From Policing to Learner Analysis: Use of ExamSoft Snapshots to Analyze Students' Exam Taking Behaviors in Live and Online Exams

Dan Cernusca, Ph.D.
North Dakota State University, College of Health Professions, School of Pharmacy

Abstract

A Doctor of Pharmacy program is using a specialized examination software, ExamSoft® to addesss accreditation requirements. The primary objective of this study was to use ExamSoft snapshots log to provide empirical evidence of exam taking behaviors and the potential impact of the pandemic context on exam taker behaviors across different examination contexts. Significant behavioral differences were found between high and low achieving students as well as between face-to-face and online live monitored exams.

Motivation and Objective of the Study

The use of multiple-choices exams combined with the introduction of computer-testing platforms increased educational researchers' interest in test-taking behaviors, changes in these behaviors, and their relationships with the assessment activities (Olev & Must, 2013; Plake & Wise,1988; Tamowsky, 1977). As interest in test-taking behaviors increased, researchers started to focus on more specific impact areas such as the potential impact of test-taking behaviors on students' cognitive test performance (e.g., Heinonen et al., 2011), the role of self-efficacy in mediating the impact of test-taking behaviors on test performance (e.g., Birenbaum & Alhija, 2013), or the association between test-taking behaviors and test performance for students with special needs (e.g., Pohl et al., 2016).

Pharmacy programs are required by an accreditation agency to provide specific data to support students' progress (ACPE, n.d.). To address these requirements, a PharmD program began using a specialized examination software, ExamSoft®, that allowed instructors to link learning objectives and outcomes to assessment items. ExamSoft was integrated with the learning management system and provided valuable statistical data used to monitor the quality of individual exams items and the whole exam. ExamSoft includes security features such as device lock during the exam, item randomization, and snapshots that record all choices students make during an exam.

The primary objective of this study was to analyze students' snapshots and provide empirical evidence of exam taking behaviors, their potential relationship to the exam performance, and the potential impact of the pandemic context on exam taker behaviors.

Context of the Study

At the beginning of Fall 2020 the university set up pandemic restrictions that allowed both a mixed-attendance and a fully synchronous online attendance option. Both students and the instructor had the option to use either the face-to-face or online format of the course. For live attendance and examination options, the classroom seating capacity was limited and face masks were required.

To address the needs of fully online students, ExamSoft Exam Integrity functionality offered strong security options including offline video, audio recording, and flagging of potential breaches of integrity. For the target pharmaceutical course, in the first half of the semester, when COVID 19 pandemic restrictions were more relaxed the course was live and additional classrooms accommodated a live first exam for the entire class. For the second part of the course, the conditions worsened and the instructor decided to teach and administer the second examination fully online for the entire class. This situation offered a unique opportunity to analyze the exam taking behaviors of the same group of students both in the live and online context and analyze the potential impact of the pandemic-related conditions on the examination process.

Research Focus and Methodology

Research Design

We used a sequential mixed-methods design that built on a qualitative analysis of the raw ExamSoft data to identify significant examination behaviors followed by a comparative qualitative analysis of the identified significant exam-taker behaviors. The dependent variables used in this study were behavioral patterns, assessment item difficulty, and the maximum number of choices an exam taker made for each assessment item throughout the examination period. The independent variables were grade level (high, within top 20% of the exam grades or low, within the bottom 20% of the grades), nature of the exam (face-to-face or online with digital exam monitoring), and the correctness of the final answer (correct or wrong).

Data Collection

ExamSoft records the entire activity of the exam taker during the exam in what is known as "Exam Taker Snapshots". This detailed information was intended to serve as a means to provide support for the instructor when checking and identifying potential attempts of cheating. While this "policing" tool is considered a needed and useful security feature, the Exam Taker Snapshots (or simply ExamSoft snapshots) can also provide support for tracking students' observed exam-taking behaviors (Cernusca & Friesner, 2019). That is, for multiple-choice assessment items ExamSoft snapshots include all choices students made (e.g., view an assessment item, select a specific answer, change a previously selected answer) and the time-stamp associated with each choice.

The ExamSoft snapshot logs are available through the software interface (Figure 1) and allow for basic sorting of the information by assessment item or time stamp. However, the ExamSoft snapshot logs can also be downloaded as an Excel spreadsheet and used for further more detailed analyses.

