Comparing Measures of Student Performance in Hybrid and MOOC Physics Courses

Trevor A. BALINT[1]
Raluca TEODORESCU[2]
Kimberly COLVIN[3]
Youn-Jeng CHOI[4]
David E. PRITCHARD[5]

[1]Department of Physics, George Washington University, Washington, DC
[2] Science, Engineering and Technology Unit, Montgomery College, Takoma Park, MD
[3]Department of Educational and Counseling Psychology, University at Albany SUNY, Albany, NY
[4]Department of Educational Studies in Psychology, Research Methodology and Counseling, University of Alabama, Tuscaloosa, AL
[5]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA

(Received: 28.11.2015, Accepted: 27.12.2015)

Abstract
In this paper we use seven quantitative measures of student performance to compare the performance of students enrolled in three physics courses (two hybrid and one MOOC) that have some common features. We find that, despite the fact that these courses have different audiences, aims, and methods, the measures presented here place the students from all three courses on the same scale and reveal performance similarities. All measures are compared pairwise and the sign of the correlation between each pair is consistent for all courses. The percentage-based measures all positively correlate with each other and with Item Response Theory measure, while the measures based on average number of submissions positively correlate together but anti-correlate with some percent-based and IRT measures. Our findings suggest that for all course types students who get a higher fraction of problems correct tend to use fewer submissions to do so and have a higher IRT skill, while students in a MOOC choose more frequently to not attempt a problem upon opening it than students enrolled in hybrid courses.
Keywords: Problem solving, Data mining, Student performance

INTRODUCTION

With the advent of online Course Management Systems (CMSs), much richer data are available now than ever before for instructors and researchers. This allows for more in-depth analysis of how students behave when they solve physics problems. However, determining how successful students are in our courses and how successful our courses are at teaching students are complex tasks that are often correlated with final course grades (Beichner, Saul, & Abbott, 2007; Beuckman, Rebello, & Zollman, 2007; S. W. Bonham, Deardorff, & Beichner, 2003; Breslow et al., 2013; Cheng, Thacker, Cardenas, & Crouch, 2004; J. Docktor, Heller, Henderson, Sabella, & Hsu, 2008; J. L. Docktor & Mestre, 2014; Finkelstein & Pollock, 2005; S. J. Pollock, 2009; Sadler & Tai, 2001); exam grades (S. W. Bonham et al., 2003; Deslauriers, Schelew, & Wieman, 2011; J. Docktor et al., 2008; J. L. Docktor & Mestre, 2014; Mestre, Hart, Rath, & Dufresne, 2002); or FCI, FMCE, or other research-validated assessment scores (Bao et al., 2009; Beichner et al., 2007; S. W. Bonham et al., 2003; Cheng et al., 2004; J. Docktor et al., 2008; J. L. Docktor & Mestre, 2014; Finkelstein & Pollock, 2005; Harper, Etkina, & Lin, 2003; Henderson, 2002; Perkins, 2005; S. J. Pollock, 2009; Steven J. Pollock & Finkelstein, 2008; Taasoobshirazi & Sinatra, 2011). Research points out that analyses based solely on these measurements "provide limited and sometimes misleading information about student learning." (Fraser et al., 2014) In this paper, we show how data available in CMSs can supplement and enrich the existing methods. We use data from three physics courses offered on two CMSs (LONCAPA (Kortemeyer et al., 2003, 2008; "LON-CAPA," n.d.) and edX ("edX," n.d.; S. Rayyan, Seaton, Belcher, Pritchard, & Chuang, 2013)) to compare seven measures of student success and discuss their relationships and what they may say about student performance. We also use the variety of course types for which we have available data to explore the differences between student performances in different types of courses.

As CMSs are in use in a variety of different course types, we perform this analysis for a Hybrid course that has in-person lecture and lab sessions with online-only homework, a Hybrid course in a flipped classroom with no labs and online-only homework, and for a Massive Open Online Course (MOOC) that is entirely online. This not only gives us a way to examine our measures of success in a variety of contexts but it also allows us to look at the differences in skill present between these types of courses. All the courses considered in this study are introductory mechanics courses and their details are presented below.

Course A: Hybrid Algebra-Based Course. The first course we have analyzed is an introductory algebra-based physics course offered in 2008 at a research university. Split across two sections, this course had 208 students finish at least one homework assignment, with the 187 of those who passed the course being the focus of our analysis. The course had peer-instruction lecture, lab, and recitation sessions, as well as in-person class quizzes and exams, but the homework was administered and graded using the CMS LON-CAPA.

