

PEER REVIEWED

Does Self-Regulated Test Duration Correlate with Vision Science Test Score in First-Year Optometry Students?

Patricia M. Cisarik, OD, PhD, FAAO, and Melissa Powers, MS

Abstract

We explored the relationship between self-regulated test duration and test performance in first-year optometry students using test administration software. The scores for two midterms (MT1 and MT2) and the final exam for a single course (Visual Sensation and Perception) and time from password submission to exam upload were obtained for 132 students in spring 2019. Statistical analysis suggests that the relationship between self-regulated test duration and test performance is inconsistent across the semester for this group. Exploration of self-regulated test duration and test performance for other courses and throughout the optometric program may provide additional insight.

Key Words: test duration, multiple choice, test score, digital administration, performance

Background

Students in the United States generally have experienced limited duration examinations several times before entering an optometric degree program (e.g., SATs, ACTs, OATs and in undergraduate course work). For optometric education didactic courses, typical practice is for students to be given tests of knowledge that are of limited duration (e.g., 60 minutes). Although students may apply for extended time for examinations if they have a qualifying condition, most students must complete their didactic course tests during the allotted time. Personal observation (in three different optometric courses) is that, for an allotted examination time of one hour, some students will complete the examination in approximately 15 minutes, many complete the exam within 45 minutes, and a handful of students remain for the entire test period. Whether the time they spend on the test (self-regulated test duration) is related to their test performance is unknown.

In a paper presented at the Association for Institutional Research Annual Forum, Chicago, IL, June 2010, Hosch examined time on test, student motivation and performance on the Collegiate Learning Assessment (CLA), which Hosch states is a “low stakes” examination.¹ The CLA is a timed essay test used to assess student learning at the undergraduate level. Hosch found a strong relationship between time students spent on taking the test and test performance ($p < .001$). In a different study involving students in an undergraduate statistics course, Landrum et al. found that the students’ self-reported test completion time was sometimes, but not consistently, negatively correlated with test performance.² A literature review did not find any studies that examined time taking the test and test performance in graduate health professional programs or on tests with only multiple-choice items.

With digitally administered tests, information about how much time each student spends taking the test can be objectively obtained. Given the paucity of published research on the relationship between a student’s self-regulated test duration and test performance, reporting any identified relationship between

these variables is of interest to both educators and students because the knowledge may help improve test performance.

We aimed to examine the relationship between self-regulated test duration on a test of known time limit and test performance on multiple-choice examinations in a required basic science class for first-year students in an optometric program.

Methods

After affirmation by the Institutional Review Board of Southern College of Optometry (SCO) that the protocol met the requirements for review exemption, a retrospective review of already-existing data was performed. The multiple-choice test scores for two midterm examinations (MT1 and MT2) and a final examination for the first-year course “Visual Sensation and Perception” and the time spent taking each test were obtained for the first-year students at SCO in spring 2019. All tests were digitally administered via ExamSoft (ExamSoft Worldwide LLC, Dallas, TX). All tests had been administered in the same large lecture hall with consistent environmental conditions. Students had an allotted test duration of 60 minutes each for MT1 and MT2 (both administered from 11 a.m. to noon) and 120 minutes for the final examination (administered from 12:30 p.m. to 2:30 p.m.). The software automatically uploaded the responses of any student who had not self-uploaded their responses prior to the expiration of the test's time limit. MT1 and MT2 each included 40 items; the not directly cumulative final examination included 60 items. Each test was worth 20% of the total course grade. A member of the Information Technology staff at SCO (second author) provided the instructor of record for the course (first author) with each student's test scores for the three tests (percent correct) and the total test time (in minutes) between the password input by the student to begin the test and the completed upload of the student's examination responses.

SPSS v.26.0 (IBM) was used for statistical analysis. The Shapiro-Wilk test was used to test the shape of the data distributions to determine the appropriate tests for evaluating the relationships between self-regulated test duration and test score. Related samples Wilcoxon signed rank test, related samples Friedman's two-way analysis of variance (ANOVA) by ranks, Mann-Whitney U test, and independent samples t-test were used, as appropriate for data distribution, for data comparisons. All reported P values are two-tailed.

Results

With respect to the class results as a whole, the Shapiro-Wilk tests for the distributions of self-regulated test times and test scores for each examination indicated that only the test scores for the final exam were approximately normally distributed. For the grouped data described for Analysis 1 and Analysis 2 below, all self-regulated test duration distributions were not normally distributed within each group, but test score distributions within each group were approximately normal.