Figure 1

Sample Snapshot Viewer in ExamSoft

Question numbering in the snapshot reflected the default organization of the exam even when questions were randomized. The spreadsheet format of the snapshots allowed for the organization of choices by question number and time stamp to reflect the exam taker behaviors for each question. In addition to the snapshot viewer the item analysis output from ExamSoft provided the correct answer and assessment item difficulty (Figure 2)

Figure 2

ExamSoft Item Analysis Output Sample

Participant Selection

The snapshot of seven high performers that were placed in the first 20% of the class and seven low performers placed in the lower 20% of the class from both live and online exams were selected for this research study.

This number of participants was considered optimal considering that each exam used for this study had between 35 and 50 assessment items and for each assessment item the ExamSoft snapshot log recorded between two and 10 behavioral actions.

Data Analysis

Data analysis involved a qualitative phase that helped identify significant behavioral patterns and a quantitative phase where identified behavioral patterns quantified and used for statistical analysis using SPSS® (https://www.ibm.com/products/spss-statistics) statistical software.

Qualitative Analysis Phase

The structure of the ExamSoft Snapshots allowed for a qualitative analysis of the observed exam taking behaviors followed by grouping of these behaviors into specific behavioral patterns. As a first step in the qualitative analysis the snapshot Excel output was organized by assessment item and then by the choice (recorded exam taker action). The first choice for each assessment item was then color coded and a column with the correct answer for each item was added to help analyze the sequence of the choices made by the exam taker (see Figure 3).

Figure 3
ExamSoft Snapshots Organized by Item and Choice

	Item#	Snapshot#	Item Type	Time Stamp	Trigger	Response	correct_answer
1	Item#	Snapshot#	Item Type	Time Stamp	Trigger	Response	correct_answer
2	3	1	Choice	4:04:43 PM	Answered		A
3	3	2	Choice	4:14:16 PM	Answered	Choice(s): A	
4	3	3	Choice	4:33:04 PM	Final	Choice(s): A	
5	4	1	Choice	4:09:16 PM	Answered	Choice(s): A	C
6	4	2	Choice	4:33:04 PM	Final	Choice(s): A	
7	5	1	Choice	4:05:11 PM	Answered		A
8	5	2	Choice	4:14:36 PM	Answered	Choice(s): A	
9	5	3	Choice	4:33:04 PM	Final	Choice(s): A	
10	6	1	Choice	4:07:18 PM	Answered	Choice(s): C	C
11	6	2	Choice	4:33:04 PM	Final	Choice(s): C	
12	7	1	Choice	4:11:31 PM	Answered		A
13	7	2	Choice	4:29:29 PM	Answered	Choice(s): D	
14	7	3	Choice	4:33:04 PM	Final	Choice(s): D	
15	8	1	Choice	4:03:43 PM	Answered	Choice(s): A	A
16	8	2	Choice	4:22:48 PM	Answered	Choice(s): B	
17	8	3	Choice	4:33:04 PM	Final	Choice(s): B	
18	9	1	Choice	4:04:22 PM	Answered	Choice(s): D	D
19	9	2	Choice	4:33:04 PM	Final	Choice(s): D	
20	10	1	Choice	4:04:15 PM	Answered	Choice(s): C	C
21	10	2	Choice	4:33:04 PM	Final	Choice(s): C	

This structure of the snapshot log allowed to identify several micro-behaviors: viewing a question without selection, selection of an answer, change of an answer from correct to wrong, or change of a wrong answer to a correct one. These micro-behaviors were then used to defined 10 overall behaviors that reflected potential sequencing across observed exam taker choices for the dataset used in this study. Examples of these behaviors include: correct answer from first choice to the final choice; view without a choice then correct answer to final choice; wrong answer as the first choice then correct answer to final choice; or wrong answer at the first choice, changed to correct answer and then wrong answer to final choice. The identified basic behaviors were mixed into behavioral patterns that included repetitive patterns. For the scope of this study, the behaviors were grouped in four outcomes, two beneficial and two damaging.

