Comparing Student Performance on Proctored and Non-Proctored Exams in Online Psychology Courses

Lee William Daffin, Jr. and Ashley A. Jones
Washington State University

Abstract

As online education becomes a more popular and permanent option for obtaining an education after high school, it also raises questions as to the academic rigor of such classes and the academic integrity of the students taking the classes. The purpose of the current study is to explore the integrity issue and to investigate student performance on online examinations. Utilizing a sample of about 1,700 students who took online psychology classes of varying difficulty at Washington State University from spring 2015 to spring 2016, we found that students performed 10–20% better and took about twice as long on non-proctored versus proctored exams. The effect held when we compared our in-house proctoring service used during this time against ProctorU, used for one semester in fall 2012. To ensure the most robust design possible, we also rotated the proctored exam in each class at least once and then compared performance on an exam when it was proctored versus when the same exam was non-proctored. Results showed better performance when the exam was non-proctored than when it was proctored. Finally, since instructors changed over the four semesters our study ran, we wanted to ensure that the results were not due to differences in teaching style. This potential confounding variable was eliminated. We discuss possible reasons for the difference in performance, to include student academic misconduct, and offer suggestions for ensuring we have both academic rigor and integrity in online courses.

Keywords: cheating, online education, proctored exams, time measure

Comparing Student Performance on Proctored and Non-Proctored Exams in Online Psychology Courses

Online/distance learning has become an increasingly popular method of receiving a higher education, often serving an off-campus population who cannot attend class on campus. The platform allows for flexibility for the student (Stack, 2015), and as Hannay and Newvine (2006) found in a 22-question survey of 217 students taking undergraduate criminal justice courses, 88% of students said they take online classes because classroom course schedules do not always fit into their busy lives, while 20% said the choices of traditional classes are at times

limited. Furthermore, 90% of online students were found to read the required text compared to just 60% of their classroom counterparts, who apparently were waiting on their instructors to "feed" the information to them. Overall, the authors found that students in their sample earned higher grades, believed they learned more, thought exams were easier, spent more time on classes, found the text more useful, and perceived classes to be of higher quality, but overall they saw classes as harder in the online learning environment (Hannay & Newvine, 2006).

Despite the results presented above, a criticism of online learning is that academic rigor may not be up to par with classroom courses. With more and more universities developing online courses and degree programs, this criticism should in time disappear. There is also research dedicated to addressing this issue and to finding ways to effectively engage online students so that their experience is nearly identical to that of classroom students. A second criticism, and one that is currently being addressed, is that there is a lack of academic integrity in the online environment. We live in the age of the information superhighway and students can find whatever they want to on the Internet with just a click or two of a mouse. This makes researching for classes very easy and should enhance learning. Of course, the downside is that papers can be easily bought online from sites such as Course Hero, and when students take exams they can find the answers to questions, despite admonitions against using outside resources, such as the textbook, websites, or other students. Some websites even post publisher test-bank questions and answers. Possibly the most flagrant offense is the existence of companies willing to take the class for the student, because "Life is too short to spend on classes you have no interest in. Focus on whats really important in your life" (We Take Your Class, 2012). They guarantee an A or B or your money back. As a consequence, the student you think you are interacting with on a weekly basis may not actually be the student at all. This is a serious issue for academic integrity in the online environment.

So why do students cheat? Cizek (1999) identified common justifications students use: it's easy to do, they need to pass the class, the course is too hard, the instructor is unfair, no one cares about cheating, everyone else is cheating, and there's no time to study. Cheating can be classified in one of two ways (Bunn, Caudill, & Gropper, 1992). *Planned cheating* is premeditated and involves developing cheat sheets, plagiarizing a paper, or copying homework. *Panic cheating* occurs during a test when the student realizes he or she does not know the answers to the questions, which leads to wandering eyes and the copying of answers. Planned and panic cheating are both issues in the classroom environment, but planned cheating is more likely the only issue online since students complete exams in isolation. In terms of cheating online, Rowe (2004) said students may wait to take their exam so that they can get answers from other students, obtain unauthorized help during the exam, or make false claims so they can retake the exam. Other methods include setting up two computers (one for taking the exam and one for looking up answers), stating that computer problems prevented completion of the exam, and purchasing answers from the Internet (Moten et al., 2013).

This leads us to wonder how prevalent instances of cheating are in the online learning environment. Grijalva, Nowell, and Kerkvliet (2006) estimated that 3% of students cheated in a single online class and that this rate is not quantitatively different than instances of cheating in a traditional classroom. They conclude that as online education grows there is no reason to assume that academic dishonesty will grow as well. Likewise, Stuber-McEwen, Wiseley, and Hoggatt (2009) administered the Academic Dishonesty Survey to 225 students to determine the frequency and type of dishonest academic behaviors they engaged in. Students reported higher cheating

behavior in the traditional classroom than in the online setting, leading the authors to conclude, much like Grijalva, Nowell, and Kerkvliet (2006), that instances of academic misconduct among online students may not be as high as thought. In line with these studies, students' perception of the frequency of cheating in online classrooms versus face-to-face classrooms showed no significant difference (Spaulding, 2009).

