Alessio, H., & Maurer, K. (2018). The impact of video proctoring in online classes. *Journal on Excellence in College Teaching*, *29*(3&4), x-x.

The Impact of Video Proctoring in Online Courses

Helaine Alessio
Karsten Maurer
Miami University

With the meteoric rise of online course offerings, the development and adoption of software tools has been integral to the quality of the educational experience. There are substantial concerns regarding academic integrity during online test taking, which sacrifices much of the instructor's control over the testing environment in favor of allowing students the asynchronous flexibility that is essential in courses administered online. Video proctoring software has been a key technology for returning some control over the testing environment to the instructor. The authors analyzed the change in grade distributions across 29 courses and instructors on a college campus before and after video proctoring was made widely available. They found that the average course grade point average scores were significantly reduced after the adoption of the proctoring software, suggesting that academic integrity is compromised when online exams are left unproctored.

Advances in technology, computers, internet access, and pedagogical tools have contributed to raising the quality of the educational experience of distance or online learning to match that of traditionally taught courses. A meta analysis by the U.S. Department of Education (2010) concluded that courses taught online and in a blended learning format were increasing in popularity and, if properly implemented, were as likely or more likely to promote student learning as traditional formats. This conclusion bodes well for the increased interest and acceptance of online education. The Online Learning Consortium reported that nearly 30% of all students in 2015 took at least one online course, which totalled over six million students, half of whom took all of their classes online (Online Consortium, 2017). While most college students enroll in traditional courses, online examinations are becoming more commonly employed in courses taught via distance, blended, and face-to-face classrooms (Kemp & Grieve, 2014). Online exams are increasingly preferred by faculty regardless of class format due to the convenience of asynchronous exam taking (Billings, 2004), ease of grading, and benefits of immediate feedback (Butler, Pyzdrowski, Goodykoontz, & Walker, 2007). Online assessments provide flexibility; however, faculty must ensure an environment that promotes academic integrity and prevents cheating when students take exams asynchronously and out of sight of the instructor.

Actual cases of academic dishonesty have rarely been systematically reported, and claims of increased dishonesty in colleges and universities are not always supported in the literature (Brown & Emmett, 2001). Results and conclusions are mixed for research comparing student grades and academic integrity violations in face-to-face versus online class settings (Alessio, Malay, Maurer, Bailer, & Rubin, 2017; Beck, 2014; Harmon & Lambrinos, 2008; Watson & Sottile, 2010). Tollman's (2017) review of the literature challenges the perception that academic dishonesty occurs more frequently in online courses. However, Alessio et al. (2017) reported significant grade disparities—averaging 17 points in proctored compared versus unproctored online exams—in multiple sections of the same course, raising suspicions about student cheating when taking unproctored online exams. Cheating is difficult to document, especially on a large scale and across many different courses. Grade distributions can vary dramatically from one course to another and from one instructor to another; nevertheless, substantial grade disparities from one class to another, especially when some classes are proctored and others are not, may be used as a proxy for suspected cheating. On the other hand, if cheating is deterred, grade distributions between proctored and unproctored should be similar or shift toward lower scores when proctored.

When comparing results from several studies on grade differences in proctored versus unproctored exams in multiple sections of one online course, Alessio and colleagues (2017) discovered a pattern of differing grades between proctored and unproctored exams as well as between more and less difficult exams. This pattern emerged both between and within class sections, suggesting that powerful forces affect student behaviors and contribute to academic dishonesty. When students feel a need to cheat and the opportunity arises, which could occur in a non-proctored environment, exam grades are remarkably and suspiciously higher than exam grades in a proctored environment. The grade disparity between proctored and unproctored environments is exacerbated by exam difficulty level.

