TIME NEEDED FOR UNDERGRADUATE BIOMECHANICS EXAMS

ChengTu Hsieh

Department of Kinesiology, California State University, Chico, CA, USA

The purpose of this study was to determine the appropriate amount of time for a biomechanics exam that consists of different types of questions other than multiple-choice questions. Eighty-nine students enrolled in two introductory biomechanics classes were recruited in the spring and fall of 2017. Students' exam performance and time from three separate exams were obtained. All descriptive statistics were reported. Pearson's correlation and factorial ANOVAs were performed to examine the associations between the exam and time used and the difference between the genders and among exams. There was no association between time used and exam performance in both male and female students. The concept of angular kinematics and kinetics in addition to the fluid mechanics could be a difficult area for students to comprehend and need more time on the exam.

KEYWORDS: learning, pedagogy, teaching, exam time.

INTRODUCTION: Determining the appropriate amount of exam time for students is not an easy task when considering the content and students' ability. Over the years, studies have taken two different approaches to examine the factor of time needed to complete academic exams, speededness and effects of extended time (Mandianch, Bridgeman, Cahalan-Laitusis, & Trapani, 2005). Educational Testing Service (ETS) defines an unspeeded multiple-choice (MC) test as when all test-takers answer 75% of the items and 80% of the test-takers reach the last question (Swineford, 1974). Thus, speededness of a test is the influence of time limits on the performance of the test taker (Hailey, Callahan, Azano, & Moon, 2012). On the other hand, studies also point out the need for time extension as an accommodation for students with disabilities to measure their mastery of the material equitably (e.g., Lovett, 2010). Furthermore, the differences between genders in STEM subject have been widely studied including participation and performance (e.g., Miyake et al., 2010). In an effort to minimize the performance gap between genders, the Oxford University provided extra time on math and computer science exams for female students (Brown, 2018).

The inconsistent results within the literature regarding the effects of timing on exam performance make generalizations difficult. MC items have been heavily favored by instructors for several reasons since they are easy to design and score. Also, MC items can accurately determine students' breadth of knowledge, especially given increasing class sizes and decreasing instructional support. As such, some instructors strive to follow a one MC question per minute of exam duration as a rule-of-thumb (Brothen, 2012; McKeachie, 2002; Renner & Renner, 1999). However, depending solely MC items as a measure of students' knowledge has received criticism for failing to adequately assess higher-order thinking skills (Darling-Hammond, Ancess, & Falk, 1995). To assess various levels of learning, it would be essential for a test to include a combination of MC and open-ended items (Crowe, Dirks, & Wenderoth, 2008). Therefore, classroom-based biomechanics exams should include open-ended response items to assess deep conceptual knowledge. However, these types of items require more time than what is recommended for MC items.

Therefore, how much time is reasonable for an undergraduate biomechanics exam? Unfortunately, a "one-size fits all" approach cannot be applied to the diverse assessment situations encountered in an undergraduate biomechanics curriculum. Since there is very limited empirical research on the time required for an exam consisting of different types of questions, instructors need to be receptive to a wide range of variables when determining the amount of time required for any assessment of student knowledge and learning. Therefore, the purpose of this study was to examine the appropriate time constraints needed by undergraduate students for introductory biomechanics exams. It was hypothesized that the performance of students who don't need special accommodation would have a negative association with the amount of time to complete an exam for both men and women.

METHODS: Eighty-nine students enrolled in two introductory biomechanics classes in Spring and Fall of 2017 were recruited. Data from two students were excluded due to special accommodations for taking exams in a separate testing center that resulted in 41 female and 46 male students in the study. All the policies and procedures for the use of human subjects were followed and approved by the university's institutional review board. Throughout both semesters, all students were required to take three different non-cumulative exams. Each exam was given using a traditional paper-pencil format which consisted of 17-20% of true/false, 70% multiple choice, and 10-13% problem solving (short answer) questions. The difficulty of these questions was designed based on the six levels of Bloom's taxonomy (Crowe et al., 2008).