Course B: Hybrid Calculus-Based Course. The second course we have analyzed is a calculus-based physics course offered at a top research university. This course is offered every spring, for students who earned less than a C in the similar prior course offered in the fall semester. We use data from the Spring 2013 offering of the course. This course used the MAPS pedagogy (Pawl, Barrantes, & Pritchard, 2009; Saif Rayyan, Pawl, Barrantes, Teodorescu, & Pritchard, 2010) which was proven to improve students' learning attitudes about science, performance on the course material, and future physics grades. The grading policy was flexible, students having the freedom to earn "homework" points by choosing to solve easy, medium, and hard problems that have certain points associated with them – the higher the difficulty of the problem, the higher the number of points associated with it. Consequently, the percent of students attempting the online homework problems was variable. From 47 students enrolled in the course, only 35 completed over 50% of available items and are the subject of our analysis.

Course C: MOOC - 8.MReVx. The third course we have analyzed is the 2013 offering of the introductory physics MOOC 8.MReVx designed by the RELATE (REsearch in Learning Assessing and Tutoring Effectively) Group ("RELATE | Research in Learning, Assessing and Tutoring Effectively," n.d.) at the Massachusetts Institute of Technology and offered on the edX platform. ("Mechanics Review," n.d.) Our analysis focuses on 1,080 participants (out of 16,787 enrolled in the course) who attempted more than 50% of available problems, with a certificate of completion being offered to students who achieved more than 60% of available points

(N=1,033). This course used the same physics items as Course B and thus provides an excellent comparison to the performance of students in a hybrid course.

Table 1 has a comparison of the courses in order to highlight some of their pertinent similarities and differences.

Previous Research

A previous study comparing the students' performance of Courses B and C found that students in the MOOC had a higher average skill than students in the on-campus course.(K. Colvin, Champaign, & Liu, 2014)

Table 1. *Information about each course*

	Course A	Course B	Course C
Course size	187	35	1033
HW Format	Online – End-of-unit problems	Online – End-of-unit problems	Online – Both end-of-unit and conceptual mid-unit problems
Lecture Format	Live given by a single instructor	Flipped Classroom	Lecture videos, with and without a narrator
Labs Format	Traditional – Small groups led by a TA	None	Digital – In the style of PhET
Exams	Traditional – In-person during regular class time	Traditional – In-person during regular class time	Optional midterm and final exam
Topics & Order	Forces, kinematics, circular motion, energy, momentum, fluids, and waves	Forces, kinematics, momentum, energy, rotational motion, and oscillations	Forces, kinematics, circular motion, energy, momentum, and rotational motion

Other previous research by Lieberman et al(Dubson, Johnsen, Lieberman, Olsen, & Finkelstein, 2014; Lieberman, Dubson, Johnsen, Olsen, & Finkelstein, 2014) has found that a MOOC designed with exactly the same materials as an offline course had no worse educational outcomes than the offline course when comparing the measures of various standardized exams and the assignments given to each group of students. Similar research by Konstan et al(Konstan, Walker, Brooks, Brown, & Ekstrand, 2015) has found that in another identical MOOC/offline course pair the primary predictor of success was effort put into the course and not the type of course taken. An additional study found that student performance on online homework is no worse than performance on homework collected and graded by hand.(S. Bonham, Beichner, & Carolina, 2001)

Measures of Student Success

Using the data commonly available in a CMS, we define several measures of success - some of them can be determined with data available even from a traditional, fully offline course, while others can only be obtained in a CMS. In a CMS, students are typically given multiple attempts for solving a given item. For all three courses analyzed here, a student is considered to have gotten the right answer if he or she did so on any attempt made on the item.

We define an "item" as an individual question with a single response field that may or may not be embedded within a larger problem. The response field may be free-response, radio buttons, checkboxes, or a field for the student to draw in, but each item only has one response field.

The first measure we define is simplest and most closely related to the student's grade in the course. It measures how many items were solved correctly out of how many were offered, called by us "Percent Correct" or "% Corr":

$$\%\text{ Corr} = \frac{\text{Number of Items Solved Correctly}}{\text{Number of Items In Course}} \tag{1}$$

This measure can be easily calculated in a traditional course, and as such it and its relationships to the other measures may be of particular interest to instructors.

The second measure is calculated from the number of items students solved out of how many items they opened, and is called "Percent Correct of Accessed" or "% Corr Access":

$$\%\text{ Corr Access} = \frac{\text{Number of Items Solved Correctly}}{\text{Number of Items Accessed}} \tag{2}$$

This measure can only be available in CMSs. It has the potential to discriminate between different types of learners; for example, students with a low score on this metric may belong to Shell's profile of the "surface learning" student or "apathetic" student. (Shell & Husman, n.d.; Shell & Soh, 2013)

The next measure is defined as the number of items the student solved incorrectly out of how many items they tried, called "Percent Wrong of Attempted" or "% Wrong Att":

$$\%\text{ Wrong Att} = \frac{\text{\# of Items Attempted But Not Solved Correctly}}{\text{Number of Items Attempted}} \tag{3}$$

Though of course students have a wide range of values of this measure, students in traditional courses tend to attempt every available item since every item is graded and adds to the student's final score, even if they may not get the problem correct. This differs from a MOOC where students can open a problem and decide not to attempt the problem because the course is graded on a pass/fail scale, (Ghadiri, Qayoumi, & Junn, 2013; JO, PARK, KIM, & SONG, 2014; Konstan et al., 2015; Patterson, 2014) causing this measure to be lower for these MOOC students.

