Currently, we have means to identify two types of such methods.

1. **Cheating Using Multiple Accounts:** This refers to users who maintain multiple accounts: A *master* account that receives credit, and *harvesting* accounts/s used to collect the correct answers (typically by relying on the fact that many questions provide the full answer, or at least True/False feedback, after exhausting the maximum number of attempts) [3,18]. We note that in this method the multiple accounts are used by the same person. Using the algorithm described in [18,4], we identified 65 ($\sim$13%) of the certificate earners who used this method. Hereafter we use the term CAMEO that was suggested by [15] for this phenomenon.

[4] `https://courses.edx.org/courses/MITx/8.MReVx/2T2014/course/`

2. **Collaborators:** MOOC learners might work in study groups or with peers to submit assignments together. These associations are found using the algorithm described in [19] that relies on dissimilarity metrics and a data-driven method to find accounts that tend to submit their assignments in close proximity in time. Sometimes these associations represent real learning collaboration between peers taking a MOOC together and working towards a common goal, in other occasions they may represent more unethical and systematic dishonest behaviors, such as one learner passing the correct quiz responses to a friend every week. Overall, we identified 20 (∼4%) of the certificate earners who submitted a significant portion of their assignments with peers. As there might be some overlapping between the detection of the two methods, we give the CAMEO algorithm priority as it represents a more specific behavioral pattern. Among the collaborators, 11 also used the CAMEO method. Hereafter we refer as 'collaborators' to the 9 accounts who were not CAMEO users.

3 Results

The results are organized as follows. First, we examined the differences between fake and true learners with respect to fundamental behavioral characteristics. Then, we examine the effect of these differences on correlations that seek to associate behavior and performance.

3.1 Differences in Behavioral Characteristics

Time on Course Resources The first measure that we examine is the amount of time that the fake learners spent on different course resources, compared to the true learners. We quantify *Reading Time* (time that the users spent on explanatory pages), *Watching Time* (time spent on videos), and *Time on Homework* (time spent in pages that contain homework items). Table 1 presents, per resource type, the mean time spent by fake/true learners, and *p-value* for the hypothesis that the fake learners spent less time on this type of resource.

Table 1. Time on resources.

Item Type	True learners	Fake Learners	*p-value*
Reading time	17.8	9.9	<0.001
Watching time	3.4	2.1	<0.1
Homework time	14.4	9.2	<0.001

From the table, it is quite clear that fake learners spent less time on the instructional resources. A more detailed illustration of the differences between the groups, also separating the fake learners into their subgroups, is presented in Figure 1.

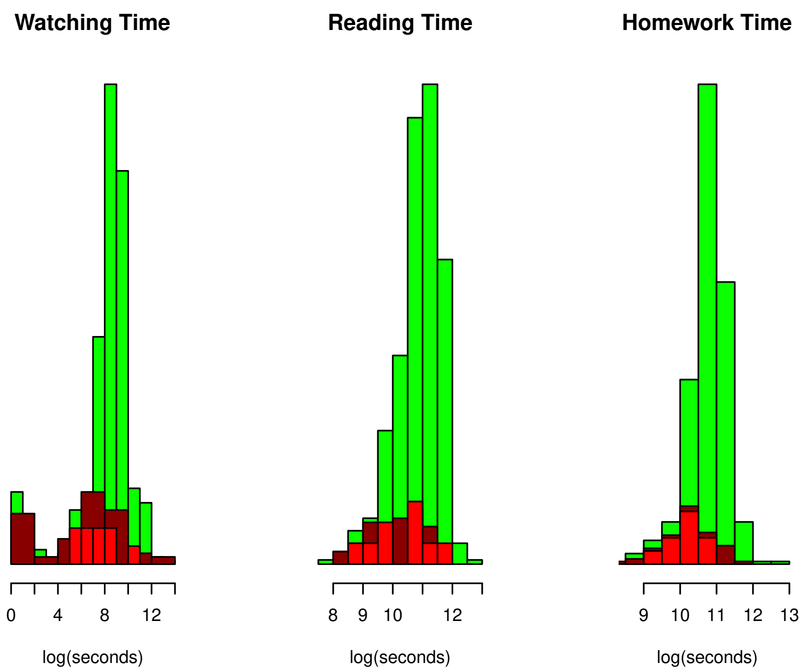

Fig. 1. Time on Resources: True learners in green; CAMEO in dark-red; Collaborators in red

Overall, the behavior of the two subgroups within the fake learners cohort – cheaters and collaborators, is quite similar (confirmed with *t-test*).