It may be that the rate of cheating online is as low as these studies suggest or that most cheating online goes undetected—and so unpunished—leading to repeated performance of the act, and that the academic community grossly underestimates it (Haney & Clarke, 2007). In a hybrid information systems class consisting of 300 students, and in which online quizzes were used to reinforce text material and made up just 10% of the final grade, or 1% for each quiz, Milliron and Sandoe (2008) convicted 15% of the students for quiz cheating. The quizzes were low stakes, easy, and information was covered in a very fundamental way. Students were told explicitly that were not allowed to consult with others. Results showed that students still congregated both on and off campus to take the quizzes together.

Several studies have shown that students perform better on non-proctored exams than proctored exams. Using a sample of 120 students in a doctor of pharmacy program and randomly assigned into either an online self-study group with online proctored quizzes, online self-study group with non-proctored quizzes, or textbook-based self-study group with proctored quizzes, Wellman and Marcinkiewicz (2004) found that the online, non-proctored students had the highest average scores of the three groups. They were also found to have the lowest performance on learning based on their pre- versus posttest change score. Similar results were found in a study of 300 undergraduate students taking a cognitive achievement test. Students in the non-proctored group performed better than those in the proctored group, but also of interest was of the fact that there was no difference whether the exam was non-proctored online or administered via pencil-and-paper. The authors suggest that the mode of administration is not as important as a lack of proctoring (Carstairs & Myors, 2009). Furthermore, Richardson and North (2013) found the same pattern of higher non-proctored exam scores in a sample of 65 students scattered across four online business administration courses at both the undergraduate and graduate level, though the difference in exam performance was smaller for the graduate students.

Harmon and Lambrinos (2008) drew data from two online classes in a principles of macroeconomics course taught during summer 2004 and the same class taught again in summer 2005. Each course consisted of four exams: three non-proctored exams and one comprehensive final. The classes were identical in structure and content, with the exception that the 2004 class had a non-proctored final exam, and the 2005 class had a proctored final exam. Results indicated significantly higher scores in the non-proctored test group versus the proctored test group, and the researchers took this as a case of cheating, since human capital variables, such as age, class rank, and GPA, explained test score variation less effectively in the non-proctored format than the proctored format. In contrast, Hollister and Berenson (2009) found a greater variability in test scores among a non-proctored group versus a proctored group, though upon further analysis the researchers found no evidence of cheating.

Stack (2015) found no significant difference between tests administered in a proctored setting at a university and tests administered via a lockdown browser without a proctor. Students in the lockdown browser group were given randomized test questions and were unable to backtrack to previous questions. This finding is important, as it could help solve the problem of needing to have all tests be proctored in order to ensure academic integrity. Perhaps a lockdown

browser, randomized test questions, and the inability to backtrack could be sufficient to reduce academic dishonesty.

Time is another important factor related to exam scores for proctored versus non-proctored tests online. Students who take longer to complete exams may be engaging in forms of academic dishonesty, such as looking up answers or consulting with a peer. A study by Hylton, Levy, and Dringus (2016) randomly assigned students to either a treatment or control group. Both groups were enrolled in the same course at a private university in Jamaica and took their exams online from the same pool of questions. The treatment group was monitored via webcam by a Web-based proctor, while the control group was not monitored at all. The results showed that students who took the non-proctored exam scored significantly higher than those taking the proctored exam and that the non-proctored students took significantly longer to complete the exam. The researchers also learned that students who were not monitored perceived that they had a greater chance to collaborate, and those who were monitored felt a greater level of deterrence from cheating.

So the studies above produce mixed findings about whether students are more likely to cheat online or not. Some indicate that the rate of cheating is no higher online than in the classroom, with classroom being higher at times (Grijalva et al., 2006; Stuber-McEwen et al., 2009; Spaulding, 2009). Other studies show that students perform better when taking an online non-proctored exam than when taking an online proctored exam (Milliron & Sandoe, 2008; Wellman & Marcinkiewicz, 2004; Richardson & North, 2013; Harmon & Lambrinos, 2008; Stack, 2015; Hylton et al., 2016).

The studies mentioned have several common limitations. First, generalizability is an issue, as many studies had very small sample sizes or samples from specific demographics of the population. Most studies have included no more than a few hundred students from one or two sections of a class. Second, cheating was assessed via self-report, making social desirability bias a real issue. Third, in most studies, only one class from that field was utilized. Classes within a major vary in difficulty level, so more difficult classes may, by their nature, lead to a greater level of student cheating than easier classes. Fourth, the content tested in each exam will likely vary, with some exams being more difficult than others, such as in an introductory psychology course. The biological basis of behavior chapter is usually more challenging for students than the chapter on human development. As such, students may be led to cheat more on the former exam than the latter. This limitation can be eliminated by simply rotating which exam is proctored across semesters.