The following two measures involve multiple attempts to correctly solve an item and are only available for online homework, which allows for multiple submissions while traditional homework does not. (Kortemeyer, 2014a) The first measure is the average number of tries per correct submission named "Subm Per Correct Item":

$$\text{Subm Per Corr} = \frac{\text{\# of Subm Used on Corr. Answered Items}}{\text{\# of Correctly Answered Items}} \tag{4}$$

and the second is a measure of the average number of tries out of all attempted items named "Submissions Per Attempted Item":

$$\text{Subm Per Item} = \frac{\text{Number of Submissions Used}}{\text{Number of Items Attempted}} \tag{5}$$

These measures begin to determine how skilled a student may actually be, since the most skilled students are expected to have low values on both of these measures as they do not require as many submissions to get a correct answer. Students who only attempt problems that they expect to be able to answer correctly will likely have very close values for these two measures, while students who attempt every problem regardless of their skill may have different results. (Ghadiri et al., 2013; JO et al., 2014; Konstan et al., 2015; Patterson, 2014) In the courses considered, the numbers of submissions allowed on homework items range from 3 to 10; for quiz and assessment items, this range is lower (around 2 to 4).

Another measure we consider here is the students' ability determined by a Two-Parameter Logistic Item Response Theory (IRT). (Bergner et al., 2012; K. F. Colvin et al., 2014; Kortemeyer, 2014b) These skills together with the IRT question's difficulty can be used to determine the probability that a student will correctly answer an item. The student skill is constrained to be a normal distribution with a mean of 0 and a standard deviation of 1. For the purposes of this study, the students from Course B were put on the same IRT scale as those from 8.MReVx using the anchor item IRT equating method. This was possible as the two courses shared a majority of items.

In the follow up sections, tables and figures we will use the labels we have introduced above to help the reader follow our work.

RESULTS

Figures 1-3 show the correlation matrices for the student body on each pair of measures for Course A, Course B, and the MOOC (Course C) respectively. The values in the top right of each grid of graphs show the Pearson r correlation coefficient (Fraenkel, Wallen, & Hyun, 2014) of that pair of measures. For example, in Fig. 1 r = 0.236 corresponds to the comparison of % Wrong Att and Subm Per Corr, while in Fig. 3 the r-value of 0.942 is for Subm Per Corr vs. Subm Per Item. The possible values for r vary from -1 (perfect anti-correlation) to 1 (perfect correlation), and the standard error values for a given r-value (discussed in the caption for each figure) are determined using Eq. 7, where n is the sample size: (Bowley, 1928; Efron, 1979; Ellison, 2006; Park, EunsikLee, 2001; Shieh, 2010)

$$\sigma = \frac{1 - r^2}{\sqrt{n-2}} \tag{6}$$

An important result is that all three courses show the same sign of the correlation and general behavior of the pair's relationship for every pair of measures. The percentage correct measures both positively correlate with each other and with IRT, while the two average submission measures and the percentage wrong measure positively correlate together but negatively correlate with the two percent correct and the IRT measures. These results in general imply that, for our students, those who get a higher fraction of problems correct will use fewer submissions to do so and have a higher IRT skill. Additionally, students who tend to get low fractions of problems correct tend to have low IRT skill and use a higher number of submissions to do so. This is true for all three courses examined, and is independent of course type.

A numerical comparison of the three courses is provided in Table 2, which gives the values of the divisions between quartiles for each measurement for each course. Note that the distributions of students are skewed with a narrower distribution of high-skill students near the limit of each measure and a wide distribution of low-skill students.

We now choose several examples to illustrate how the measures we introduced can help instructors better understand some aspects of student performance. A strongly correlated pair of measures in Course A and 8.MReVx and weakly correlated in Course B is % Corr vs. % Corr Access with r-values of 0.722 in Course A, 0.298 in Course B, and 0.807 in 8.MReVx. We attribute this low correlation to the differences in grading standards between the MOOC and the hybrid courses and discuss this difference further below.