Proportion of Items Solved Next, we measure the proportion of assessment items that students attempted (either correct or incorrect). As explained in Subsection 2.1, the course contains mainly three types of assessment items: Checkpoint, Homework, and Quiz. The reason for analyzing them in separate is that their different characteristics with respect to weight (points for solving them), and the easiness of getting the correct answer without effort (e.g., whether it is possible to receive the full answer after exhausting the possible attempts). Our assumption is that fake learners would factor that into their decision of whether to spend time on these items. For example, as checkpoint items have low weight, we assume that fake learners would show less interest in solving them. Quiz items have high weight, but are harder to cheat upon (no 'show answer', only True/False feedback). Homework offers relatively high weight and have 'show answer' enabled, which probably makes them ideal for fake learners (high 'return on investment'). Table 2 contains the proportion of items solved by each group, and the *p-value* that the fake and true learners have a different distribution.

Table 2. Proportion of items solved.

Item Type	True learners	Fake Learners	*p-value*
Quiz	0.63	0.58	<0.001
Homework	0.49	0.46	<0.01
Checkpoint	0.73	0.67	<0.01

Again, there is a clear difference between the groups, with fake learners trying less items. We also examine the distribution in more detail, and separate the fake learners into their two subgroups. This is shown in figure 2. Again, we do not see a significant difference between cheaters and collaborators in each of these metrics. In Section 4 we analyze these results and discuss the characteristics which make certain questions more attractive for fake learners.

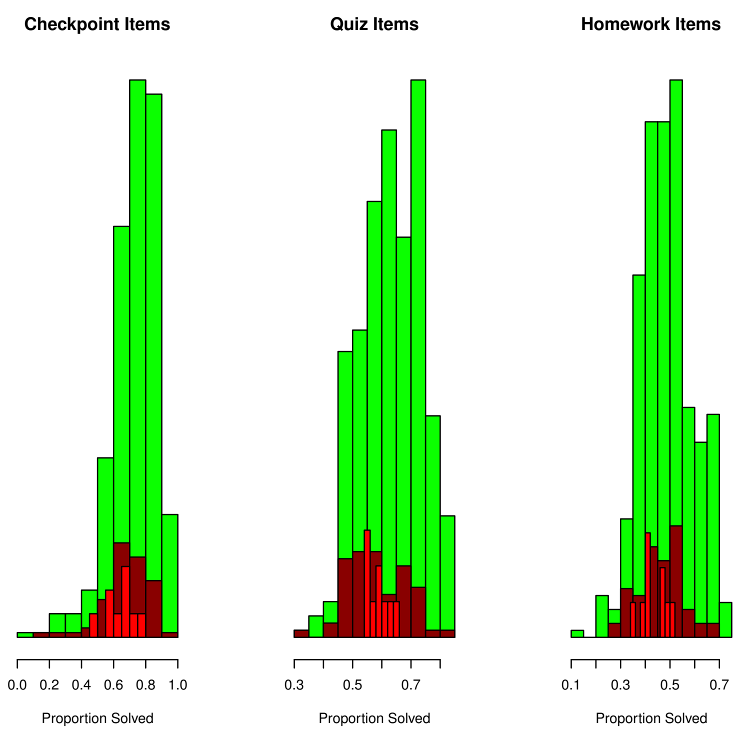

Fig. 2. Proportion of Items Solved: True learners in green; CAMEO in dark-red; Collaborators in red

Performance Measures Student performance can be measured in various ways. We focus on the following metrics.

