

How is students' online learning behavior related to their course outcomes in an introductory physics course?

Geoffrey Garrido

Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, Florida, 32816

Matthew W. Guthrie, and Zhongzhou Chen

Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, Florida, 32816

This study investigates patterns in students' learning and problem-solving behavior as they proceed through a sequence of 10 mastery-based online learning modules and how these patterns correlate with overall course outcome. Students' interaction with each module, as measured by analyzing the platform log data, was categorized into nine different states. The student population was divided into top, middle and bottom cohorts based on their total course credit, and we visualized each cohort's distribution among the nine states over the 10 modules using a series of parallel coordinates graphs. We found that the patterns of interaction were mostly similar on the first six modules, but are significantly different on modules 7-10. For the later modules, the top cohort mostly concentrated on the state corresponding to high problem-solving effort after learning, while the majority of the bottom cohort did not access the learning materials after multiple failed attempts.

I. INTRODUCTION

The large amounts of data generated by online learning systems provide researchers with ample opportunities to observe and measure various aspects of students' learning behavior, such as accessing learning resources [1–3], problem-solving [4–7], decision making [8], and engagement in online discussion [9,10]. There is a significant and rapidly growing body of research that examines the relationship between students' online learning behavior and course outcomes. For example, Bowman et. al. [4] found that the average time spent on online homework problems positively correlated with final course grade in an introductory chemistry course, while using multiple attempts for each problem correlated negatively. Kortemeyer observed that regular access of online materials is a predictor for success on physics exams [11], and that the frequency of physics related discussion posts in an online discussion is a strong predictor of both course success and a positive attitude towards physics [10]. In an online psychology course, Koedinger et. al. estimated that students who are engaged in more activities learned six times more than students watching more videos or reading more pages [8]. While many earlier studies described student behavior via one or more variables averaged over the entire course, more recent studies have attempted to extract detailed patterns in sequential learning activities over an entire course [5] or learning session [12]. The implicit assumption behind those studies seems to be that differences in student behavior are largely being determined by the characteristics of the students, and less so by the differences in the instructional resources presented to the students. Such an assumption is acceptable when the behavioral patterns are observed over a large number of learning resources such as an entire course. However, this type of analysis cannot reveal how students' learning behavior changes as they interact with different learning resources in the course, and therefore has limited ability to inform instructors which resources in the course are more likely to be challenging to low performing students.

The current study investigates how students' behavior changes as they proceed through a relatively small number of learning resources and problems in sequence, and whether some activities in the sequence lead to larger behavioral differences between high and low performing students than others.

The learning resources and problems are organized and presented to students in the form of a sequence of 10 Online Learning for Mastery (OLM) modules [13–15], designed based on the idea of mastery learning [16–18]. We categorized students' interaction with each OLM module into 9 interaction states, allowing each student's changes in interaction state when proceeding through the 10 modules to be visualized via a series of parallel coordinates graphs. The major patterns of change between different states for a given

student population can be identified using a hierarchical clustering algorithm.

We divided the students in a calculus based introductory class into top, middle and bottom cohorts according to their total course credit earned at the end of the semester, and compared the major interaction patterns of the three cohorts to answer the following two questions:

1. How do students change from one interaction state to another on different modules? Is one cohort more stable in terms of their behavior than another?
2. On which of the modules do the top and bottom cohorts occupy different interaction states, and on which modules are the differences greatest?

II. METHODS

A sequence of 10 OLM modules on the topic of conservation of mechanical energy was assigned as homework to students enrolled in a calculus based introductory physics course at University of Central Florida, to be completed in two weeks. Students were allowed 5 attempts on the assessment component of each module. On the first three attempts students were given slightly different isomorphic problems, while on the final two attempts the problems from attempts 1 and 2 were repeated. A total of 230 students attempted at least one module, and 223 students attempted all 10 modules. The modules are implemented on an open source platform, Obojobo [19], developed by the Center for Distributed Learning at the University of Central Florida.