The two beneficial behavioral patterns were:
       (b1-cc) correct - correct, correct answer from the beginning to the end of the exam
       (b2-wc) wrong - correct, wrong answer in the beginning changed into a correct answer by the end of the exam.
The two damaging behavioral patterns were:
       (d1-cw) correct - wrong, correct answer in the beginning changed to a wrong answer
       (d2-ww) wrong - wrong, wrong answer from the beginning to the end.

Qualitative Analysis Phase

For the quantitative phase of the analysis, two exams in the same course that were administered during the COVID-19 pandemic were considered for this study. The first one was administered during the first stage of the pandemic when the face-to-face attendance was acceptable with distancing and class occupancy and masking requirements were strictly observed. The second exam was administered at the time when the pandemic requirements were tightened and remote instruction was in place. For each of the exams, seven high achieving students placed in both exams within top 20% of the grades as well as seven low achieving students placed within the bottom 20% of the scores were selected for this study. For each exam each student's snapshots were coded at the assessment item level for the four observed behavioral patterns identified in the previous step and then the percentage of each behavior relative to the total number of assessment items was computed.

In the next step, a synthesis Excel spreadsheet was created by recording for each student the percentages for each of the four observed behavioral patterns computed in the previous step, the maximum number of choices, the correct/wrong answer code, the grade level and the exam code (1-live, 2-remote). Figure 4 shows a snapshot of the synthesis Excel spreadsheet generated and used for this phase of the data analysis.

Because of the small number of participants, a non-parametric test, Kruskal-Wallis, was used to compare the considered dependent behavioral variables and number of choices by exam type and respectively achievement/grade level.

Figure 4

Synthesis Spreadsheet for Statistical Data Analysis

1	StudID	ex_id	grade_level	b1_cc	b2_wc	d1_cw	d2_ww	max_choices
2	fs20_413_ex1_live_high_1	1	1	96.43	0.00	0.00	3.57	3
3	fs20_413_ex1_live_high_2	1	1	96.43	0.00	0.00	3.57	3
4	fs20_413_ex1_live_high_3	1	1	96.43	0.00	0.00	3.57	3
5	fs20_413_ex1_live_high_4	1	1	96.43	0.00	0.00	3.57	3
6	fs20_413_ex1_live_high_5	1	1	89.29	3.57	0.00	7.14	4
7	fs20_413_ex1_live_high_6	1	1	100.00	0.00	0.00	0.00	3
8	fs20_413_ex1_live_high_7	1	1	82.14	7.14	0.00	10.71	10
9	fs20_413_ex1_live_low_1	1	2	46.43	14.29	3.57	35.71	4
10	fs20_413_ex1_live_low_2	1	2	32.14	21.43	28.57	17.86	4
11	fs20_413_ex1_live_low_3	1	2	64.29	0.00	0.00	35.71	3
12	fs20_413_ex1_live_low_4	1	2	50.00	7.14	7.14	35.71	5
13	fs20_413_ex1_live_low_5	1	2	46.43	7.14	7.14	39.29	5
14	fs20_413_ex1_live_low_6	1	2	53.57	3.57	7.14	35.71	5
15	fs20_413_ex1_live_low_7	1	2	60.71	0.00	3.57	35.71	4
16	fs20_413_ex2_rem_high_1	2	1	89.66	6.90	0.00	3.45	7
17	fs20_413_ex2_rem_high_2	2	1	68.97	31.03	0.00	0.00	10

Findings

When the observed beneficial behavioral patterns were analyzed by the student achievement levels, a Kruskal-Wallis test indicated that high achieving students ranked statistically significant higher in beneficial correct-correct behavioral pattern (21.5) than low achieving students (7.5), $H(1) = 20.37$, $p < 0.001$. However, the two groups ranked about the same in beneficial wrong-correct behavioral pattern ($p = .0.083$).

For the damaging behavioral patterns, the Kruskal-Wallis test indicated the same type of differences between low and high achieving students:

- for damaging correct-wrong behavioral pattern the low achieving students ranked statistically significant higher (20) than high performing students (9), $H(1) = 16.16$, $p < 0.001$;

- for damaging wrong-wrong behavioral pattern, the low achieving students ranked statistically significant higher (21.5) than high performing students (7.5), $H(1) = 20.59$, $p < 0.001$;

When only the high achieving students were considered for the analysis, a Kruskal-Wallis test indicated a statistically significant higher ranking only for damaging wrong-wrong behavioral pattern during the live exam (10.1) when compared to the remote exam (8.9), $H(1) = 6.04$, $p < 0.05$. No statistically significant differences were found for the two positive behavioral patterns and second damaging behavioral pattern (correct-wrong).