- Grade: Total points earned in the course (60 points is the threshold for certificate)
- Ability: Student's skill in a 2PL Item-Response Theory (IRT) model, based on first attempt, with population containing the certificated users *(N=502)*, and items that were answered by at least 50% of these users. We chose IRT because students' IRT ability scores are known to be independent of the problem sets each student tried to solve [9]. Missing items were imputed using a mean imputation. We used R's *TAM* package [5].
- Weekly Improvement: Per student, this is interpreted as the slope of the regression line fitted to the *weekly* IRT ability measures (e.g., fitting 2PL IRT on each week of the course in separate). One of the important issues that must be addressed during the calculation of the IRT slopes is to set up the common scale across weekly IRT scores. IRT is a latent variable model, and a latent variable does not have any inherent scale. Therefore, each IRT estimation defines its own scale for the latent variable. Equating is the process of transforming a set of scores from one scale to another. We used mean and sigma equating to set up a common scale across weekly IRT scores. The equated IRT slope captures *the change* in students' relative performance during the course. For example, a student who has average performance in all the weeks, will have 0 relative improvement.
- Proportion Correct on First Attempt (CFA): The proportion of items, among the items that the student attempted, that were answered correctly on the first attempt.
- Mean Time to First Attempt (TTF): The average time it took a student between seeing the item (operationalized as entering into the page in which the item resides, or in case of multiple items in page, answering the previous item), and making the first attempt.
- Mean Time on Task (TOT): The average time the student spent on an item (e.g., sum of time for all attempts).

The mean values for these performance measures, and the *p-value* for the hypothesis that fake and true learners have different distribution, are presented in Table 3.

According to the table, fake learners are significantly faster (on both measures), but on the other metrics do not differ significantly from the true learners. However, it turns out that on these metrics there is a significant difference *within* the fake learners cohort, between the CAMEOers and the collaborators. This is demonstrated in Figure 3. CAMEOers have higher grade (0.85 vs. 0.77), ability (0.21 vs. -0.66), and CFA (0.79 vs. 0.67), than collaborators, all with significant *p-values.*

In fact, on ability and CFA, we get that CAMEOers > true learners > collaborators, with ability = (0.21, -0.07, -0.66), and CFA = (0.79, 0.76, 0.67),

[5] `https://cran.r-project.org/web/packages/TAM/TAM.pdf`

Table 3. Performance of true and fake learners.

Measure	True learners	Fake Learners	*p-value*
Grade	0.85	0.83	0.27
Ability	-0.07	0.1	0.23
Weekly Improvement	0.01	0.09	<0.05
Proportion CFA	0.76	0.77	0.42
Mean TTF	112s	72s	<0.001
Mean TOT	150s	97	<0.001

respectively. The *p-value* for these are borderline (< 0.1 for CAMEOers vs. true learners, and < 0.2 for true learners vs. collaborators), but it demonstrates that on these metrics the fake learners have different behaviors, in which their average is quite similar to the average behavior of the true learners.

The fact that CAMEOers can have higher ability, yet the same grade, as true learners, is due to the nature of IRT, which weighs items according to their empirical behavior, and due to the fact that we train the IRT models on first attempts data. CAMEOers choose items strategically, and have very high proportion of CFA.

Summary of Differences Overall, we see that fake learners spent much less time on course resources, and attempted less items. In the case of response time, we see that fake learners are much faster to solve exercises correctly. Regarding success metrics, we see that on average there is no significant difference between true and fake learners with respect to grade, ability, and CFA. However, a finer look into the subgroups reveals that CAMEOers have higher ability and CFA, and collaborators have lower ability and CFA, than true learners (though strictly speaking the *p-value* for this ordering is slightly above the 0.05 customary threshold).

3.2 Correlation Study

Next, we examine the effect of the differences in the behavioral metrics presented above on fundamental relationships – between response time and success, and between resource use and aggregated performance in the course.

Response Time vs. Success One of the issues of interest in education research is the relation between *response time*, and the likelihood of making a correct attempt. On one hand, better students might be faster (between-person differences), but on the other hand, spending more time on the question increases the probability of finding the correct solution (within-person effect) [11]. This is under the assumption that students try to learn. However, the performance of fake learners is affected by other factors. Figure 4 (left) shows the relation between *proportion CFA*, and *mean time to first attempt*, for the fake and true learners (red and light-blue dots, respectively). The red, steep regression line is