A. Classification of interaction states

Each OLM module consists of an instructional component (IC) containing learning resources, and an assessment component (AC) containing 1-2 multiple-choice problems (FIG 1). To pass, students must answer all questions correctly. Upon accessing a new module, students are required to make one attempt at the AC before being able to access the IC. The IC contains learning resources directly related to solving the problem in the AC. 77% of students accessed the IC after their first attempt, and 13% accessed the IC after their second. A student can proceed onto the next module in the sequence after either passing the AC or using up all the attempts. Depending on when or whether the

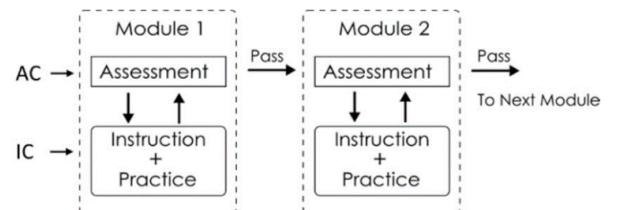


FIG 1: Schematic outline of an OLM module sequence

student studied the IC of the module, students' interaction can be sorted in to 4 general categories:

- I. Initial pass: The student passed the AC on his/her 1st or 2nd attempt without accessing the IC.
- II. Pass after study: The student studied the IC after failing the 1st or 2nd attempts on the AC, then passed within 2 additional attempts after accessing the IC.
- III. Fail after study: The student studied the IC after failing the 1st or 2nd attempts on the AC, but also failed the next two attempts.
- IV. Multiple attempts without study: The student made 3 or more attempts on the AC without accessing the IC.

We believe that category IV is different in nature from category I, because if a student deliberately chose not to access the IC after more than 2 failed attempts, they are more likely to be guessing than actually trying to learn from the modules.

In addition, previous research into online learning has repeatedly demonstrated that some students will submit answers to problems in unusually short amounts of time, which are often attributed to either random guessing or answer copying [20–22]. Therefore, a student who passed a module in an abnormally short attempt is likely interacting with the AC of the module qualitatively differently from someone who spent a normal amount of time on an AC attempt. Different studies have suggested different cutoff times for classifying abnormally short attempt times, ranging from 20 s to 60 s [14,20,21]. In the current study, we use 40 s as a cutoff to distinguish "Brief" attempts from "Normal" attempts, which was measured to be approximately the time required to open the AC, briefly look at the problem body, and submit a random choice.

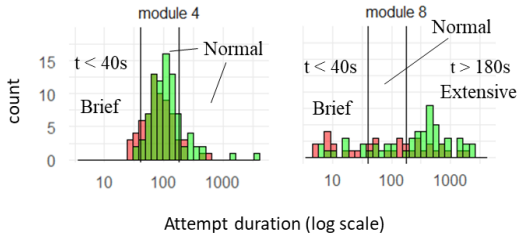


FIG 2: Two examples of AC attempt duration distribution plotted on a log scale. Passing attempts are plotted in green and overlaid on top of failed attempts plotted in red. The vertical lines are drawn at $t = 40$ s and $t = 180$ s respectively.

Finally, when examining the distribution of duration for students' AC attempt after studying the IC, we found that on modules 1-6, the distribution roughly follows a single log-normal distribution, but on modules 7-10, there is a separate sub-group of students who spent > 180 s on their AC attempt, as shown in the two examples in FIG 2. In addition, on modules 7, 8, and 10, students spending > 180 s have a higher percentage of correct responses than those who spent < 180 s. A possible explanation is that the AC of modules 7-10 contains complex calculation problems, whereas the ACs of modules 1-6 contains either simple calculation problems

or conceptual problems. Therefore, we categorized student AC attempts on modules 7-10 after studying the IC as "Extensive" if the attempt duration is > 180 s.

Combinations of the "Brief," "Normal," and "Extensive" labels for each attempt and the four general interaction categories resulted in 9 interaction states as listed in Table 1. The ordering of those states roughly reflects the level of effort in interacting with the module, with state 8 being the highest and state 0 being the lowest. In addition, neighboring states are more similar in problem-solving duration than distant states. One exception is that students in state 4 (initial pass with normal attempt duration) likely spent less effort than those in states 2 and 3 by passing the AC in their initial attempt and skipping the IC. We ranked it as state 4 because those students likely have a higher level of incoming knowledge on the subject, and are qualitatively very different from students in state 1, who passed the AC in a Brief attempt, and are more likely to have guessed in their attempt.

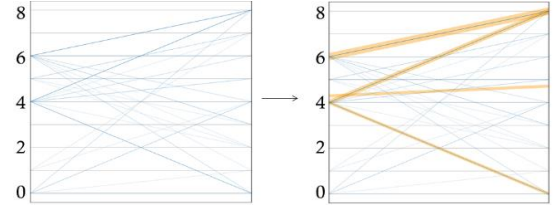


FIG 3: Transition paths for every student between two modules (left) and highlighted major transition paths after clustering algorithm (right).