Table 1. [Click to enlarge](#)


Figure 1(a) and **Table 1** (column 2) show the distributions of the self-regulated test durations for each of the tests. The related samples Wilcoxon signed rank test indicated that the median of the differences in the time spent on MT1 compared to the time spent on MT2 was significantly different from zero (standardized test statistic=-4.22, $P<.001$). The mean times spent on MT1 and on MT2, respectively, were 42.0 ± 11.0 and 39.1 ± 11.5 minutes. The proportions of the allotted time spent on each test that the means represent were, for MT1, MT2 and the final examination respectively, 0.7, 0.65 and 0.43.


Figure 1(b) and **Table 1** (column 3) show the distributions of the test scores for the two midterms and for

the final examination. The mean test scores (percent correct) for MT1, MT2 and the final examination, respectively, were 80.3±10.4, 78.0±12.1 and 64.5±10.9. A related samples Friedman’s two-way ANOVA by ranks test indicated that the distributions of test scores for MT1, MT2 and the final examination were significantly different (test statistic=102.07, P<.001). Pairwise comparisons indicated that the distributions of test scores were different between MT1 and MT2, MT2 and the final, and MT1 and the final (P<.04 for each, with a Bonferroni correction for multiple tests).

 **Figure 1.** Boxplots showing the distribution of (a) self-regulated test times and (b) test scores across the three multiple-choice tests administered digitally in the course. Error bars represent 1 standard deviation from the mean. The median of the differences in self-regulated test time between midterm 1 (MT1) and midterm 2 (MT2) was significantly different from zero (P<.001), and the distributions of test scores were significantly different across the three tests (P<.001). [Click to enlarge](#)

 **Figure 2.** Scatterplots, with regression line and R2 values, showing the relationship between self-regulated test time and test score for (a) midterm 1 (MT1), (b) midterm 2 (MT2), and (c) the final examination. Statistical analysis showed no significant linear relationship between the two variables for any of the three tests. [Click to enlarge](#)

 **Figure 3.** Boxplots showing the distribution of (a) self-regulated test times and (b) test scores for the three tests across two groups defined by the proportion of the allotted test time that the subject spent taking the test. The distributions of self-regulated test durations (a) were significantly different for the two groups for all three tests; the means of the test scores (b) were significantly different only for midterm 1 (MT1) (P=.020) and the final examination (P=.022). [Click to enlarge](#)

 **Figure 4.** Boxplots showing the distribution of (a) self-regulated test times and (b) test scores for the three tests across two groups defined by the median self-regulated test duration for each test. The distributions of self-regulated test durations (a) were significantly different for the two groups for all three tests; the means of the test scores (b) were not significantly different for any of the tests (P>.05 for all). [Click to enlarge](#)

For eight students who used the entire exam period for MT1 and for nine students who used the entire exam period for MT2, the self-regulated test duration registered was >60 minutes due to the upload times; however, none of these students had test durations registered that were more than four minutes past the allotted time; therefore, their data were included in the analysis. **Figures 2(a-c)** show that the relationship between test score and self-regulated test duration was not linear for any of the three tests (MT1: R²=.022, P=.091; MT2: R²=.001, P=.70; final: R²=.025, P=.072).

TABLE 2
Proportions (%) of Students with Test Scores of "B-" and "B" Based on Test Time Group

Score	MT1 – Allotted Time Groups		MT1 – Median Time Groups		MT2 – Allotted Time Groups		MT2 – Median Time Groups		Final – Allotted Time Groups		Final – Median Time Groups	
	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
< 80	6	54	28	32	15	53	33	35	96	25	60	61
≥ 80	15	57	40	32	15	48	30	34	6	5	4	7
< 83	8	73	36	43	21	64	41	44	90	26	61	63
≥ 83	13	38	30	21	9	36	22	25	4	4	3	5

Table 2. Only the proportion of students scoring above 83 on midterm 1 (MT1) differed between the two groups defined by the time allotted for the test (Chi-square=5.70, P=.017) [Click to enlarge](#)

To further explore the relationship between self-regulated test duration and test score, two different analyses were done in which, for each test, the students were divided into two groups. For Analysis 1,

those whose self-regulated test durations were less than half the allotted time were assigned to Group 1, and those whose self-regulated test durations were greater than or equal to half the allotted test time were assigned to Group 2. For Analysis 2, those whose self-regulated test durations were less than half the median time for each test were assigned to Group 1, and those whose self-regulated test durations were greater than or equal to the median time for each test were assigned to Group 2. The distributions of the self-regulated test durations were compared between the two groups formed for each analysis to assure that test time distributions were significantly different for the two groups compared. The test scores were then compared between the two groups for each analysis. Table 1 shows the means (\pm SD) of the self-regulated test durations (columns 4, 5, 8 and 9) and test scores (columns 6, 7, 10 and 11) for all three tests for each group for both analyses.