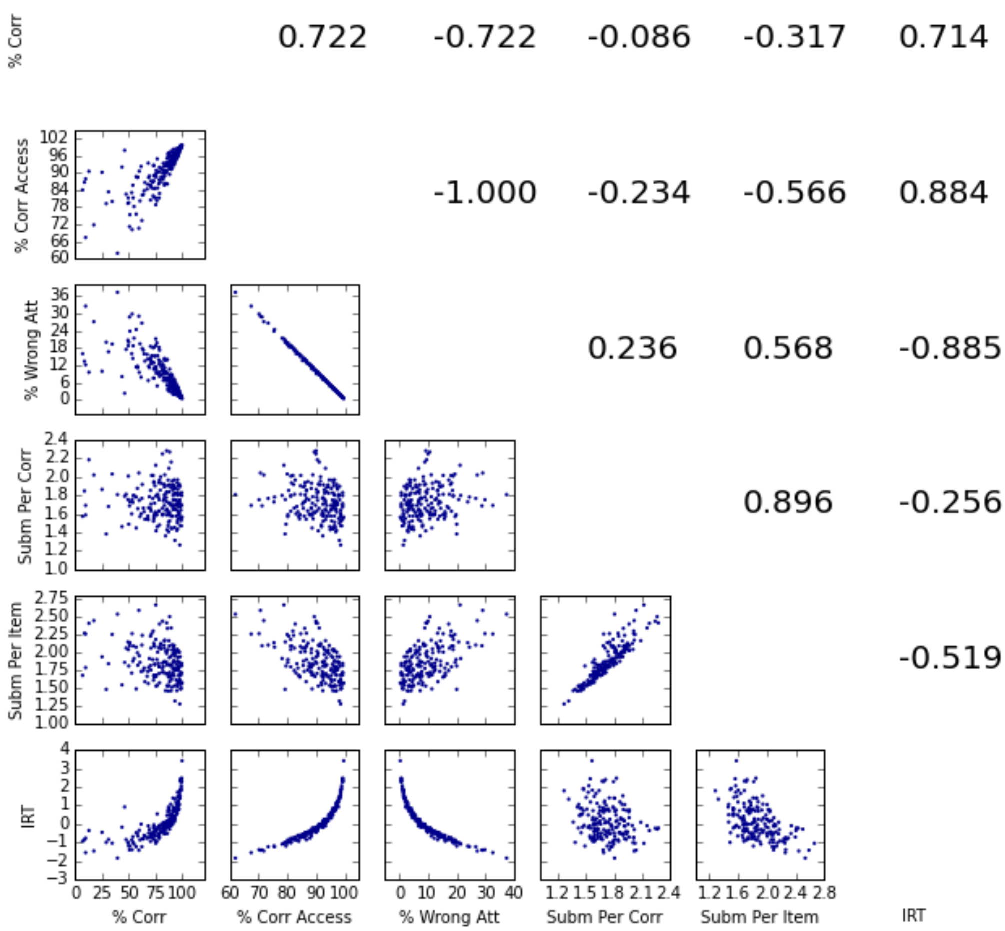

Figure 1. Distributions of the student body across all measures for Course A (N=187). Error values vary from 0.01 to 0.07 for |r|=0.897 to |r|=0.086 respectively.

An anti-correlated pair of measures that appears in all courses is % Corr Access and Subm Per Corr with values of r = - 0.234, r = - 0.558, and r = - 0.673 in Course A, Course B, and the MOOC respectively. This anti-correlation implies that students having a lower average number of submissions on correctly answered items tend to have a higher percentage correct of problems that they attempted. The most striking result from these comparisons of measures lies not in the correlation coefficients for a single course but the consistency of the distributions of the pairs of measures across the three courses.

Another small but still significant difference lies in the relationship between % Wrong Att and % Corr Access. In the Hybrid courses, these two measures have a strict correlation of r = -1.000. In 8.MReVx these measures have a correlation of r = -0.921 which, while still high, is not a strict correlation. We attribute this difference (although small) to the pass/fail nature of the online course allowing the students to decide not to solve a problem, whereas that is not a choice that many students make in the traditional courses as every problem counts towards a students' final grade.