B. Visualization of transition between interaction states on adjacent modules

Each student's transition from one interaction state to another between two adjacent modules can be visualized by a path on a parallel coordinates graph, as illustrated in FIG 3. The left axis represents student states for the first module and the right axis represents the state for the next module. A horizontal path indicates that the student remained in the same interaction state on two consecutive modules.

Table 1: List of 9 interaction states categorizing students' interaction with each OLM. * States 7 and 8 are only applicable to modules 7 - 10

State	Attempt Duration	Interaction Category
8*	Extensive	II: Pass after study
7*	Extensive	III: Fail after study
6	Normal	II: Pass after study
5	Normal	III: Fail after study
4	Normal	I: Initial pass
3	Brief	II: Pass after study
2	Brief	III: Fail after study
1	Brief	I: Initial Pass
0	N/A	IV: Multiple attempts without study

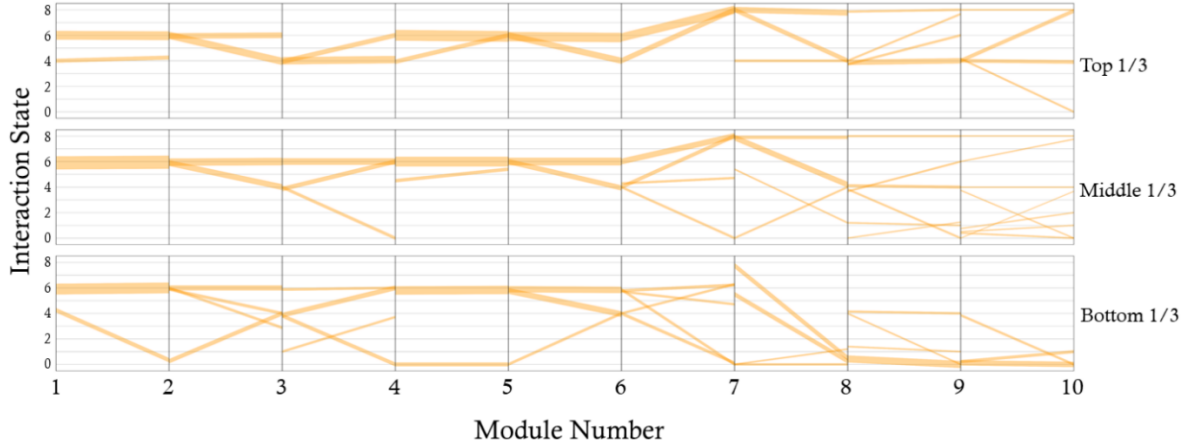


FIG 4: Parallel coordinates graph of all major interaction state transition paths for the top, middle, and bottom cohorts. Note that states 7 and 8 are only applicable to module 7 – 10.

The student population is divided into top, middle and bottom cohorts according to their final course credit, which includes homework, classroom activities, labs, and exams. The OLM homework accounts for less than 2% of the of the total grade.

To identify the major transition paths for a given student population between any two adjacent modules, an agglomerative hierarchical clustering algorithm was employed [23]. Complete-link clustering was performed on the Euclidian distance between students, using their interaction state on every pair of adjacent modules as location variables. Each cluster, identified at a cut-height of 0.02, can represent either a single path or an average between two or more neighboring paths. The most populated clusters accounting for up to half of the population in a given cohort are highlighted by a yellow line, as shown in FIG 3 and FIG 4. The width of each line is proportional to the fraction of students in the cohort that belongs to this cluster. When a cluster contains more than one path, the start and end point of the highlighted path is the average of the paths in the cluster, weighted by the population of each path.

Data analysis, hierarchical clustering, and statistical testing were conducted using R [24] and the tidyverse package [25], while the visualization was created using Python’s Matplotlib [26].

III. RESULTS

The major transition paths of interaction states across the 10 modules for the top, middle, and bottom cohorts are plotted in FIG 4. Individual paths were hidden for clarity. For all cohorts, most major paths start and end on or close to four interaction states: 0, 4, 6, and 8. Only a few paths start or end between two neighboring states.

There are some remarkable differences between the top and bottom cohorts in terms of their major paths between

adjacent modules. On one hand, the top cohort mostly transitioned between states 4 and 6 for modules 1-6, and transitioned between states 8 and 4 for modules 7 -10. Only on the last two modules did a major transition pathway end on state 0.