The questions in the exams were approximately distributed in level 1, knowledge (15%); level 2, comprehension (20%); level 3, application (30%); level 4, analysis (20%); level 5, synthesis (5%); and level 6, evaluation (10%). Table 1 shows the content covered, number of questions, and time allotted for each exam based on the university course schedule. Time used to complete the exam (exam time) was measured with a stopwatch from the beginning of the exam to the moment when a student submitted his/her exam to the instructor. Time was recorded in minutes and seconds. All students started the exam with the package placed on their table at the same time. Students received a reminder when the halfway point of the time allotted was reached.

All statistical assumptions were checked before the statistical analysis. The Shapiro-Wilk test was used to examine the normality of the data sets. To examine the association between the exam time and performance overall and separately for both genders, Pearson's correlation and coefficients were performed. Since there are six pairs of correlations for exam performance and exam time used, the statistical significance was set at 0.008 to control type I error. Two factorial ANOVAs were performed to examine the effect of genders and exam content (2 X 3) on time to completion and exam performance. Tukey HSD post-hoc test was applied as needed. The statistical significance was set at 0.017 due to three planned contrasts were performed.

Table 1: Exam content, number of questions, and time allowed

Exams	Content	# of Questions	Time (min)
Exam 1	Linear kinematics & kinetics	50	75
Exam 2	Angular kinematics, kinetics, & fluid mechanics	50	75
Exam 3	Functional anatomy & internal biomechanics	60	90

RESULTS: All students were able to complete the exams within the allotted time frame for exams 1 and 3. A few students were unable to complete exam 2 during the allotted time and they were allowed extra time to finish all questions without being rushed or reminded of the time limit. The descriptive statistics for the exam performance and time used were listed in Tables 2 and 3. A normality check indicated all exam scores were normally distributed (*P*-values > 0.05). There was no significant association between time used and exam performance with all *r*-values ranging between -0.04 and -0.13 (Table 4). Additionally, no significant association between time used and exam performance was found in either gender independently.

Table 2: Mean and SD for the exam performance in %

	Exam 1	*Exam 2*	*Exam 3*
Male	**79.8 ± 12.7**	74.8 ± 15.8	**75.9 ± 10.1**
Female	**72.4 ± 10.1**	68.7 ± 12.3	**70.4 ± 10.3**
Total	76.3 ± 12.1	71.9 ± 14.5	73.3 ± 11.0

Note: Bolded numbers present significant difference between men and women ($P < 0.01$).

Table 3: Mean and SD for time used in minutes:seconds

	*T1**	*T2*	*T3*
Male	56:15 ± 9:51	70:26 ± 8:09	68:18 ± 9:51
Female	56:12 ± 10:23	69:05 ± 8:56	69:04 ± 4:01
Total	56:28 ± 10:17	69:31 ± 8:17	68:30 ± 15:16

Note: * represent a significant difference from *T1* to either *T2* or *T3* in all categories ($P < 0.01$).

Factorial ANOVA showed no significant effects of gender and the interaction between gender the time used in exams ($P > 0.05$). The only significant difference was found in time used for three exams. Exam 1 took the shortest amount of time to complete for all students when compared to Exam 2 and 3 ($P < 0.01$). The amount of time used to finish exams 2 and 3 were similar ($P > 0.05$). The second factorial ANOVA indicated no significant effects of exam performance and the interaction between gender and exams performance ($P_{values} > 0.05$). The effect of gender showed that male students did better than their female counterparts on exams 1 and 3 with *P*-values < 0.01. Additionally, the exam 1 performance was significantly associated with exam 2 and 3 performance separately (Table 4, $P_{Values} < 0.008$).

Table 4: Associations (*r*) between exam performance and time used

	Exam 1	*Exam 2*	*Exam 3*	*T1*	*T2*	*T3*
Exam 1	1.00					
Exam 2	*0.24	1.00				
Exam 3	**0.28	***0.59	1.00			
T1	-0.13	0.04	0.04	1.00		
T2	0.08	-0.13	0.06	0.10	1.00	
T3	-0.04	-0.18	-0.04	0.07	***0.68	1.00

Note: * represents $P < 0.05$, ** represents $P < 0.01$, and *** represents $P < 0.008$.