The purpose of this study is to add to the body of literature investigating performance differences between proctored and non-proctored exams, and to expand on this research, much like Hylton et al. (2016), by offering an objective measure of why such differences exist. We hypothesize that students will perform better on non-proctored exams than on proctored exams, and that students will take more time to complete non-proctored exams than proctored exams.

Methods

Participants

Our study included data from 1,694 students taking online classes at Washington State University (WSU) in Pullman, WA. Data were collected from the spring 2015 semester to the spring 2016 semester and included students taking classes over the compressed, 12-week

summer 2015 session. During the fall and spring semesters, which last 15 weeks, online classes are restricted to Global Campus students unless a course does not fill to capacity, but during the summer, students from the Pullman, Tri-Cities, Vancouver, and/or Spokane campuses take online courses too.

Typically, our online students are working adults with families, are in their 30s or 40s, have been out of school for a period of time, and attend half- to full-time, all while juggling their many life responsibilities. This is contrasted with campus students, who are generally coming straight out of high school; attending school full-time; juggling extra-curricular activities, campus life, and work; and are away from home for the first time in their life.

Materials

The only materials used in this study were course exams, most of which were in multiple-choice format. In a few classes, short answer and matching questions were used too. Our online program includes a diverse range of courses at all academic levels, from freshman to senior, and all difficulty levels, from foundational courses (such as PSYCH 105) to more advanced courses (such as PSYCH 491). The list of courses included in this study are described in Table 1.

#	Catalog Title	Study n	Catalog Description
105	Introductory Psychology	135	Survey of the basic terms, processes, principles, and theories related to the scientific study of human behavior. Offered: Fall, Spring, Summer
230	Human Sexuality	132	Sexuality in personal development; personal, cultural, biological influences on sexual identification and behavior; fertility, reproduction, sexual functioning, sexuality and personality. (Crosslisted course offered as PSYCH 230, WOMEN ST 230) Offered: Fall, Spring, Summer
265	Biopsychological Effects of Alcohol and Other Drugs	85	Biopsychological effects of the major classes of abused and psychotherapeutic drugs, including alcohol, stimulants, sedatives and hallucinogens. Offered: Fall and Spring
321	Introduction to Personality	225	Theories, concepts, methods, discoveries in psychology of personality. Offered: Fall, Spring, Summer
324	Psychology of Gender	223	Contemporary overview of the psychological theory and research on sex and gender. (Crosslisted course offered as PSYCH 324, WOMEN ST 324) Offered: Fall, Spring, Summer
333	Abnormal Psychology	225	Problems of abnormality from traditional and evolving points of view; types, therapies, outcomes, preventive techniques. Offered: Fall, Spring, Summer

350	Social Psychology	263	Attitude changes, conformity, interpersonal relations, groups and social influences explored to give a coherent view of social psychology. (Crosslisted course offered as PSYCH 350, SOC 350). Offered: Fall, Spring, Summer
361	Principles of Developmental Psychology	140	Psychological processes of aging; changes in sensory, motor, cognitive, motivational and personality characteristics; research methodologies for the study of aging Offered: Fall, Spring, Summer
363	Psychology of Aging	17	Psychological processes of aging; changes in sensory, motor, cognitive, motivational and personality characteristics; research methodologies for the study of aging. Offered: Summer
372	Biological Basis of Behavior	35	Functional relationship between nervous system and behavior; integrated organ systems, sensory processes, and investigative procedures. Offered: Spring
401-Pre	Historical Development of Psychology	31	Concepts, methods, theories, trends, and systems.
401-Post		55	Offered: Fall, Spring, Summer
464	Behavior Disorders of Children and Adolescents	12	Theoretical and empirical approaches to the description, etiology, and treatment of behavior disorders in children and adolescents. Offered: Summer
490	Cognition and Memory	91	Human information processing, memory, and cognition. Offered: Fall, Summer
491	Principles of Learning	25	Principles of learning from a behavioral perspective using the experimental analysis of behavior. Offered: Spring

Table 1. Descriptions of Courses Offered Online in the Psychology Department at Washington State University

The course designs held constant during the 2015–2016 year (with the exception of PSYCH 401, and we separated the data for this class into pre- and post-design change), while the instructor of record for each course changed at least once. In general, the same exam was proctored during the three semesters making up 2015, and then in the spring 2016 we switched which exam was proctored.