On the other hand, for the bottom cohort, there are multiple major paths that start or end on state 0 across all 10 modules. In particular, on the last three modules, the horizontal path that starts and ends both on state 0 is the most populated path, following a significant shift from states 8, 6, and 5 to states 0 and 1 on module 7. In addition, no major path passed through either state 8 or state 6 on the last three modules.

The transition patterns for the middle cohort seem to be a mix between the top and bottom cohort, with many more pathways identified for the last three modules.

Table 2: p values for Fisher’s Exact test conducted on the correlation between interaction states and student cohorts. Values < 0.01 are highlighted by two asterisks.

Module	p-values	Adjusted p-values
1	0.063	0.31
2	0.33	0.67
3	0.20	0.50
4	< 0.01**	< 0.01**
5	0.14	0.41
6	0.98	0.98
7	< 0.01**	< 0.01**
8	< 0.01**	< 0.01**
9	< 0.01**	< 0.01**
10	< 0.01**	< 0.01**

The difference in the interaction states occupied by the three cohorts on each module can be verified by conducting an extended Fisher’s Exact Test [27–29] using the 8 states as dependent variables. To account for Type 1 errors caused by conducting multiple statistical tests, Hommel corrections were applied to the p-values [30], the results of which are

shown in Table 2. At $\alpha = 0.01$ level, interaction states occupied by the three cohorts are significantly different on modules 4, 7, 8, 9, and 10, according to adjusted p -values.

IV. DISCUSSION

We visualized major patterns in students' change in interaction states as they proceed through 10 OLM modules. Under the current classification of interaction states, we saw significant differences between the top and bottom cohorts.

The interaction states of the top third cohort are remarkably stable across all 10 modules, transitioning between states 6 or 8, which are passing the module on either a normal or extensive attempt after studying the IC, and state 4 which is passing the module on a normal initial attempt before studying the IC. This pattern is consistent with the intended behavior in a mastery learning design, in which students determine their learning effort based on their own level of mastery.

In contrast, the behavior of the bottom cohort is different in two ways. First, on modules 1-6, there were multiple pathways starting or ending on state 0 or between states 0 and 1. Second, on the last three modules, no major transition paths involved states 5-8, and the most populated paths started and ended on state 0 or between states 0 and 1. Given that state 0 is making more than 2 failed attempts before accessing the IC, and state 1 is initial passing on a Brief attempt, this pattern suggests that students in this cohort are less motivated to study the IC. One possible reason is that some of those students started working on the modules closer to the due date, and thus had insufficient time to study the IC, especially on the last few modules. This explanation can be easily verified in the future by linking the current data set with students' start times for each module. Another possible explanation is that some students in the bottom cohort had less success in learning from the IC on earlier modules, and were discouraged on later modules. This could explain why a major shift toward state 0 or 1 took place between modules 7 and 8, since module 7 had the lowest passing percentage of all modules.

Those two possible explanations suggest two potential strategies for helping struggling students persist in productive learning behavior. First, one could introduce mechanisms such as extra credit for early completion to encourage students to start early. Second, one could improve the instructional quality of module 7, or add an easier module between modules 6 and 7. Both strategies are currently being developed for future implementations of the modules.

Finally, the behavior pattern of the middle cohort is more similar to the top cohort on the first six modules, with fewer pathways involving state 4. This suggests that those students have weaker incoming knowledge but were equally successful in learning from the IC. The paths on the last few modules resembled a mix of the top and the bottom cohort, indicating that the behavior of this cohort becomes more diverse towards the end of the sequence.

V. CONCLUSION AND FUTURE DIRECTIONS

The current study shows that the difference in learning behavior between high and low performing students can become significantly larger on more challenging contents near the end of an OLM sequence. The results also suggested detailed strategies that could remediate the behavioral gap between top and bottom students.

As an exploratory effort, the current analysis also leaves several major caveats that can be examined and addressed in more extensive future follow up studies.

First of all, the current categorization scheme focused on the duration of AC attempts, while other important aspects of learning behavior such as duration of study on the IC were either simplified or neglected. Future studies should include more aspects of student learning behavior to gain a more comprehensive understanding of the relation between learning behavior and performance. In addition, future studies involving more modules could also track students' transition in learning behavior over longer periods on time, such as over an entire semester.

Second, the ordering of the interaction states in this study grouped attempts with similar durations closer than those with similar outcomes. While the grouping does not affect the statistical test outcomes, it does impact the clustering outcomes as well as the visual representation. How the results would change under different ordering of states needs to be examined carefully in more extensive follow up analyses.