The Higher Education Opportunity Act of 2008 supports online learning for the flexibility and convenience it provides to college students and faculty. However, the Act states that institutions offering online courses must have in place effective strategies to ensure academic integrity and prevent cheating (The Higher Education Opportunity Act, 2008). In order to be effective, an institutional culture of academic integrity needs to be created and supported by a combination of stated commitments, including honor codes and integrity practices, and a system that supports programs to prevent cheating and encourage honest academic performance (Lee-Post & Hapke, 2017). The use of technology to prevent cheating and promote a fair test-taking environment has become a practical option for institutions, especially those in which online exams are used, regardless of the class format. Proctoring software that uses webcams to videotape students immediately prior to and during online exams provides authentication of the identity of the student and furnishes the instructor with feedback and video footage of suspicious behavior.

Proctoring software that has been implemented in specific classes has been found to have an effect on student grades, with significantly lower grades for students who were proctored than for those who were not (Alessio et al., 2017; Alessio et al., in press). Results of grade disparities when proctoring software is used in one or two classes have provided evidence of its utility in promoting academic integrity, although its effectiveness on a larger scale, when an institutional practice to prevent cheating has been implemented, needs to be determined. We compared grades in a variety of online courses offered throughout Miami University, a mid-sized public university in the Midwestern United States, one year before and one year after the campuswide implementation of proctoring software.

Methods

The protocol used in this study for the collection and examination of data was approved by the University's Internal Review Board in the Research Ethics and Integrity Office. Records of the courses and instructors who taught online classes in the academic years 2016-17 and 2017-2018 were provided by the Miami University Center for eLearning. All student names were removed from the data used for analysis. Final grade distributions were provided by the Office of the Registrar. By combining these data sources, we identified 29 cohorts in which the instructor had taught at least one section of the course without video proctoring, then used video proctoring in at least one other section.

We aimed to determine the effectiveness of an institutional change in policy that included implementing video proctoring software for all departments at no cost in order to deter cheating during online exams. Measuring such a definitionally clandestine activity directly is not possible, and differentiating between academically honest and dishonest students is similarly uncertain. We applied statistical methods to test whether final class grade point averages (GPAs) decreased within specific courses after implementing the video proctoring software, and we used statistical meta-analysis across many courses to determine whether grades in online courses were generally depressed after implementing video proctoring software.

The analytical units of interest, the *cohorts,* represented a group of students with the same course and instructor combination. For example, all students enrolled in Personal Nutrition (KNH 101) with Instructor A were considered cohort KNH101:A. Note that a cohort does not equate to a classroom, because the students in a cohort may have been from different enrollment sections; however, the assumption was that students taking the same course with the same instructor would be graded in a consistent manner.

We tested for a significant change in final class grades among a cohort after adding video proctoring. The course grade was converted from letter grades to GPA values using a standard conversion ("A+" or "A" = 4, "A-" = 3.7, "B+" = 3.3, "B" = 3, "B-" = 2.7, "C+" = 2.3, "C" = 2, "C-" = 1.7, "D+" = 1.3, "D" = 1, "D-" = 0.7, "F" = 0). We considered cases where the instructor had taught one or more sections of a course without proctoring, then added video proctoring software in additional sections. We partitioned the students into those who were proctored and those who were unproctored, then examed for a significant difference in final grades by running a two-sample *t* test assuming unequal sample sizes and unequal variances (Welch, 1947). We allowed for unequal sample sizes due to differences in section enrollment sizes, and we allowed for unequal variances in the case that proctoring impacted both the center *and* the spread of the grade distribution. We defined the null hypothesis as "no difference in average grade with or without video proctoring" and a one-sided alternative as "average grade lower with video proctoring." The test statistic for testing cohort was as follows:

$$t_k = (\bar{Y}_{1k} - \bar{Y}_{2k}) \,/\, s_k^2 \text{, where } s_k^2 = s_{1k}^2/n_{1k} + s_{2k}^2/n_{2k}$$