Procedure

All online courses were taught in Blackboard Learn with the exception of one semester of data. Course instructors were graduate students in at least their second year of study at WSU. They were drawn from both our clinical and experimental PhD programs. A few courses were taught by department faculty, to include the first author of this study. Instructors were not

informed of the details of this study so that they could not influence their students' performance on exams.

In each online class, one exam is proctored and the remaining exams are non-proctored. Proctored exams are administered by WSU's in-house proctoring service, Global Campus Proctoring Service (GCPS). Students sign up for a time during the exam period, which runs Monday to Sunday in all classes. This gives students the opportunity to find a day and time that fits their busy schedule and which fits the hours during which GCPS is open. When their exam day and time arrives, students contact GCPS, provide proof of identity, and then use their webcam to show their testing space. The student takes the exam free of interference from the proctor unless an issue arises. When finished, the student informs the proctor, and the session ends. All proctored exams are password protected, and only GCPS and the instructor have this password. Hence, a student cannot preview an exam on his or her own time. Proctored exams include the following instructions: "The exam is closed book, with no aids/notes. A dry erase white board is allowed for making notes. The proctor will verify that the student has wiped the board clean upon exam completion." In terms of non-proctored exams, students can take the exam again anytime during the exam period. The same guidelines prohibiting the use of outside materials, people, or books apply, as stated in the course syllabus.

Data were collected at the end of each semester after all final grades were submitted by downloading the gradebook in each class. The names of the individual students were omitted from the data file analyzed via SPSS and were replaced by a participant number. Also, any work unrelated to exams was deleted from the final data file.

Analysis

We utilized a series of paired-samples *t*-tests to investigate several hypotheses. First, we examined differences in student performance on proctored and non-proctored exams based on percentage earned and the time it took to take the exam. We hypothesized that students would earn a higher score on non-proctored exams than on proctored exams and that students would take more time to complete non-proctored exams than proctored exams.

To rule out the possibility that poor performance is somehow due to the in-house proctoring service currently in use, we examined student exam performance from the fall 2012 semester, when ProctorU was utilized. We hypothesized that students would earn a higher grade on non-proctored exams than proctored exams and analyzed exam scores from four psychology courses offered at the time: 105, 321, 324, and 350.

We next wondered how students performed on an individual exam when it was non-proctored one semester and then became proctored another semester. To this end, we looked at the specific exams in each class that underwent a switch. We hypothesized that when an exam switched from proctored to non-proctored, student performance would increase significantly. Similarly, we hypothesized that if an exam was initially non-proctored, student performance would decrease significantly when it became proctored.

Lastly, we wanted to ensure that the performance was not linked to the instructor of record for each class. It is possible that some online instructors do a better job of explaining concepts in the discussion board, which then leads students to a clearer understanding of the material and higher exam grades. By examining performance for each exam in the class across semesters, we could see where differences might lie and then explain whether this is attributable

to the instructor and not potential cheating. To accomplish this, we used one-way ANOVAs with Tukey's post hoc test and had two hypotheses in relation to exams, semesters, and percentages. First, we hypothesized that exam differences would exist only—or mostly—with exams that had a shift from non-proctored to proctored. Those without a shift would not be significantly different from one another. Second, for semesters, most differences would lie with spring 2016 and all or most of the 2015 semesters, as the greatest number of shifts in the most classes occurred during spring 2016.

Results

Analysis Level 1: Testing Exam Score and Time

We wanted to first see how students performed on proctored and non-proctored exams in online psychology classes at WSU during the four semesters from spring 2015 to spring 2016. A paired samples t-test revealed that the mean percentage earned on proctored exams ($M = 64.37$, $SD = 15.47$) was significantly lower than the percentage earned on non-proctored exams ($M = 77.12$, $SD = 11.45$), $t(1693) = -37.19$, $p = .000$, d = 0.90, 95% CI [-13.42, -12.08], thereby supporting our hypothesis. This result held not only across all online psychology courses pooled together in one analysis but also across individual courses analyzed alone. It would be expected, and was found, that in some classes the difference between proctored and non-proctored exam was greater than in other classes and that this is a function of class difficulty. Please see Table 2.