For Analysis 1, the number of subjects in each group varied depending on the test. For MT1, N=21 for Group 1 and N=111 for Group 2; for MT2, N=30 for Group 1 and N=102 for Group 2; for the final, N=30 for Group 1 and N=102 for Group 2. **Figure 3** and Table 1 (columns 4-7) show the results for Analysis 1 (groups defined by allotted time). Mann-Whitney U test indicated that the distributions of self-regulated test durations were significantly different for the two groups for all three tests (MT1: U=2331, Z=7.26, P<.001; MT2: U=3060, Z=8.31, P<.001; final: U=3060, Z=8.31, P<.001). Independent t-test indicated that the mean test scores were significantly different for MT1 (P=.020, 95% CI .93, 10.53), not significantly different for MT2 (P=.95, 95% CI -4.81, 5.14) and significantly different for the final (P=.022, 95% CI -9.56, -.76).

For Analysis 2, the number of subjects in each group varied slightly depending on the test. For MT1 (median test time=41 minutes), N=68 for Group 1 and N=64 for Group 2; for MT2 (median test time=38 minutes), N=63 for Group 1 and N=69 for Group 2; for the final (median test time=48 minutes), N=64 for Group 1 and N=68 for Group 2. **Figure 4** and Table 1 (columns 8-11) show the results for Analysis 2 (groups defined by median test time). Mann-Whitney U test indicated that the distributions of self-regulated test durations were significantly different for the two groups for all three tests (MT1: U=4352, Z=9.92, P<.001; MT2: U=4347, Z=9.91, P<.001; final: U=4352, Z=9.91, P<.001). Independent t-test indicated that the means of the test scores were not significantly different between groups for any of the tests (MT1: P=.09, 95% CI -.46, 6.64; MT2: P=.85, 95% CI -4.59, 3.76; final: P=.88, 95% CI -3.47, 4.06).

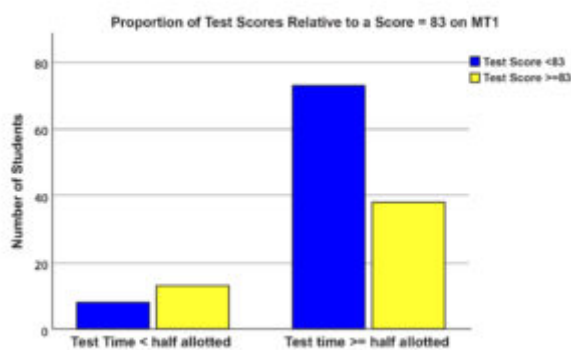


Figure 5. Bar graph showing the number of students who scored below 83 vs. 83 or above on midterm 1 (MT1) across the two groups defined by the proportion of the allotted test time that the subject spent taking the test. The proportion of students scoring above 83 was significantly different for the two groups only for MT1 (Chi-square=5.70, P=.017). Proportions of students scoring above 83 were not significantly different between the two groups for any of the other tests, nor for any test when compared between the two groups defined by median self-regulated test duration.

[Click to enlarge](#)

Finally, we investigated whether the proportion of students scoring 80 or above (grade letter B- or better) and 83 or above (grade letter B or better) on each test differed between the two groups defined for Analysis 1 and Analysis 2 (**Table 2**). The results indicated that only the proportion of students scoring above 83 on MT1 differed between the two groups defined by the time allotted for the test (Chi-square=5.70, P=.017) (**Figure 5**).

Discussion

The main finding of this study was that when the amount of time spent taking a multiple-choice test in a basic science course in the optometric program was less than half the allotted time for the test, this group of first-year optometry students demonstrated a mean test score that was significantly higher for MT1, significantly lower for the final examination, and not significantly different for MT2, compared to the scores of their classmates who spent more than half the allotted test time completing the test. Additionally, no linear relationship was found between self-regulated test durations and test score for any of the three tests. Finally, the proportion of students achieving a test score of 83 or above (equivalent to a letter grade better than B-) was significantly different between those spending more vs. less than half the allotted test time completing the exam only for MT1.