Third, a number of decisions such as using 40s as the cutoff between Brief and Normal attempts or dividing the population into three cohorts were made in order to categorize complex student behavior into a reasonable amount of interaction states. How each of those decisions impact the outcome can be tested in multiple future studies, as well as investigating whether there are more "natural" cutoffs between different student populations or different types of learning behavior.

Finally, the current analysis cannot identify one kind of state that could potentially be highly populated: if students transitioned from multiple different states on a previous module into a single state on the current module, then in turn transitioned from that state to multiple different states on the subsequent module, then that single "hub" state in the current module will not be highlighted, as it is not involved in a major path. Whether those type of states exist and how different cohorts populate such states could also serve as an interesting topic for future investigations.

ACKNOWLEDGMENTS

The authors would like to thank the Learning Systems and Technology team at UCF for developing the Obojobo platform. This research is partly supported by NSF Award No. DUE-1845436 and the Advancement of Early Career Researchers (AECR) Program at the University of Central Florida.

- [1] D.T. Seaton, Y. Bergner, G. Kortemeyer, S. Rayyan, I. Chuang, D.E. Pritchard, The Impact of Course Structure on eText Use in Large-Lecture Introductory-Physics Courses, 2013 Phys. Educ. Res. Conf. Proc. (2014) 333–336. doi:10.1119/perc.2013.pr.071.
- [2] D.T. Seaton, Y. Bergner, I. Chuang, P. Mitros, D.E. Pritchard, Who does what in a massive open online course?, Commun. ACM. 57 (2014) 58–65. doi:10.1145/2500876.
- [3] S.Y. Lin, J.M. Aiken, D.T. Seaton, S.S. Douglas, E.F. Greco, B.D. Thoms, M.F. Schatz, Exploring physics students' engagement with online instructional videos in an introductory mechanics course, Phys. Rev. Phys. Educ. Res. 13 (2017) 1–18. doi:10.1103/PhysRevPhysEducRes.13.020138.
- [4] C.R. Bowman, O. Gulacar, D.B. King, Predicting Student Success via Online Homework Usage, J. Learn. Des. 7 (2014). doi:10.5204/jld.v7i2.201.
- [5] J. Guerra, S. Sahebi, P. Brusilovsky, Y. Lin, The Problem Solving Genome : Analyzing Sequential Patterns of Student Work with Parameterized Exercises, Proc. 7th Int. Conf. Educ. Data Min. (2014) 153–160.
- [6] R.E. Teodorescu, D.T. Seaton, C.N. Cardamone, S. Rayyan, J.E. Abbott, A. Barrantes, A. Pawl, D.E. Pritchard, N.S. Rebello, P. V. Engelhardt, C. Singh, When students can choose easy, medium, or hard homework problems, in: Phys. Educ. Res. Conf., 2011: pp. 81–84. doi:10.1063/1.3679999.
- [7] G. Alexandron, J. Antonio Ruiperez Valiente, Z. Chen, D.E. Pritchard, Using Multiple Accounts for Harvesting Solutions in MOOCs, in: Proc. Third ACM Conf. Learn. @ Scale, Edinburg, Scotland UK, 2016: pp. 63–70. doi:10.1145/2876034.2876037.
- [8] K.R. Koedinger, E.A. McLaughlin, J. Kim, J. Zhuxin Jia, N.L. Bier, J.Z. Jia, E.A. McLaughlin, N.L. Bier, Learning is Not a Spectator Sport, in: Proc. Second ACM Conf. Learn. @ Scale - L@S '15, ACM Press, New York, New York, USA, 2015: pp. 111–120. doi:10.1145/2724660.2724681.
- [9] D. Gašević, S. Dawson, G. Siemens, Let ' s not forget : Learning analytics are about learning, TechTrends. 59 (2015) 64–71. doi:10.1007/s11528-014-0822-x.
- [10] G. Kortemeyer, Correlations between student discussion behavior, attitudes, and learning, Phys. Rev. Spec. Top. - Phys. Educ. Res. 3 (2007) 1–8. doi:10.1103/PhysRevSTPER.3.010101.
- [11] G. Kortemeyer, Work Habits of Students in Traditional and Online Sections of an Introductory Physics Course: A Case Study, J. Sci. Educ. Technol. 25 (2016) 697–703. doi:10.1007/s10956-016-9624-6.