Class	n	Proctored Exams		Non-Proctored Exams		M_D	t	df	d	95% CI
		M	SD	M	SD					
All Classes	1694	64.37	15.47	77.12	11.45	-12.75	-37.19	1693	0.90	[13.42, -12.08]
105	135	67.11	14.52	73.17	13.16	-6.06	-6.89	134	0.59	[-7.80, -4.32]
230	132	63.77	15.36	77.80	10.94	-14.03	-10.58	131	0.92	[-16.65, -11.40]
265	85	61.67	11.82	73.63	9.69	-11.96	-12.35	84	1.34	[-13.88, -10.03]
321	225	73.03	12.97	80.66	9.68	-7.63	-9.50	224	0.63	[-9.21, -6.05]
324	223	68.47	16.21	75.88	11.74	-7.41	-7.89	222	0.53	[-9.26, -5.56]
333	225	59.96	11.33	77.79	11.95	-17.83	-19.49	224	1.30	[-19.64, -16.03]
350	263	58.68	13.98	73.95	10.99	-15.27	-20.21	262	1.25	[-16.76, -13.79]
361	140	72.15	12.50	82.19	9.23	-10.04	-12.11	139	1.02	[-11.68, -8.40]
363	17	58.35	14.62	77.76	6.19	-19.41	-7.62	16	1.85	[-24.81, -14.01]
372	35	58.17	16.08	75.56	11.72	-17.39	-8.41	34	1.42	[-21.60, -13.19]
401-Pre	31	53.16	19.94	80.31	12.19	-27.15	-7.91	30	1.42	[-34.16, -20.15]
401-Post	55	60.52	18.65	76.18	10.52	-15.66	-7.39	54	1.00	[-19.91, -11.41]
464	12	63.00	11.58	79.17	11.12	-16.17	-4.15	11	1.20	[-24.74, -7.59]
490	91	59.78	18.22	79.46	12.85	-19.68	-9.87	90	1.04	[-23.64, -15.72]
491	25	56.25	18.69	75.57	10.32	-19.32	-5.28	24	1.06	[-26.87, -11.76]

Note. All p values are < .01.
Table 2. Differences in Student Performance on Proctored vs. Non-Proctored Exams in Online Psychology Course

Secondly, we wanted to know if there was a difference in the amount of time it took students to complete their exams, whether it was proctored or non-proctored, and so obtained the average time of completion for the class from Blackboard under Item Analysis. In most online courses, students were given 60 minutes to complete exams. A paired samples *t*-test revealed that the mean time taken on proctored exams (*M* = 27.69, *SD* = 5.17) was significantly lower than the time taken on non-proctored exams (*M* = 48.18, *SD* = 5.13), *t*(38) = −17.55, *p* = .000, d = 2.81, 95% CI [−22.85, −18.13]. This supported our hypothesis and showed that even though students were given 60 minutes to take all exams, they averaged 20 minutes more time on non-proctored than proctored exams. Please see Table 3.

Class	*n*	Proctored Time (minutes)	Non-Proctored Time (minutes)	M_D
All Classes	1694	27.69	48.18	-20.49
105	135	25.75	46.50	-20.75
230	132	21.67	45.75	-24.08
265	85	21.00	52.44	-31.44
321	225	24.00	43.75	-19.75
324	223	28.25	48.08	-19.83
333	225	29.00	56.13	-27.13
350	263	28.80	46.27	-17.47
361	140	28.00	49.50	-21.50
363	17	24.00	40.33	-16.33
372	35	23.00	50.67	-27.67
401-Pre	31	30.00	51.00	-21.00
401-Post	55	35.33	56.42	-21.08
464	12	43.00	37.50	-5.50
490	91	30.33	47.17	-16.83
491	25	35.00	48.67	-13.67

Table 3. Mean Time Taken on Proctored vs. Non-Proctored Exams in Online Psychology Courses

Analysis Level 2: Proctoring Service

A paired samples *t*-test revealed that the mean percentage earned on proctored exams (*M* = 65.37, *SD* = 14.73) was significantly lower than the percentage earned on non-proctored exams (*M* = 81.91, *SD* = 10.17), *t*(258) = −19.25, *p* = .000, d = 1.20, 95% CI [−18.22, −14.84], thereby confirming our hypothesis. Please see Table 4.

Class	*n*	Proctored Exams		Non-Proctored Exams		*t*	*df*	*d*	95% CI
		M	*SD*	*M*	*SD*				
All Classes	259	65.37	14.73	81.91	10.17	-19.25	258	1.20	[-18.22, -14.84]
105	46	57.48	12.25	73.43	12.40	-11.22	45	1.65	[-18.82, -13.09]
321	70	57.54	14.92	85.75	6.87	-16.69	69	1.99	[-31.58, -24.83]
324	77	69.04	10.54	82.62	9.67	-12.35	76	1.41	[-15.77, -11.39]
350	66	74.91	13.14	82.90	8.73	-5.46	65	0.64	[-10.91, -5.07]

Note. All *p* values are < .01.

Table 4. Differences in Student Performance on Proctored vs. Non-Proctored Exams in Select Online Psychology Courses Taken During Fall 2012

The results for exam percentage were identical to the 2015–2016 data, suggesting that the proctoring service used was not a factor in student performance.

Analysis Level 2: Exam Difficulty Affecting Student Performance

Table 5 shows that when an exam was proctored, students performed worse than when it was non-proctored, as predicted. This was the case in all psychology classes for the four semesters under study. We did not include PSYCH 361, 372, 401 Pre, 464, and 491 since we had only one semester of data to examine.