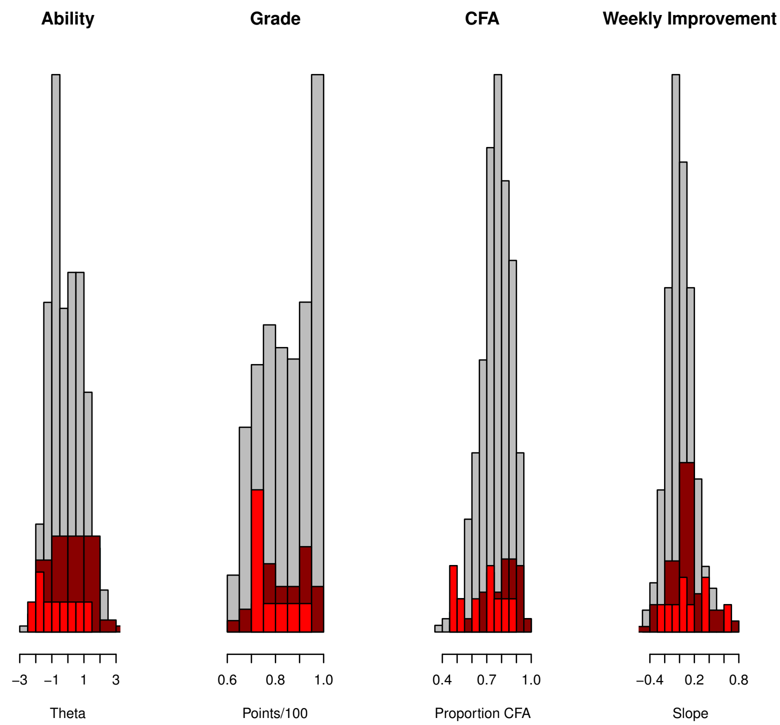

Fig. 3. Performance Measures: True learners in gray; CAMEO in dark-red; Collaborators in red

of fake learners; the blue, moderate line is of true learners; the dashed line is for the entire population. The difference between the blue and the dashed lines is how much the fake learners 'pull' the correlation down. Figure 4 (right) shows the same for the effect of fake learners on the relation between *IRT ability*, and *mean time on task* (the difference between *time on task* and *response time* is that the former is the time for all attempts, while the latter is only the time till the first attempt; most items in the course allow multiple attempts, and there is no penalty for using them).

As can be seen in both figures, the relationship between speed and performance is very different for fake and true learners, however the fake learners cohort is not big enough (about 15%) to change the overall trend dramatically.

Resource Use and Aggregated Performance Measures The relation between the time students spend on different types of instructional materials, and their performance on various metrics, was studied in [7]. This is one of the first MOOCs EDM research papers, and it studied core questions related to the effectiveness of online learning materials. Here, we replicate the specific relationships