DISCUSSION: Time limits on exams are enforced for reasons of practicality and efficiency. However, a potential association between exam time and performance could be a concern and a potential threat to quality assessment of student learning in biomechanics. If students cannot complete exams, their performance is negatively impacted and getting a clear picture of what the student knows is confounded by the aspect of speed. The findings of the current study indicated that student performance was not associated with the time used to complete the exam (r = -0.04 to -0.13, $P_{values} > 0.05$). This supports Brothen's (2012) finding that in a proctored environment students' exam performance had very low or even no association with the time used for the exam. According to ETS guideline for an unspeeded test, 80% of the test-takers must reach the last question which represents about ± 1.28 standard deviation of the normal distribution. For exam 1 to meet the ETS guideline, the time needed would be approximately 70 minutes, leaving an additional 5 minutes to assess more content. For exam 3, about 88 minutes would be needed to meet guideline set forth by ETS and only 2 minutes would be available to answer additional questions. Exam 2 was a speeded exam since only 70% of students completed the exam within the time allotted. Thus, items may need to be altered in format or be removed to meet the ETS requirements for these students and biomechanics content.

In order to determine the appropriate amount of exam time for students, instructors need to consider two factors related to a number of questions: type of questions and content of questions. In postsecondary institutions, it is necessary to provide assessments that evaluate the spectrum of Bloom's taxonomy to get a rich understanding of student learning. For a topic like biomechanics, it is not merely enough for students to demonstrate memory of information (breadth); it is an applied scientific topic and requires depth of understanding. Crow et al. (2008) recommended the inclusion of open-ended items to accurately assess higher order thinking skills and overall student mastery. However, testing higher order thinking skills (e.g., synthesis, analysis, evaluation) is time-consuming (Brady, 2005) and requires adjustment of

the number and types of items for them to be an efficient assessment of students' proficiency of biomechanical concepts.

In addition to the type of item given on the exam, the content of the item also impacts the speed or pace at which the student requires to successfully complete the exam. The speed at which the test-taker works during the exam and the amount of effort required by the type of items will dictate how many items can be included on an exam to align with the ETS guideline. Although exams 1 and 2 had the same number of questions, students spent 23% more time to complete exam 2. In addition, the exam 3 had 20% more questions than the exam 2, students spent the similar amount of time to complete the exam. This would suggest that the content of angular kinematics and kinetics in addition to the fluid mechanics are difficult concepts and require more time and effort for these students to comprehend in the challenging environment of the exam. Finally, the results from the current study also supported pedagogical logic that the content of exams in biomechanics are likely dependent on previous content and academic performance, given there were moderate correlations between performance in all three exams (Table 4).

The limitations of this study are but not limited to 1) the time used for different types of biomechanical questions is unclear, 2) the true time needed for each student is difficult to determine, 3) unknown reliability of the exam, 4) inability to equalize the difficulty of the content across the three exams, and 5) sample size for each semester was small. One suggestion for future study is to examine student performance and time needed between MC and open-ended questions separately.

CONCLUSION: This study found that when ample time was provided the students' exam performance was not associated with the time used to complete the exam. The concepts for angular kinematics and kinetics in addition to fluid mechanics were difficult for these students to comprehend and apply in a problem-solving situation during an exam, consequently, more time or a revised set of questions may be appropriate.

REFERENCES

Brady, A. M. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, *5*, 238-242. doi:10.1016/j.nepr.2004.12.005

Brothen, T. (2012). Time limits on tests: Updating the 1-minutes rule. *Teaching of Psychology*, *39*(4), 287-292. doi:10.1177/0098628312456630

Brown, A. (2018, January 23). Oxford University extends maths exams to improve women's grades. *The Oxford Student News Paper*. Retrieved from http://oxfordstudent.com/2018/01/23/oxford-university-extends-maths-exams-improve-womens-grades/

Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sciences Education*, *7*(4), 368-381.

Darling-Hammond, L., Ancess, J., and Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.

Hailey, E., Callahan, C. M., Azano, A., & Moon, T. R. (2012). An evaluation of test speededness in an assessment for third-grade gifted students. *Journal of Advanced Academics*, 23(4), 292-304.

Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities: Answer to five fundamental questions. *Review of Educational Research*, *80*(4), 611-638.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330, 1234–1237.

Swineford, F. (1974). *The test analysis manual (ETS SR 74-06)*. Princeton, NJ: Educational Testing Service.