Class	n	Exam	When Proctored		When Non-Proctored		M_D	$S_{(M1-M2)}$	t	p	df	95% CI
			M	SD	M	SD						
105	135	2	67.96	15.21	77.98	10.50	-10.02	2.72	-3.69	.000	133	[-15.40, -4.65]
		4	64.85	12.41	72.45	15.80	-7.61	2.90	-2.62	.010	128	[-13.35, -1.87]
230	132	5	63.50	16.25	69.95	13.90	-6.45	2.86	-2.25	.026	130	[-12.12, -0.79]
		2	64.31	13.57	85.18	12.34	-20.87	2.36	-8.86	.000	130	[-25.54, -16.21]
265	85	2	64.92	11.74	72.43	11.31	-7.51	2.51	-3.00	.004	83	[-12.50, -2.53]
		3	58.91	11.29	76.31	12.50	-17.39	2.58	-6.74	.000	83	[-22.53, -12.26]
321	225	3	73.88	13.29	87.25	10.48	-13.37	1.86	-7.21	.000	223	[-17.03, -9.72]
		4	70.91	11.97	83.89	12.27	-12.98	1.80	-7.21	.000	223	[-16.53, -9.43]
324	223	4	70.58	15.55	78.41	13.83	-7.83	2.37	-3.30	.001	221	[-12.50, -3.16]
		2	61.85	16.60	72.66	16.30	-10.81	2.56	-4.22	.000	221	[-15.86, -5.77]
333	225	3	59.29	10.74	83.36	12.69	-24.07	1.71	-14.08	.000	223	[-27.44, -20.70]
		2	61.85	12.76	78.19	12.68	-16.34	1.92	-8.49	.000	223	[-20.14, -12.55]
350	263	2	57.37	13.99	69.90	12.93	-12.53	2.05	-6.12	.000	261	[-16.56, -8.50]
		4	63.31	13.03	74.88	12.97	-11.57	1.93	-5.98	.000	258	[-15.38, -7.76]
361	140	3	72.51	12.38	82.59	8.39	-10.08	2.56	-3.94	.000	138	[-15.13, -5.03]
		4	70.56	13.14	81.43	10.79	-10.87	2.45	-4.44	.000	137	[-15.72, -6.03]
401-Post	55	3	59.76	18.44	77.21	11.27	-17.45	4.04	-4.32	.000	53	[-25.54, -9.35]
		4	60.86	18.98	82.67	16.86	-21.82	5.36	-4.07	.000	53	[-32.56, -11.07]
490	91	3	59.02	19.03	74.69	17.49	-15.67	3.93	-3.99	.000	89	[-23.48, -7.86]
		1	60.88	17.16	76.23	13.09	-15.34	3.17	-4.83	.000	89	[-21.65, -9.04]

Note: All p values are < .01.
Table 5. Examining Differences in Student Performance on the Proctored and Non-Proctored Versions of Class Exams

Analysis Level 2: Instructor of Record Affecting Student Performance

The results of one-way ANOVAs with Tukey's post hoc test showed that differences only occurred with the exams for which a switch from non-proctored to proctored occurred. This confirmed our hypothesis. Also, in all but two online courses, only the exams which had a switch showed significant differences across semesters. The two classes in which more than the two exams were significant included PSYCH 265 and PSYCH 361. The result from PSYCH 265 is not surprising. The same instructor taught the class each semester and routinely modified his exams by adding new questions. He made the decision as to which questions to keep and which

to discard based on item analysis of each exam. Each exam changed slightly from semester to semester; hence, three of the four exams were significantly different across this period. As for PSYCH 361, four of the five exams came up significantly different. Exams 3 and 4 experienced the switch, and so significance was to be expected. Exams 1 and 2 did not switch; therefore, significance was unexpected.

Interestingly, when the post hoc tests revealed differences between semesters other than spring 2016 and another, it was always between summer 2015 and another 2015 semester. This was the case for Exam 2 in PSYCH 361 and can be easily explained: WSU students taking online classes in the summer semester of any year tend to be of a different demographic. Many online students take the summer off, while on-campus students, in turn, take online classes. This affords on-campus students the opportunity to go home but still make progress toward completing their degree.

Discussion

Main Findings and Previous Research

Consistent with previous research and in support of our hypotheses, we found that students performed significantly worse on proctored exams than non-proctored exams (Wellman & Marcinkiewicz, 2004; Harmon & Lambrinos, 2008; Hollister & Berenson, 2009; Richardson & North, 2013; Stack, 2015). In spring 2016, the proctored exam in each PSYCH class was switched, meaning the results held across 14 classes, four semesters, and at least one proctored exam switch. Switching proctored exams was an important manipulation in this study. It could be that we unintentionally selected an exam to be proctored that was generally more difficult for students, resulting in poor performance and making it appear that students did better on non-proctored exams. The main take-home point is this: an exam students did poorly on when proctored transformed into a greatly improved performance when non-proctored. Likewise, an exam that students did well on when non-proctored became a blemish on their overall class performance when it was proctored.