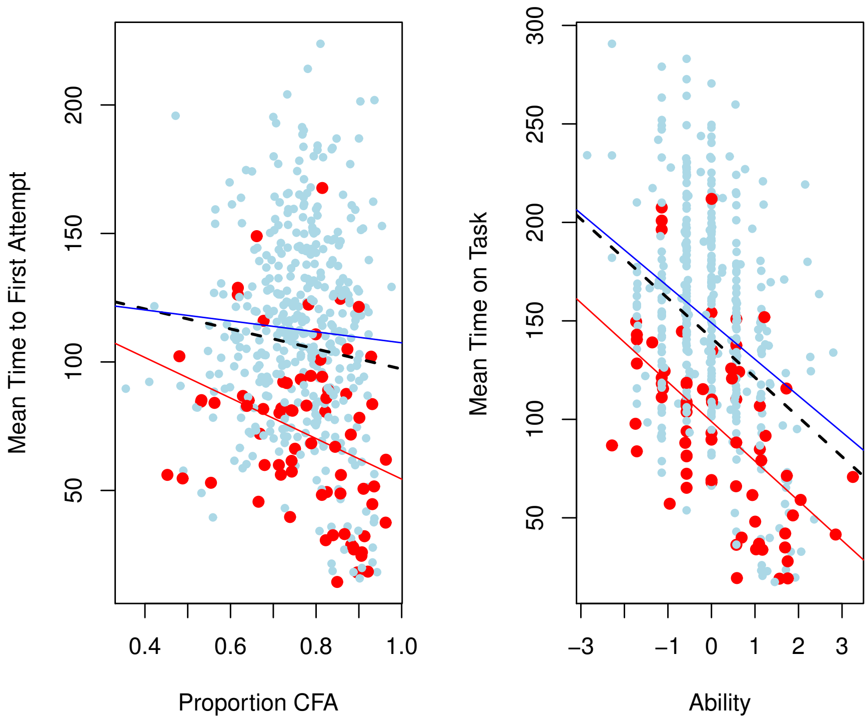

Fig. 4. The effect of fake learners on the relation between time and measures of skill

studied in that paper, and how they change when removing fake learners. The results are presented in Figure 5, which also adopts the visualization style used in [7]. It shows the relation between the amount of time spent on various course resources, and certain performance metrics. For each pie, the outer circle is the whole group, and the inner is the same measure, **after removing the fake learners from the data**. The angle of the piece is the size of the correlation. Clockwise angle represents positive correlation (colored with green), and counter clockwise represents negative correlation (colored in red). Gray color means $p - value > 0.05$. **The difference between the angle of the outer circle, and the angle of the inner one, is the effect of fake learners' data on the correlation.**

Let us examine the correlations with $p{-}value < 0.05$ (colored with red/green). With respect to Grade vs. Homework and Reading Time, there is almost no effect (angle of inner and outer piece is almost identical). With respect to Ability vs. Homework and Reading Time, we see a *negative* correlation, which is *reduced* when removing fake learners. With respect to Weekly Improvement vs. Homework Time, we see a *positive* correlation, which *increases* when removing fake learners.

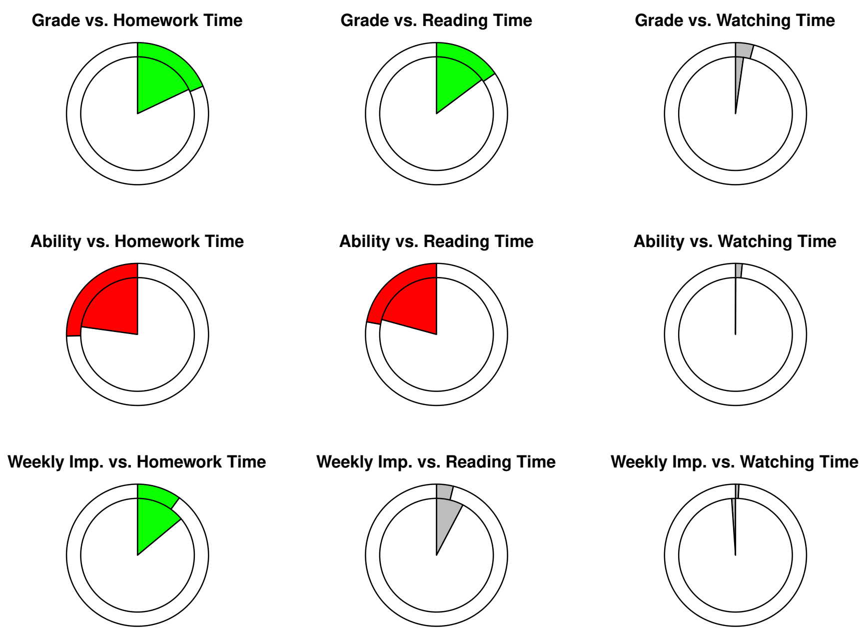

Fig. 5. Effect of fake learners on correlation between performance and time on course resources

4 Discussion

The findings reported in Section 3.1 indicate that fake learners tend to spend much less time than true learners on course materials, and to attempt fewer assessment items. Most likely, this is the result of being interested in easy ways to achieve a certificate. Since they have means other than learning to find correct answers, they can score well without spending a lot of time on the learning materials. Since the threshold for earning a certificate is 60% of the points, they can be selective with the items they choose to solve, and concentrate on ones they can solve more easily, either legitimately or not. This explains why they attempt less items, and also their *performance metrics* – why their *grade* is slightly lower, and why their *IRT ability* and *proportion CFA* are slightly higher, than the true learners (*grade* is very sensitive to the number of items solved, but

IRT and CFA are basically not). Since they solve many of the questions using non-legitimate means, their response time is much faster than the other learners (very fast response time is a hallmark of cheating [17]).

