Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 20, Number 10, April 2015 ISSN 1531-7714

On the Integrity of Online Testing for Introductory Statistics Courses: A Latent Variable Approach

Alan Fask, Fred Englander, Zhaobo Wang, *Fairleigh Dickinson University*

There has been a remarkable growth in distance learning courses in higher education. Despite indications that distance learning courses are more vulnerable to cheating behavior than traditional courses, there has been little research studying whether online exams facilitate a relatively greater level of cheating. This article examines this issue by developing an approach using a latent variable to measure student cheating. This latent variable is linked to both known student mastery related variables and variables unrelated to student mastery. Grade scores from a proctored final exam and an unproctored final exam are used to test for increased cheating behavior in the unproctored exam.

There has been a significant growth in the development and expansion of distance learning courses in higher education over the last ten to fifteen years. There has also been research evidence in this time period that a high proportion of college students are inclined to cheat on exams and other graded assignments and that the willingness of college students to cheat has increased in recent years. Despite indications that faculty and students alike believe that distance learning courses are more vulnerable to cheating behavior than traditional, face-to-face courses given that distance learning courses very often rely on unproctored, online exams rather than traditional, proctored exams, there has been relatively little research addressing the issue of whether this growth of distance learning facilitates a greater prevalence of student cheating. This paper presents a straightforward and minimally invasive method of determining if the use of online examinations in a given college setting facilitates a higher level of student cheating.[1] This approach allows, for the purposes of evaluating the academic integrity of a course or program, a judgment to be made as to whether a greater use of unproctored, online exams represents a threat to that integrity. Such an analytical method could well become valuable to a college or university (or any program or sub-division therein) faced not only with the desire to internally maintain control over the quality of its academic program and reputation but also faced with the obligation, under federal Higher Education Opportunity Act (2008), to avoid "separate procedures, or policies for the evaluation of distance education" relative to traditional education (§ 3325). Further, that legislation expresses concern about the possibilities that such distance education may be more susceptible to cheating by mandating oversight of colleges and

[1] For this paper, "cheating" is conceptually defined as any action which breaks explicit or implicit rules relating to the conditions under which an examination is taken. Such rules serve to prevent some students from gaining an unfair advantage relative to their peers. Operationally, "cheating" will be defined here as any statistically significant difference, directly or indirectly measured, by which the online test scores exceed the in-class test scores, when the tests, themselves, are equivalent. This is the operational definition of cheating utilized by Peng (2007), Harmon and Lambrinos (2008), Hollister and Berenson (2009) and Yates and Beaudrie (2009).

universities offering distance education such that "the institution establishes that the student who registers in a distance education or correspondence education course or program is the same student who participates in and completes the program and receives the academic credit (§ 3325)."

This article utilizes an approach based on the use of a latent variable for student cheating to evaluate the hypothesis that there is a greater relative presence of cheating in online exams than in proctored, in-class exams. The relationship of such a cheating variable to both known student mastery variables and variables unrelated to student mastery is examined. (In the present study, individual student grade point average (GPA) and student class attendance are included under the umbrella term of student mastery variables, i.e., variables that can reasonably be expected to increase the students' genuine comprehension and learning of the course material). Grade scores from a proctored, in-class final exam and an unproctored, online final exam are used to test for evidence of cheating behavior in the unproctored exam. The relationship of this methodology to previously suggested methods of detecting student cheating for online exams is discussed. A novel aspect of this article is that it provides an example in which a structural equation approach and a stacked regression approach may both be used to analyze the same latent variable problem. This article offers the opportunity to explore the relationships between both approaches. It is hoped that the statistical approach outlined in this article will provide a useful approach for instructors or administrators to determine if the use of online examinations is subject to a higher incidence of student cheating than comparable, proctored examinations. Such a determination would be valuable in maintaining the academic integrity of a course or academic program and potentially satisfy governmental or accreditation authorities and demonstrate that quality efforts to monitor the integrity of online offerings were in place.

Review of the Literature

There has been an unmistakable growth of distance learning in post-secondary education in recent years. From the fall of 2002 to the fall of 2012, the proportion of students taking at least one online course at postsecondary institutions has increased from 9.6 percent of total enrollment to 33.5 percent of total enrollment. From the fall of 2003 to the fall of 2012 the proportion of surveyed chief academic officers that expressed the view that learning outcomes for online education was inferior to traditional, face-to-face education declined from 42.8 percent to 23 percent and the proportion of higher educational institutions that believed that online education was a key element of the long-term strategy for their respective institutions increased from 53.5 percent to 69.1 percent (Allen and Seaman, 2014). This provides additional evidence that the importance of online education most likely will continue to grow for the foreseeable future.