This leads to the question of why this difference exists. It could be due to students' general anxiety about taking an exam, which is further exacerbated by being watched while they do so. Or it could be due to the fact that while students are taking their exam under the vigilant eye of a proctor, they cannot use notes, Internet sources, other classmates, and/or their textbook, which they may use in a non-proctored exam despite instructions not to do so. We also found that students took almost twice as long to complete non-proctored exams, much like the findings of Hylton et al. (2016). Students might be taking advantage of not being monitored during all but one exam in their class, but this cannot be confirmed.

Strengths and Limitations

This study stands apart from past research for several reasons. First, our data set included a sample of approximately 1,700 students taking online psychology courses over four semesters, representing the largest sample size and spanning the longest period to date. Second, a wide array of classes of varying difficulty levels were used. Third, a comparison of two proctoring services was undertaken to ensure the effect was not just an artifact of our in-house proctoring service. Fourth, we manipulated which exam was proctored to see if the effect held. Fifth, we explored whether shifts in course instructor could be the cause of our differences in exam

performance and were able to account for and offer explanations when differences emerged. We believe these five strengths compensated for the limitations of past research, such as small sample sizes, using only one class, restricted class types, and varying exam difficulty due to content tested.

Of course, our study is not without limitations. First, our focus was only on students taking online psychology classes and offers no comparison with classroom psychology courses or courses in other fields. Second, our sample consisted of online students, who mostly represent a different demographic than classroom students. Third, though we did compare two proctoring services, our study with ProctorU utilized data from almost three years before the main study and included just four classes. We need to better generalize the effect across proctoring services. Finally, we did randomize which exam was proctored in all classes, but we need to do so in more than just one semester.

Implications and Future Directions

Hylton et al. (2016) reported that students who were not monitored felt that they had more of a chance to collaborate with others and use unauthorized resources, while those who were monitored by a Web-based proctor felt that they could not engage in misconduct while taking their exam. Though we did not ask students for their impressions of proctored and non-proctored testing conditions, our results showed that students did earn higher scores on non-proctored exams and took almost twice as long as they did on proctored exams. Importantly, this effect held when an individual exam switched from proctored to non-proctored, or vice versa.

We need to realize that students have the potential to use their textbooks and notes when taking online exams, so we need to design courses with the risk of academic dishonesty in mind (Trenholm, 2007). How so? We have to establish clear course learning and behavioral objectives, show students what a class can offer them, explain tracking tools used to monitor activity in the course, create policies for Internet service provider (ISP) crashes, and create a large test bank to draw questions from (Christe, 2003). Additionally, we need to include essay or short answer questions and not just multiple-choice style questions, make sure students are aware of the policy on and penalties for cheating, give shorter time frames for completing exams, and randomize questions and answer choices (Moten et al., 2013). Milliron and Sandoe (2008) suggest verifying the test taker's identity and reshaping the attitudes and perceptions of students as they relate to cheating and assessment. Stiff penalties for cheating could be implemented at the university level in terms of permanent notations on the student's record (Kitahara, Westfall, & Mankelwicz, 2001) and at the course level in terms of an F for the assignment/exam or an F in the course. It is important to note that the perceived *severity* of sanctions can be just as effective at reducing dishonest behavior as the *certainty* of sanctions (D'Arcy, Hovav, & Galletta, 2008), further supporting the need to be clear on such policies and to educate students on them.

There is a justifiable need to proctor a minimum of one online exam each semester, and at best, all online exams in a class, just as would be the case in a classroom course. The issue with this is that there is a cost for each exam that is proctored, with the average being about $20 per exam. With the growing cost of textbooks, it is prudent for universities not to further burden students financially by charging what may seem (to the university) like a modest fee for proctoring exams. Of course, in light of research such as ours, universities could justify it as a cost of doing business, but at the same time, we do not want to make education inaccessible for those who are already financially challenged.

Another possibility suggested in the literature is the use of code-of-conduct or honor statements (Christe, 2003; Milliron & Sandoe, 2008). Most universities have already developed such statements, and course instructors and designers could add them to the beginning of each exam as a reminder to students to behave ethically. Inclusion of such statements does not cost the student or instructor anything in terms of money and may be just as effective as proctoring an exam. This is supported in the research of Ariely and colleagues, who showed that asking participants to recall the Ten Commandments prior to doing a task or being reminded of an honor pledge (Mazar et al., 2008), signing an honesty pledge at the top rather than the bottom of a page (Shu et al., 2012), and presenting participants with an out-group cheater (Gino et al., 2009), all significantly reduced cheating and can be used to revise current intervention and policy (Ayal et al., 2015). Hence, a viable future direction, and one that our research team is currently pursuing, is to investigate the utility of conduct statements for reducing cheating during online proctored exams.