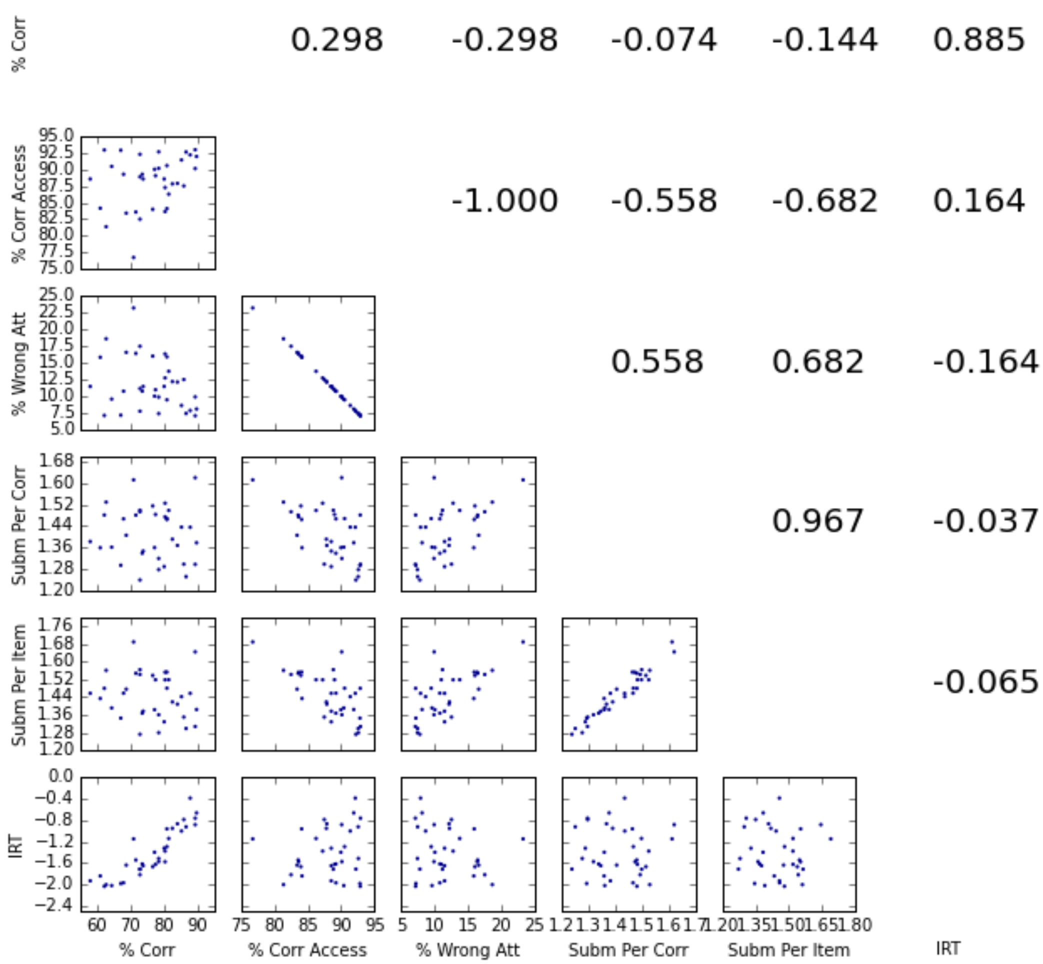

Figure 2. *Distributions of the student body across all measures for Course B (N=35). Error values vary from 0.01 to 0.17 for |r|=0.967 to |r|=0.074 respectively.*

***Table 2**. Statistics for each measure of student success for each course. The three numbers given are the divisions between quartiles; for example, 82.5% is the cutoff between the highest and second-highest quartiles in Course B in Percent All. We analyze these statistics rather than the mean/median and standard deviation because the distributions of students on all of these measures are skewed: the distribution of low-skill students is typically wider than the distribution of high-skill students.*

Measures of student success	Course A (N=187)			Course B (N=35)			8.MReVx (N=1080)		
	Q_1/Q_2	Q_2/Q_3	Q_3/Q_4	Q_1/Q_2	Q_2/Q_3	Q_3/Q_4	Q_1/Q_2	Q_2/Q_3	Q_3/Q_4
% Corr	70.7%	85.0%	92.7%	70.9%	77.5%	82.5%	61.5%	71.9%	84.2%
% Corr Access	86.3%	92.2%	96.4%	84.2%	88.8%	91.4%	71.8%	81.6%	87.7%
% Wrong Att	1.7%	7.6%	13.6%	8.1%	10.8%	12.5%	3.9%	8.18%	15.0%
Subm per Corr	1.59	1.71	1.85	1.33	1.41	1.48	1.17	1.24	1.33
Subm per Item	1.68	1.83	2.04	1.37	1.46	1.54	1.20	1.30	1.42