However, when only the low achieving group was considered for the analysis, the Kruskal-Wallis test indicated:

- a statistically significant lower ranking for correct-correct beneficial behavioral pattern for the remote exam (4.4) when compared to the live exam (10.6), $H(1) = 7.65$, $p<0.01$, and respectively

- a statistically significant higher ranking for wrong-wrong detrimental behavioral pattern for the remote exam (11) when compared to the live exam (4), $H(1) = 10.27$, $p<0.01$. No statistically significant differences for the beneficial wrong-correct and detrimental correct-wrong behavioral patterns between the two types of exams.

Finally, when the maximum choices by question were analyzed across the two types of exams (live and remote), the Kruskal-Wallis test indicated a statistically significant ranking for the remote exam when compared to the live exam for both the high achieving group ($p<0.05$) and the low achieving group ($p<0.01$).

Discussions

By expanding the use of ExamSoft snapshots from exam security purpose to the analysis of exam taker behaviors we were able to identify behavioral patterns and their variation across student achievement levels and across different examination contexts. The findings from this study indicated that, as expected, high achieving students made better answer choices and these choices were reflected in more beneficial outcomes while the low achieving students had behaviors that resulted in an increased level of damaging outcomes.

When students were placed in an online exam format due to the COVID 19 pandemic-related constraints, students in low achieving group were more likely to exhibit behaviors that decreased the beneficial outcomes and increased the damaging ones. Finally, we found that during the online examination students in both achievement groups made more maximum choices by exam questions which can be explained through a combination of a potential increase in the level of anxiety and less distractions during the exam compared to the face-to-face context. However, additional research is needed to analyze the factors that impact, positively or negatively, assessment in a remote testing format.

References

ACPE – Accreditation Council for Pharmacy Education (n.d.). https://www.acpe-accredit.org/pharmd-program-accreditation/

Birenbaum, M. & Alhija, F. N-A. (2013). Self-Efficacy Appraisals and Test-Taking Behavior of Students from Culturally Diverse Population. *ISRN Education*. Article ID 839657, 6 pages, http://dx.doi.org/10.1155/2013/839657.

Cernusca, D. & Friesner, D. (2019). Do Specific Behavioral Patterns in High Stakes Online Testing Environments Lead to Better Test Outcomes? Empirical Evidence from a Doctor of Pharmacy Program. *MBAA International Conference*, Chicago, IL, March 27-29.

Heinonen, J., Aro, T, Ahonen, T., & Poikkeus, A-M., (2011). Test-Taking Behaviors in a Neurocognitive Assessment: Associations with School-Age Outcomes in a Finnish Longitudinal Follow-Up. *Psychological Assessment*, 23(1), 184-192, DOI: 10.1037/a0021291.

Olev, M. & Must, A. (2013). Changes in Test-Taking Patterns Over Time. *Intelligence,* 41 (6), 780-790, http://dx.doi.org/10.1016/j.intell.2013.04.005.

Plake, B.S. & Wise, S.L. (1988). Test-Taking Behavior Under Formula and Number-Right Scoring Conditions. *Bulletin of the Psychonomic Society*, 26(4), 316-318, DOI:10.3758/BF03337668

Pohl, S., Sudkamp, A., Hardt, K, Carstensen, C.H., & Weinert, S. (2016). Testing Students with Special Educational Needs in Large-Scale Assessments – Psychometric Properties of Test Scores and Associations with Test Taking Behavior. Frontiers in Psychology, 7, article 154, 14 pages, http://dx.doi.org/10.3389/fpsyg.2016.00154.

Tamowsky, H. J. (1977). *Test-taking Behavior of Adult Learners in a Formal Education Setting.* (Order No. 7808434) [Doctoral dissertation, The Pennsylvania State University]. ProQuest Dissertations & Theses Global. (302835508). Retrieved from https://ezproxy.lib.ndsu.nodak.edu/login?url=https://www.proquest.com/dissertations-theses/test-taking-behavior-adult-learners-formal/docview/302835508/se-2?accountid=6766