Due to these differences, fake learners can bias various statistics and affect data-driven research results and decision-making processes. This depends not only on the behavioral differences, but also on the size of the cohort. In the course that we study, the population of the fake learners is roughly 15% of the certificate earners, which is a considerable sub-population.

Our results show mild effect on the strength of the relationships, so our conclusion is basically that the *correlation study* was robust against fake learners. This suggests that even a sub-population of ∼15% with very different characteristics is still not a threat to 'average' correlations. However, attempts to study selected groups like 'efficient learners' (e.g., learners who are fast and successful) would be very prone to distortions due to fake learners cohorts. We would also caution against doing expert-novice studies by classifying the very top students (containing a high percentage of fake learners, especially CAMEOers) as representative of 'experts'. Also, we can expect that the percentage of fake learners, and subsequently their effect, may rise as the reward for good performance is raised (e.g. in getting a grade from a college) [17].

From a systematic point of view, using 'black box' computational models that rely on data requires taking proper steps to verify that the data are trustworthy. This was already acknowledged in other domains (and is considered a major challenge), but to date received only minor attention in the LA and EDM research community. We believe that verification, validation, and quality assurance of LA and EDM models and results should receive much more attention, and that this is an important part in the process of becoming a mature field. Our study makes an initial stride in this direction.

The main limitation of our research is that it is based on a single course and examines a limited set of learning analytics. Future research can examine a wider set of courses and challenge additional reported studies that could have been affected. Also, while the definition of 'fake learners' is broad and refers to various types of cynical learning behaviors, our results are based on the limited set of such behaviors that we currently know how to detect. We hope that future research will shed light on more types of 'fake learning' behaviors, and on ways to detect and prevent them.

5 Related Work

EDM and LA are emerging disciplines that aim to make sense of educational data in order to better understand teaching and learning, with the applied goal of improving the pedagogy of online learning environments, and developing 'smart' content and tools [20,10]. In particular, open learning environments such as MOOCs, where the large enrollment, wide scope (typically, a few weeks course), variety of learning materials, the relative freedom for learners to navigate, and the high-resolution data being collected, provide "unparalleled opportunities to perform data mining and learning experiments" [7] (pp. 1). A partial list of studies includes comparing active vs. passive learning [13], how students use videos [12,1], which resources are helpful [7,2,8,14], and many others.

The basic assumption behind most EDM/LA studies (though this assumption is typically not articulated), is that the data represent genuine learning behavior of individuals. This assumption is broken by fake learners, e.g., users who succeed in the course using means such as Cheating Using Multiple Accounts [3,18,4,15], or conducting some sort of collaboration [19]. In the context of Intelligent Tutoring Systems and K12 learners, Baker *et al.* [6] defined a related phenomenon termed *gaming the system*, which they describe as "Attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning [...]". This makes this behavior a sort of 'fake learning', however, *gaming the system* is not interpreted as illegitimate, and is more associated with frustration, lack of motivation, and inadequate design of the learning environment [5].

To the best of our knowledge, the influence of fake learners (and more generally, aberrant behavior) data on the reliability of models and results was not studied within the EDM/LA community. More generally, this issue can be seen as an instance of what Cathy O'Neil calls "Weapons of math destruction" [16]: Data-driven algorithms that make wrong decisions due to bugs, wrong assumptions on the data or the process that generated them, etc. An example within the context of education is the reported incident of a teacher who was fired because of a 'performance assessment' algorithm which yielded that her class did not improve enough during the school year[6]. She argued that the previous year's test scores were artificially raised by cheating (possibly by a teacher who wanted to increase his/her evaluation). In social-media, the Facebook–Cambridge Analytica data scandal[7] demonstrates how fake accounts can be used to collect data and affect social trends.