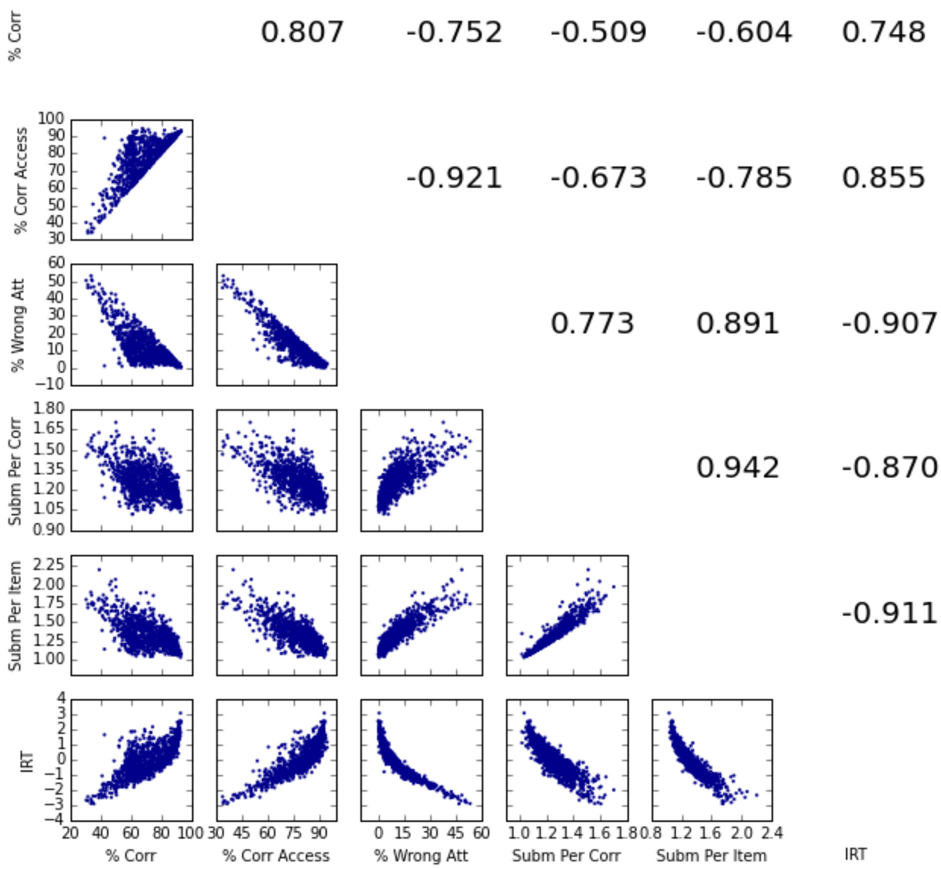

Figure 3. Distributions of the student body across all measures for 8.MReVx (N=1080). Error values vary from 0.003 to 0.02 for |r|=0.942 to |r|=0.509 respectively.

An additional percentage-based measure was chosen and analyzed but not shown in the above tables. This measure was chosen as a way to determine the fraction of problems that a student opened and decided to attempt, called "Percent Attempted" or "% Att":

$$\%\ \text{Att} = \frac{\text{Number of Items Attempted}}{\text{Number of Items Accessed}} \tag{7}$$

This measure is interesting because it is the only measure presented in this study that discriminates between types of courses. For both traditional courses this measure is >98% for every student, while for the MOOC this measure has a wide distribution that is shown (with a comparison to % Corr) in Fig. 4. This finding suggests that students in a MOOC choose more frequently to not attempt a problem upon opening it. It is unknown whether that is a consequence of the online nature of the course or its pass/fail method of awarding course certificates.

Despite the above-mentioned differences, we see that though these courses have significantly different audiences, aims, and methods, the measures presented here place the students from all three courses on the same scale and reveal similar distributions of students across all measures, even for a course with extremely low statistics. We believe that this shows the robustness of the measures and the correlations between measures, and the applicability of this analysis outside of the course types discussed here.

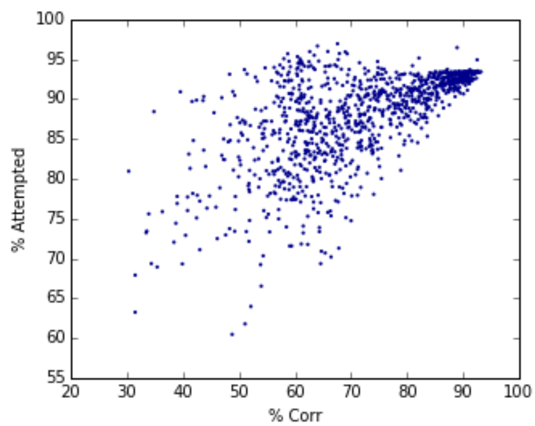

Figure 4. *Percent Attempted vs Percent Correct for all students in the MOOC (Course C). r=0.6276±0.018*

CONCLUSIONS

Through the analysis of problem performance statistics already available to Physics instructors, we have found deep connections between various simple measures of student success. We also see the relationships between these simple measures and more complex ones such as average number of attempts used when submitting homework answers and Item Response Theory. These measures are consistent across courses teaching material of varying levels to groups of students with varying levels of skill on various platforms. We see that correlations between pairs of these measures have a larger error in courses with a smaller population, as expected, but that the correlations have consistent signs between all three courses.

We have documented several measures of performance that are simple to calculate and identified which have strong correlations or anti-correlations with IRT skill. We have also shown that there are distinctions between the performances of students in fully online vs. hybrid online-offline courses that should be anticipated by instructors moving their courses into the online realm.

REFERENCES

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., … Wu, N. (2009). Physics. Learning and scientific reasoning. *Science (New York, N.Y.)*, *323*(5914), 586–7. http://doi.org/10.1126/science.1167740

Beichner, R., Saul, J., & Abbott, D. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *… -Based Reform of …*. Retrieved from http://www.percentral.com/PER/per_reviews/media/volume1/SCALE-UP-2007.pdf

Bergner, Y., Droschler, S., Kortemeyer, G., Rayyan, S., Seaton, D., & Pritchard, D. E. (2012). Model-Based Collaborative Filtering Analysis of Student Response Data: Machine-Learning Item Response Theory. *International Educational Data Mining Society*, 95–102.