$\bar{Y}_{jk}$ is the sample average GPA, s^2_{jk} is the sample variance, and n_{jk} is the sample size for proctor status (1 = video proctoring; 2 = no video proctoring) for students of cohort *k.* Under the null hypothesis of no difference, the test statistic should be *t* distributed with degrees of freedom as provided by the Welch-Satterthwaite approximation (Satterthwaite, 1946; Welch, 1947):

$$df_k = s_k^2 / \left(\frac{(s_{1k}^2 / n_{1k})^2}{(n_{1k} - 1)} + \frac{(s_{2k}^2 / n_{2k})^2}{(n_{2k} - 1)} \right)$$

Comparison of the test statistic, t_k, to this assumed *t* distribution under the null hypothesis provides us with a p-value by which to evaluate our hypotheses. The test statistic also acts as the point estimate by which we can interpret the magnitude of the video proctor effect size. Additionally, corresponding 100 • (1-1/α)% confidence intervals for the difference in average GPA after implementing video proctoring can be constructed as follows:

$$t_k \pm t_{(1-\frac{1}{2\alpha}),df_k}\sqrt{s_k^2}$$

The two-sample *t* test described above could answer whether there is significant effect of video proctoring software on the average grade for a *specific* course and *specific* instructor. We also wished to draw inference more broadly to online courses and online course instructors *in general*. To do so, we can run the cohort comparison using the method detailed above for cohorts $k = 1, 2, \ldots, K$, then combine the results using the *inverse* χ^2 meta-analysis method (Fisher 1948; Piegorsch & Bailer, 2005). The inverse χ^2 method tests whether there is significant evidence across all cohorts that the average GPA is lower with video proctoring. Using the p values from the two-sample *t* tests on each of the cohorts, P_k, we can calculate a meta-analytic test statistic as follows:

$$\chi_{agg}^2 = \sum_{k=1}^{K} (-2 \cdot log(P_k))$$

That would be χ^2 distributed, with $2K$ degrees of freedom, under the null hypothesis of "no difference in average GPA with or without video proctoring."

Results

Grades from a total of 2686 students enrolled in 29 unique course/instructor cohorts representing 10 departments were analyzed in this study. Although final course grade distributions differed by course, and grade disparities before and after proctoring software was activated were dissimilar across departments, the overall average course grade for proctored students was 0.088 points lower than for unproctored students, a 2.2% drop in GPA on the 4-point scale. This difference is likely made smaller due to the inclusion of many honest students and non-exam activities that would not be influenced by the use of proctoring software.

Figure 1 displays the 95% confidence intervals for the change in average GPA after using video proctoring for all 29 cohorts. The intervals are displayed in descending order of effect magnitude. The interval widths vary substantially because the number of students per cohort ranges n_k = 24 to 308. Seven of the 29 cohorts had a significantly higher average GPA after implementing video proctoring, nine cohorts had no significant change, and 12 cohorts had a significantly lower average GPA after implementing video proctoring.