Several reasons may explain the lack of consistency across these results. First, students in an optometric program are not only well-versed in multiple-choice test-taking strategies prior to matriculation, but have demonstrated proficiency in examinations as ascertained through their optometry school applications. Although the material may be new to them, they likely rely on strategies that have brought them success (for example, a pre-determined number of times that the student will review the entire test). Second, the material covered by the first two tests differed in nature, though the amount of material covered by MT1 and MT2 was about the same; students may have found the material on the second examination more difficult to understand. Third, for the first test given by an instructor, students have the additional unknown of the instructor's question-writing style; hence, students may have put more effort into studying for MT1 compared to MT2 and the final examination.

The overall lower scores for MT2 and the final exam compared to MT1 suggested that cumulative knowledge may have contributed to lower scores on MT2 and the final exam, even though neither MT2 nor the final exam was directly cumulative. The overall lower scores for the non-cumulative final examination may have reflected the greater amount of material covered and the students' knowledge of their grades going into the examination, with study time and/or effort on the test resulting from their calculation of the score needed to achieve their desired grade.

Another factor that may explain the better performance on MT1 compared to the other two tests is that the students had no knowledge of how well they were doing in any of their classes at the time of the first midterm. If they had performed less than desired in other courses, they may have put in less effort on MT2 and the final for this course. Finally, the students had four midterm examinations scheduled for other courses in the two days following MT1 and MT2, but the final examination was the last of seven examinations in that week; therefore, fatigue may have contributed to the lower scores for the final examination despite the proportionately larger amount of time allotted for the examination.

Although the subjects were first-year students, they had exposure to the software for test administration for all of their first-semester courses; the data collection for this study took place during the second semester. Thus, although the students were relative novices with the software, unfamiliarity with the software likely did not contribute significantly to self-regulated test duration.

The way in which test time was calculated for this study also may have contributed to inconsistent results across the tests. We used the "total time" that each student had access to the examination questions. This value was unable to reflect how much time the students actually spent with their attention on the

test. Other metrics may better reflect test scores, such as the number of times that the student viewed each question, the number of times the student viewed the more difficult questions, the number of times the student changed the answer for each question, whether or not certain features of the software were used by the student (highlighting, etc.). Further investigation into whether these indices are related to test scores may provide useful information for improving student performance.

Non-test factors may also have differed across testing days and created inconsistencies in the results, such as alertness, caffeine consumption, non-academic stressors, food consumption, medications, health on test day, etc.³⁻⁵ Additionally, for those who performed less than desired on MT1, test anxiety may have contributed to worse performance on the following tests. Finally, weaker language skills have been associated with longer completion times for multiple-choice exams;⁶ we did not use any metric of language skills to evaluate the influence of this factor.

Conclusion

Self-regulated test duration on a time-limited test as measured with digital test administration software did not show a consistent relationship with test performance on multiple-choice tests over the course of a semester in a basic science course in the first-year optometric program at SCO. Whether a more consistent relationship between self-regulated test duration and test performance would manifest when controlled for other academic and non-academic factors remains to be explored. Additionally, an analysis of self-regulated test duration and test performance in other first-year courses and in second- and third-optometric-year course work for this same class of students may provide insight into the development of test-taking strategies that may benefit new optometric program students. Further exploration of other test-taking strategy information obtained by test administration software is warranted.

Acknowledgements

We wish to thank Southern College of Optometry for permitting access to the test duration information.

References

1. Hosch BJ. Time on test, student motivation, and performance on the Collegiate Learning Assessment: implications for institutional accountability. *Journal of Assessment and Institutional Effectiveness*. 2012;2(1):55-76.
2. Landrum RE, Carlson H, Manwaring W. The relationship between time to complete a test and test performance. *Psychology Learning & Teaching*. 2009;8(2):53-56.
3. Mushtaq I, Khan SN. Factors affecting students' academic performance. *Glob J Management Bus Res*. 2012;12(9):17-22.
4. Rasul S, Bukhsh Q. A study of factors affecting students' performance in examination at university level. *Procedia – Social and Behavioral Sciences*. 2011;15:2042-47.
5. Gajghat RH, Handa CC, Himte RL. Factors influencing academic performance of the students at university level exam: a literature review. *Int J Res Eng Tech*. 2017;6(5):102-10.
6. Burnham TA, Makienko I. Factors affecting exam completion speed, exam performance, and nonexam performance. *J Mark Educ*. 2018;40(2):140-51.

Dr. Cisarik [pcisarik@sco.edu] has been on the faculty at Southern College of Optometry since 2008. In addition to clinical teaching, she is Instructor of Record for Visual Sensation and Perception, Introduction to Binocular Vision, and Neuro Eye Disease

Ms. Powers has a Master of Education in Learning Systems Technology. She has been the Systems Administrator for Learning Management at Southern College of Optometry for two years and has worked in higher education for more than 15 years.