- [12] R. Martinez-Maldonado, K. Yacef, J. Kay, A. Kharrufa, A. Al-Qaraghuli, Analysing Frequent Sequential Patterns of Collaborative Learning Activity Around an Interactive Tabletop., in: EDM 2011 - Proc. 4th Int. Conf. Educ. Data Min., 2011: pp. 111–120. http://educationaldatamining.org/EDM2011/wp-content/uploads/proc/edm2011_paper36_full_Martinez.pdf.
- [13] Z. Chen, K.M. Whitcomb, C. Singh, Measuring the effectiveness of online problem-solving tutorials by multi-level knowledge transfer, in: 2018 Phys. Educ. Res. Conf. Proc., American Association of Physics Teachers, 2019: pp. 1–4. doi:10.1119/perc.2018.pr.Chen.
- [14] Z. Chen, G. Garrido, Z. Berry, I. Turgeon, F. Yonekura, Designing online learning modules to conduct pre- and post-testing at high frequency, in: 2017 Phys. Educ. Res. Conf. Proc., American Association of Physics Teachers, Cincinnati, OH, 2018: pp. 84–87. doi:10.1119/perc.2017.pr.016.
- [15] Z. Chen, S. Lee, G. Garrido, Re-designing the Structure of Online Courses to Empower Educational Data Mining, in: K. Elizabeth Boyer, M. Yudelson (Eds.), Proc. 11th Int. Educ. Data Min. Conf., Buffalo, NY, 2018: pp. 390–396.
- [16] B.S. Bloom, Learning for Mastery, UCLA Eval. Comment. 1 (1968).
- [17] F.S. Keller, "Good-bye, teacher...", J. Appl. Behav. Anal. (1968) 79–89. doi:10.1901/jaba.1968.1-79.
- [18] C.-L.C. Kulik, J.A. Kulik, R.L. Bangert-Drowns, Effectiveness of Mastery Learning Programs: A Meta-Analysis, Rev. Educ. Res. 60 (1990) 265–299. doi:10.3102/00346543060002265.
- [19] Center for Distributed Learning University of Central Florida, Obojobo, (n.d.). obojobo.com.
- [20] G. Alexandron, J.A. Ruiperez-Valiente, Z. Chen, Pedro J. Muñoz-Merino, D.E. Pritchard, Copying @ Scale: Using Harvesting Accounts for Collecting Correct Answers in a MOOC, Comput. Educ. 108 (2017). doi:10.1016/j.compedu.2017.01.015.
- [21] R. Warnakulasooriya, D.J. Palazzo, D.E. Pritchard, Time to completion of web-based physics problems with tutoring., J. Exp. Anal. Behav. 88 (2007) 103–113. doi:10.1901/jeab.2007.70-06.
- [22] C.L. Barry, S.J. Horst, S.J. Finney, A.R. Brown, J.P. Kopp, Do Examinees Have Similar Test-Taking Effort? A High-Stakes Question for Low-Stakes Testing, Int. J. Test. 10 (2010) 342–363. doi:10.1080/15305058.2010.508569.
- [23] L. Rokach, O. Maimon, Clustering Methods, in: O. Maimon, L. Rokach (Eds.), Data Min. Knowl. Discov. Handb., Springer US, Boston, MA, 2005: pp. 321–352. doi:10.1007/0-387-25465-X_15.
- [24] R Core Team, R: A Language and Environment for Statistical Computing, Doc. Free. Available Internet [Http://Www. r-Project. Org.](http://www.r-project.org) (2019). <https://www.r-project.org/>.
- [25] H. Wickham, tidyverse: Easily Install and Load the "Tidyverse," (2017). <https://cran.r-project.org/package=tidyverse>.
- [26] J.D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. (2007). doi:10.1109/MCSE.2007.55.
- [27] C.R. Mehta, N.R. Patel, A network algorithm for performing fisher's exact test in $r \times c$ contingency tables, J. Am. Stat. Assoc. (1983). doi:10.1080/01621459.1983.10477989.
- [28] C.R. Mehta, N.R. Patel, ALGORITHM 643: FEXACT: a FORTRAN subroutine for Fisher's exact test on unordered $r \times c$ contingency tables, ACM Trans. Math. Softw. (1986). doi:10.1145/6497.214326.
- [29] D.B. Clarkson, Y. Fan, H. Joe, A remark on algorithm

643: FEXACT: an algorithm for performing Fisher's exact test in $r \times c$ contingency tables, ACM Trans. Math. Softw. (1993). doi:10.1145/168173.168412.

- [30] G. Hommel, A stagewise rejective multiple test procedure based on a modified bonferroni test, Biometrika. (1988). doi:10.1093/biomet/75.2.383.