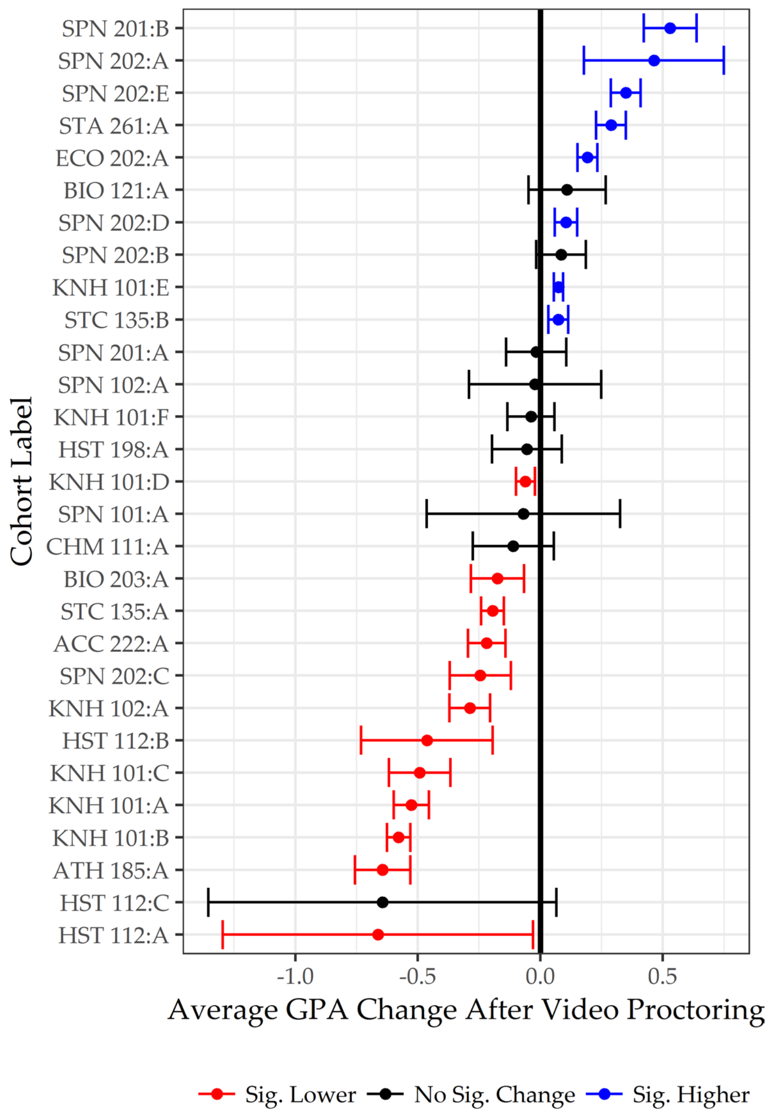
We then used the inverse χ^2 method to assess if the evidence across all cohorts suggested a more general trend in average GPA after implementing video proctoring software. We combined the 29 p values from *t* tests on each cohort into a meta-analytic test statistic: $\chi^2_{agg} = 95.20$. Compared to the distribution with 58 degrees of freedom, as expected under the null, we discovered a combined p-value = 0.0015. This result provides strong evidence that average class GPA across all cohorts was lower after implementing video proctoring software.

Discussion

Due to the challenges in detecting cheating behavior, this study's results are limited to comparing grade disparities in the same course and instructor before and after proctoring software was made available in a new campuswide policy aimed at deterring cheating and promoting academic honesty. A large and significant grade disparity of 17% in online exam scores was reported in Alessio et al.'s (2017) study of multiple sections of the same class taught by different instructors. In their investigation of one course, Medical Terminology, taken by students from a variety of majors, the time used to take an online exam was significantly longer when unproctored compared with exams that were proctored. These results raise suspicions that students left unproctored spent time looking up answers during the online exam, whereas proctored students did not, and consequently finished their exams in nearly half the time.

It is to be expected that different course content and structures as well as instructors using a variety of teaching and grading styles would affect final course grades differently in all cohorts. Student interest in the course material also appeared to play a role, with students in Spanish courses,

Figure 1
95% Confidence Intervals for the 29 Cohorts, Individually



for example, generally earning higher grades, and students in History and Kinesiology and Health courses (specifically, Nutrition courses) earning lower grades after adopting proctoring software. There were not enough other departments with multiple cohorts to compare this effect further. These results align with Beck's (2014) conclusion that multiple factors influence academic dishonesty including student major. Although academic dishonesty was not directly measured in this study, the significant grade decrease that occurred when video proctoring software was used in online exams provided evidence to raise serious suspicions that when not monitored, students taking exams online are likely to cheat. A study of nursing students reported that they admitted to being less likely to cheat when monitored with a webcam during online testing (Mirza & Staples, 2010).

A general decrease in final course grades when students were proctored compared with unproctored could be explained by the impact video proctoring software had on deterring cheating. This suggests that when the playing field is even and video proctoring software is implemented in all online exams, student performance will probably drop, and most significantly for those students inclined to cheat. Nevertheless, grades can be impacted by other factors, including changed exams and level of exam difficulty.