A final possibility is to make online exams open notes/book from the start but increase the difficulty of such exams so that they are not simply testing the recollection of facts (Feller, 1994; Williams & Wong, 2009; Stowell, 2015). Though students would be permitted to utilize outside sources, eliminating student misconduct issues, they would still need a good understanding of the material to be able to accurately apply it and could not simply look up answers as they took the exam.

Conclusion

In sum, the current study adds to the growing body of literature showing that students perform better when an online exam is not proctored. Cheating is one explanation for this behavior, but so is student-reported anxiety about test taking, whether the exam is proctored or not. Many strategies can be undertaken to ensure academic integrity in the growing area of online education if we, as educators, are willing to put the time into course development.

References

Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science, 10*(6), 738–741.

Bunn, D. N., Caudill, S. B., & Gropper, D. M. (1992). Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *The Journal of Economic Education, 23*(3), 197–207.

Carstairs, J., & Myors, B. (2009). Internet testing: A natural experiment reveals test score inflation on a high-stakes, unproctored cognitive test. *Computers in Human Behavior, 25*(3), 738–742.

Christe, B. (2003). Designing online courses to discourage dishonesty. *Educause Quarterly, 26*(4), 54–58.

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Routledge.

D'Arcy, J., & Devaraj, S. (2012). Employee misuse of information technology resources: Testing a contemporary deterrence model. *Decision Sciences, 43*(6), 1091–1124.

Feller, M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation, 20*, 235–238.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior the effect of one bad apple on the barrel. *Psychological Science, 20*(3), 393–398.

Grijalva, T. C., Nowell, C., & Kerkvliet, J. (2006). Academic honesty and online courses. *College Student Journal, 40*(1), 180–185.

Haney, W. M., & Clarke, M. J. (2007). Cheating on tests: Prevalence, detection and implications for online testing. *Psychology of Academic Cheating,* 255–287.

Hannay, M., & Newvine, T. (2006). Perceptions of distance learning: A comparison of online and traditional learning. *Journal of Online Learning and Teaching, 2*(1)*,* 1–11.

Harmon, O. R., & Lambrinos, J. (2008). Are online exams an invitation to cheat? *Journal of Economic Education,* 116–125.

Hollister, K. K., & Berenson, M. L. (2009). Proctored versus unproctored online exams: Studying the impact of exam environment on student performance. *Decision Sciences Journal of Innovative Education, 7*(1), 271–294.

Hylton, K., Levy, Y., & Dringus, L. P. (2016). Utilizing webcam-based proctoring to deter misconduct in online exams. *Computers & Education, 92-93*, 53–63.

Kitahara, R., Westfall, F., & Mankelwicz, J. (2011). New, multi-faceted hybrid approaches to ensuring academic integrity. *Journal of Academic and Business Ethics, 3*, 1.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644.

Milliron, V., & Sandoe, K. (2008). The net generation cheating challenge. *Innovate: Journal of Online Education, 4*(6), 3.

Moten, J., Jr., Fitterer, A., Brazier, E., Leonard, J., & Brown, A. (2013). Examining online college cyber cheating methods and prevention measures. *The Electronic Journal of e-Learning, 11*, 139–146.

Richardson, R., & North, M. (2013). Strengthening the trust in online courses: A common sense approach. *Journal of Computing Sciences in Colleges, 28*(5), 266–272.

Rowe, N. C. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration, 7*(2).

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*(38), 15197–15200.

Spaulding, M. (2009). Perceptions of academic honesty in online vs. face-to-face classrooms. *Journal of Interactive Online Learning, 8*(3), 183–198.

Stack, S. (2015). The impact of exam environments on student test scores in online courses. *Journal of Criminal Justice Education, 26*(3), 273–282.

Stowell, J. R. (2015). Online open-book testing in face-to-face classes. *Scholarship of Teaching and Learning in Psychology, 1*(1), 7–13.

Stuber-McEwen, D., Wiseley, P., & Hoggatt, S. (2009). Point, click, and cheat: Frequency and type of academic dishonesty in the virtual classroom. *Online Journal of Distance Learning Administration, 12*(3).

Trenholm, S. (2007). A review of cheating in fully asynchronous online courses: A math or fact-based course perspective. *Journal of Educational Technology Systems, 35*(3), 281–300.

We Take Your Class. (2012). Home. Retrieved from https://www.takeyourclass.com/

Wellman, G. S., & Marcinkiewicz, H. (2004). Online learning and time-on-task: Impact of proctored vs. un-proctored testing. *Journal of Asynchronous Learning Networks, 8*(4), 93–104.

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology, 40*, 227–236.