Beuckman, J., Rebello, N. S., & Zollman, D. (2007). Impact of a Classroom Interaction System on Student Learning. In *AIP Conference Proceedings* (Vol. 883, pp. 129–132). AIP. http://doi.org/10.1063/1.2508709

Bonham, S., Beichner, R., & Carolina, N. (2001). Online homework: Does it make a difference? *The Physics Teacher*, *39*(5), 293. http://doi.org/10.1119/1.1375468

Bonham, S. W., Deardorff, D. L., & Beichner, R. J. (2003). Comparison of student performance using web and paper-based homework in college-level physics. *Journal of Research in Science Teaching*, *40*(10), 1050–1071. http://doi.org/10.1002/tea.10120

Bowley, A. L. (1928). The Standard Deviation of the Correlation Coefficient. *Journal of the American Statistical Association*, *23*(161), 31–34.

Breslow, L., Pritchard, D. E., Jennifer, D., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Research & practice in assessment, *8*(March 2012), 13–25.

Cheng, K. K., Thacker, B. A., Cardenas, R. L., & Crouch, C. (2004). Using an online homework system enhances students' learning of physics concepts in an introductory physics course. *American Journal of Physics*, *72*(11), 1447. http://doi.org/10.1119/1.1768555

Colvin, K., Champaign, J., & Liu, A. (2014). Comparing Learning in a MOOC and a Blended On-Campus Course. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 343–344). Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/posters/19_EDM-2014-Poster.pdf

Colvin, K. F., Champaign, J., Liu, A., Zhou, Q., Fredericks, C., & Pritchard, D. E. (2014). Learning in an introductory physics MOOC: All cohorts learn equally, including an on-campus class. *The International Review of Research in Open and Distributed Learning*, *15*(4). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1902/3038

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science (New York, N.Y.)*, *332*(6031), 862–4. http://doi.org/10.1126/science.1201783

Docktor, J., Heller, K., Henderson, C., Sabella, M., & Hsu, L. (2008). Gender Differences in Both Force Concept Inventory and Introductory Physics Performance. In *AIP Conference Proceedings* (pp. 15–18). AIP. http://doi.org/10.1063/1.3021243

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics - Physics Education Research*, *10*(2), 020119. http://doi.org/10.1103/PhysRevSTPER.10.020119

Dubson, M., Johnsen, E., Lieberman, D., Olsen, J., & Finkelstein, N. D. (2014). Apples vs. Oranges: Comparison of Student Performance in a MOOC vs. a Brick-and-Mortar Course. Retrieved August 23, 2015, from http://www.compadre.org/per/items/detail.cfm?ID=13512

edX. (n.d.). Retrieved from https://www.edx.org/

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26. Retrieved from http://projecteuclid.org/euclid.aos/1176344552

Ellison, S. L. R. (2006). In defence of the correlation coefficient. *Accreditation and Quality Assurance*, *11*(3), 146–152. http://doi.org/10.1007/s00769-006-0087-y

Finkelstein, N. D., & Pollock, S. J. (2005). Replicating and understanding successful innovations: Implementing tutorials in introductory physics. *Physical Review Special Topics - Physics Education Research*, *1*(1), 010101. http://doi.org/10.1103/PhysRevSTPER.1.010101

Fraenkel, J., Wallen, N., & Hyun, H. (2014). *How to Design and Evaluate Research in Education* (9th Editio). McGraw-Hill Education.

Fraser, J. M., Timan, A. L., Miller, K., Dowd, J. E., Tucker, L., & Mazur, E. (2014). Teaching and physics education research: bridging the gap. *Reports on Progress in Physics. Physical Society (Great Britain)*, *77*(3), 032401. http://doi.org/10.1088/0034-4885/77/3/032401

Ghadiri, K., Qayoumi, M., & Junn, E. (2013). The transformative potential of blended learning using MIT edX's 6.002 x online MOOC content combined with student team-based learning in class. *Environment*. Retrieved from https://www.edx.org/sites/default/files/upload/ed-tech-paper.pdf

Harper, K. A., Etkina, E., & Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors. *Journal of Research in Science Teaching*, *40*(8), 776–791. http://doi.org/10.1002/tea.10111

Henderson, C. (2002). Common Concerns About the Force Concept Inventory. *The Physics Teacher*, *40*(9), 542. http://doi.org/10.1119/1.1534822

JO, I.-H., PARK, Y., KIM, J., & SONG, J. (2014). Analysis of Online Behavior and Prediction of Learning Performance in Blended Learning Environments. *Educational Technology International*, *15*(2), 71–88. Retrieved from http://www.newnonmun.com/article=188862