References

1. Alexandron, G., Keinan, G., Levy, B., Hershkovitz, S.: Evaluating the effectiveness of educational videos. In: EdMedia 2018 (To appear) (2018)
2. Alexandron, G., Pritchard, D.: Discovering the Pedagogical Resources that Assist Students in Answering Questions Correctly A Machine Learning Approach. Proceedings of the 8th International Conference on Educational Data Mining pp. 520–523 (2015)
3. Alexandron, G., Ruiperez-Valiente, J.A., Pritchard, D.E.: Evidence of mooc students using multiple accounts to harvest correct answers (2015), learning with MOOCs II, 2015

[6] https://www.washingtonpost.com/local/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html?utm_term=.21c5f0af7fd3

[7] https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

4. Alexandron, G., Ruipérez-Valiente, J.A., Chen, Z., Muñoz-Merino, P.J., Pritchard, D.E.: Copying@Scale: Using Harvesting Accounts for Collecting Correct Answers in a MOOC. Computers and Education **108**, 96–114 (2017)
5. Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments. Journal of Interactive Learning Research **19**(2), 162–182 (2008)
6. Baker, R.S.J.d., De Carvalho, A.M.J.B., Raspat, J., Aleven, V., Corbett, A.T., Koedinger, K.R.: Educational software features that encourage and discourage "gaming the system". In: Proceedings of the 2009 Conference on Artificial Intelligence in Education. pp. 475–482 (2009)
7. Champaign, J., Colvin, K.F., Liu, A., Fredericks, C., Seaton, D., Pritchard, D.E.: Correlating skill and improvement in 2 MOOCs with a student's time on tasks. Proceedings of the first ACM conference on Learning @ scale conference - L@S '14 (March), 11–20 (2014)
8. Chen, Z., Chudzicki, C., Palumbo, D., Alexandron, G., Choi, Y.J., Zhou, Q., Pritchard, D.E.: Researching for better instructional methods using AB experiments in MOOCs: results and challenges. Research and Practice in Technology Enhanced Learning **11**(1), 9 (2016)
9. De Ayala, R.: The Theory and Practice of Item Response Theory. Methodology in the social sciences, Guilford Publications (2009)
10. U.S. Department of Education, O.o.E.T.: Enhancing teaching and learning through educational data mining and learning analytics: An issue brief (2012)
11. Goldhammer, F.: Measuring ability, speed, or both? challenges, psychometric solutions, and what can be gained from experimental control. Measurement: Interdisciplinary Research and Perspectives **13**(3-4), 133–164 (2015)
12. Kim, J., Guo, P.J., Seaton, D.T., Mitros, P., Gajos, K.Z., Miller, R.C.: Understanding in-video dropouts and interaction peaks in online lecture videos (2014)
13. Koedinger, K.R., Mclaughlin, E.A., Kim, J., Jia, J.Z., Bier, N.L.: Learning is Not a Spectator Sport : Doing is Better than Watching for Learning from a MOOC pp. 111–120 (2015)
14. MacHardy, Z., Pardos, Z.A.: Toward the evaluation of educational videos using bayesian knowledge tracing and big data. In: Proceedings of the Second (2015) ACM Conference on Learning @ Scale. pp. 347–350. L@S '15, ACM (2015)
15. Northcutt, C.G., Ho, A.D., Chuang, I.L.: Detecting and preventing "multiple-account" cheating in massive open online courses. Comput. Educ. **100**(C), 71–80 (Sep 2016)
16. O'Neil, C.: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, New York, NY, USA (2016)
17. Palazzo, D.J., Lee, Y.J., Warnakulasooriya, R., Pritchard, D.E.: Patterns, correlates, and reduction of homework copying. Phys. Rev. ST Phys. Educ. Res. **6**, 010104 (2010)
18. Ruiperez-Valiente, J.A., Alexandron, G., Chen, Z., Pritchard, D.E.: Using Multiple Accounts for Harvesting Solutions in MOOCs. Proceedings of the Third (2016) ACM Conference on Learning @ Scale - L@S '16 pp. 63–70 (2016)
19. Ruipérez-Valiente, J.A., Joksimović, S., Kovanović, V., Gašević, D., Muñoz Merino, P.J., Delgado Kloos, C.: A data-driven method for the detection of close submitters in online learning environments. In: Proceedings of the 26th International Conference on World Wide Web Companion. pp. 361–368 (2017)
20. Siemens, G.: Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist (10), 1380–1400 (2013)