An honor code had been in place at Miami University for several years prior to the implementation of campuswide video proctoring software. Based on the grade disparities, it is clear that honor codes, alone, do not enforce academic integrity. This study is the first we are aware of to investigate the impact on final course grades of a campuswide policy implementing video proctoring software for all online exams.

Conclusions

Online testing is becoming increasingly popular as students and faculty reap the benefits of the flexibility associated with asynchronous learning. The ability to authenticate students taking online exams so that they are not cheating when not supervised by the instructor is critical to the assurance of fairness and integrity, however. Advances in technology designed to enhance academic integrity have increased over the years and include video proctoring software. The analyses of course grades in 29 online student cohorts representing 10 different departments before and after video proctoring software was provided to all departments at a midsized university indicate that significant grade disparities are likely to occur when comparing classes in which online exams were not proctored with those in which they were. These results add to the body of evidence urging that widespread adoption of controlled online testing environments using proctoring software is *necessary* to ensure academic integrity.

References

Alessio, H. M., Malay, N. J., Maurer, K., Bailer, A. J., & Rubin, B. (in press). Interaction of proctoring and student major on online test performance. *International Review of Research in Open and Distributed Learning.*

Alessio, H. M., Malay, N. J., Maurer, K., Bailer, A. J., & Rubin, B. (2017). Examining the effect of proctoring on online test scores. *Online Learning*, *21*(1), 146-161.

Billings, K. (2004). Online assessment: Perspectives of developers. *Media and Methods*, *40*(4), 26-27.

Brown, B. S., & Emmett, D. (2001). Explaining variations in the level of academic dishonesty in studies of college students: Some new evidence. *College Student Journal*, *35*(4), 529-539.

Butler, M., Pyzdrowski, L., Goodykoontz, A., & Walker, V. (2007). The effects of feedback on online quizzes. *International Journal for Technology in Mathematics Education*, *14*(4),131-136.

Fisher, R. A. (1948). Combining independent tests of significance. *American Statistician*, *2*(5), 30.

Harmon, O. R., & Lambrinos, J. (2008). Are online exams an invitation to cheat? *The Journal of Economic Education*, *39*(2), 116-125.

Online Consortium. (2017). New study: Over six million students now enrolled in distance education. Retrieved from https://onlinelearningconsortium.org/news_item/new-study-six-million-students-now-enrolled-distance-education/

Kemp, N., & Grieve, R. (2014). Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. *Frontiers in Psychology, 5,* 1-11. doi:10.3389/fpsyg.2014.01278

Lee-Post, A., & Hapke, H. (2017). Online learning integrity approaches: Current practices and future solutions. *Online Learning*, *21*(1), 135-145.

Mirza, N., & Staples, E. (2010). Webcam as a new invigilation method: Students' comfort and potential for cheating. *Journal of Nursing Education*, *49*(20), 116-119.

Piegorsch, W. W., & Bailer, A. J. (2005). *Analyzing environmental data*. New York, NY: John Wiley & Sons.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110-114.

The Higher Education Opportunity Act. (2008). Retrieved from https://www.gpo.gov/fdsys/pkg/PLAW-110publ315/pdf/PLAW-110publ315.pdf

Tolman, S. (2016). Academic dishonesty in online courses: Considerations for graduate preparatory programs in higher education. *College Student Journal*, *51*(4), 579-584.

U.S. Department of Education. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* Washington, DC: Office of Planning, Evaluation, and Policy Development.

Welch, B. L. (1947). The generalization of students' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28-35.

Helaine Alessio, *Ph.D., FACSM, is a Professor and Chair of the Department of Kinesiology and Health at Miami University. Her major discipline, Kinesiology, is a popular major in the department, and she teaches several courses, one of which is offered solely online. Her scholarship includes basic and applied exercise science in health and disease. The recent demand of online courses in her department provided an opportunity to explore the academic performance of students enrolled in multiple sections of an online class.* **Karsten Maurer** *is an Assistant Professor of Statistics at Miami University. He completed his doctoral degree in statistics at Iowa State University. His research interests are in the fields of predictive modeling, data visualization, and statistics education.*