Konstan, J. A., Walker, J. D., Brooks, D. C., Brown, K., & Ekstrand, M. D. (2015). Teaching Recommender Systems at Large Scale. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1–23. http://doi.org/10.1145/2728171

Kortemeyer, G. (2014a). An Empirical Study of the Effect of Granting Multiple Tries for Online Homework. Physics Education. Retrieved from http://arxiv.org/abs/1407.2276

Kortemeyer, G. (2014b). Extending item response theory to online homework. *Physical Review Special Topics - Physics Education Research*, *10*(1), 010118. http://doi.org/10.1103/PhysRevSTPER.10.010118

Kortemeyer, G., Albertelli, G., Bauer, W., Berryman, F., Bowers, J., Hall, M., … Speier, C. (2003). *The LearningOnline Network With Computer Assisted Personalized Approach (LON-CAPA)*.

Kortemeyer, G., Kashy, E., Benenson, W., Bauer, W., Kashy, E., & Bauer, W. (2008). Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses. *American Journal of Physics*, *76*(4), 438. http://doi.org/10.1119/1.2835046

Lieberman, D., Dubson, M., Johnsen, E., Olsen, J., & Finkelstein, N. (2014). Physics I MOOC - Educational Outcomes. In *Physics Education Research Conference*.

LON-CAPA. (n.d.). Retrieved from http://www.lon-capa.org/

Mechanics Review. (n.d.). Retrieved December 31, 2014, from https://www.edx.org/course/mechanics-review-mitx-8-mrevx

Mestre, J., Hart, D. M., Rath, K. A., & Dufresne, R. (2002). The Effect of Web-Based Homework on Test Performance in Large Enrollment Introductory Physics Courses. *Journal of Computers in Mathematics and Science Teaching*, *21*(3), 229–251. Retrieved from /p/9259/

Park, EunsikLee, Y. (2001). Estimates of Standard Deviation of Spearman's Rank Correlation Coefficients With Dependent Observations. *Communications in Statistics - Simulation and Computation*, *30*(1).

Patterson, R. (2014). Can Behavioral Tools Improve Online Student Outcomes? Experimental Evidence from a Massive Open Online Course. Retrieved from http://www.ilr.cornell.edu/sites/ilr.cornell.edu/files/cheri_wp165_0.pdf

Pawl, A., Barrantes, A., & Pritchard, D. E. (2009). Modeling applied to problem solving. In *AIP Conf. Proc.* (Vol. 1179, pp. 51–54). American Institute of Physics (AIP). Retrieved from http://dspace.mit.edu/handle/1721.1/76354

Perkins, K. K. (2005). Correlating Student Beliefs With Student Learning Using The Colorado Learning Attitudes about Science Survey. In *AIP Conference Proceedings* (Vol. 790, pp. 61–64). AIP. http://doi.org/10.1063/1.2084701

Pollock, S. J. (2009). Longitudinal study of student conceptual understanding in electricity and magnetism. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020110. http://doi.org/10.1103/PhysRevSTPER.5.020110

Pollock, S. J., & Finkelstein, N. D. (2008). Sustaining educational reforms in introductory physics. *Physical Review Special Topics - Physics Education Research*, *4*(1), 010110. http://doi.org/10.1103/PhysRevSTPER.4.010110

Rayyan, S., Pawl, A., Barrantes, A., Teodorescu, R., & Pritchard, D. E. (2010). Improved Student Performance In Electricity And Magnetism Following Prior MAPS Instruction In Mechanics, 2–5. Retrieved from http://dspace.mit.edu/handle/1721.1/63094

Rayyan, S., Seaton, D. T., Belcher, J., Pritchard, D. E., & Chuang, I. (2013). Participation And performance In 8 . 02x Electricity And Magnetism□: The First Physics MOOC From MITx. *Physics Education*, *11 October*(February). http://doi.org/10.1119/perc.2013.pr.060

RELATE | Research in Learning, Assessing and Tutoring Effectively. (n.d.). Retrieved May 14, 2015, from http://relate.mit.edu/

Sadler, P. M., & Tai, R. H. (2001). Success in introductory college physics: The role of high school preparation. *Science Education*, *85*(2), 111–136. http://doi.org/10.1002/1098-237X(200103)85:2<111::AID-SCE20>3.0.CO;2-O

Shell, D. F., & Husman, J. (n.d.). Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon.

Shell, D. F., & Soh, L.-K. (2013). Profiles of Motivated Self-Regulation in College Computer Science Courses: Differences in Major versus Required Non-Major Courses. *Journal of*

Science Education and Technology, *22*(6), 899–913. http://doi.org/10.1007/s10956-013-9437-9

Shieh, G. (2010). Estimation of the simple correlation coefficient. *Behavior Research Methods*, *42*(4), 906–917. http://doi.org/10.3758/BRM.42.4.906

Taasoobshirazi, G., & Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching*, *48*(8), 901–918. http://doi.org/10.1002/tea.20434