Despite the growth of distance learning, the obvious question of the integrity of assessments in distance learning courses, has received only modest attention. As Hollister and Berenson (2009) have written, "The most commonly reported challenge in online assessment is how to maintain academic integrity" (p. 272). This view has been echoed by Lanier (2006) and Harmon and Lambrinos (2008). There have been several studies (e.g., Charlesworth *et al.*, 2006; Grijalva *et al.*, 2006; Kennedy *et al.*, 2000; Lanier, 2006) which have reported on surveys taken of students regarding their self-reported cheating behavior in online versus face-to-face examinations or on their perceptions as to whether cheating is more prevalent in one circumstance versus the other. Unfortunately, no consensus emerges from these four student survey-based studies as to whether online assessments are more susceptible to cheating than traditional face-to-face assessments. Even if a self-report survey had not found evidence of greater cheating on online exams, various objections have been raised concerning the validity of student self-report surveys in examining student cheating. Findings by Miller, Shoptaugh and Parkerson (2008) suggest that the typical reliance on volunteer subjects in student cheating surveys creates a bias which leads to the under reporting of cheating behaviors. Such results reinforce the conclusions of Randall and Fernandes (1991) of an additional bias leading to under reporting of unethical behavior, such as student cheating. Moreover, broader criticisms of the usefulness and reliability of research based on self-report surveys have been raised by Baumeister *et al.* (2007) and Porter (2011). A separate survey of the perceptions of faculty members regarding their views as to whether online testing was more susceptible to student cheating than traditional, face to face examinations was undertaken by Rogers (2006). She

reports that although many of the faculty surveyed had concerns that online exams do facilitate student cheating, 81.8 percent of responding faculty continued to administer online exams or quizzes in an unproctored environment.

Efforts to develop statistically based methods to detect cheating go back at least to the late 1920's. Charles Bird (1927, 1929) suggested several statistical approaches to comparing the patterns of incorrect answers on objective exams provided by different students. If the incorrect answers offered by a pair of students were subject to a level of similarity that exceeded the limits indicated by statistical measures of chance, then an hypothesis positing student copying of answers would be supported. A significant number of studies (e.g., Saupe, 1960; Angoff, 1974; Frary, Tideman & Watts, 1977; Hanson, Harris & Brennen, 1987; Bellezza & Bellezza, 1989; Holland, 1996; Sotaridona & Meijer, 2002; van der Linden & Sotaridona, 2006; and, Wollack, 2006) have been published since then to develop, refine and advance Bird's (1927, 1929) seminal efforts. A good review of these statistical studies to detect cheating on objective exams is provided by Khalid, Mehmood & Rehman, 2011.

These approaches to the detection of cheating were developed primarily for the in-class proctored test environment. When students take an online test, there are potentially other sources of cheating related information than, say, another student in close physical proximity. Thus an online student with the intent to cheat, might avail him/herself by texting other students, by arranging for students or others with greater expertise to be present, or by searching internet test sites. Therefore, the response pattern of two cheating students might therefore be quite different, vitiating the validity of the earlier pairwise approach for detecting cheating.

The present study attempts to draw inferences on the relative incidence of cheating within these two testing environments based upon the statistical analysis of data gathered from students taking exams in each of the two environments. The present authors are aware of only four earlier studies that have also attempted to make such direct inferences as to whether unproctored, online testing is associated with a greater incidence of cheating--the analyses of Peng (2007), Harmon and Lambrinos (2008), Hollister and Berenson (2009), and Yates and Beaudrie (2009). In only the Harmon and Lambrinos (2008) study was the use of an unproctored, online exam approach found to have facilitated student cheating.

The suggestion of Lanier (2006), Harmon and Lambrinos (2008) and Hollister and Berenson (2009) that the issue of the integrity of online exams has not been sufficiently researched may be all the more puzzling in light of the growing evidence that academic dishonesty at the college level had become a substantial threat to academic integrity even before the question of whether a greater reliance on online exams potentially creates even greater peril. Representative research efforts by Crown and Spiller (1998) McCabe *et al.*, (2001) and Brown and McInerney (2008) document the strong and increasing prevalence of cheating behavior going back a number of decades.

One methodological approach to making inferences regarding whether there has been a higher incidence of cheating in an unproctored, online assessment environment versus a traditional, proctored environment is offered by Harmon and Lambrinos (2008). They examined the final exam scores of two groups of students taking a principles of macroeconomics course, one group in the summer 2004 and a second group in the summer of 2005. The 2004 class was given an online final exam consisting of thirty randomly selected multiple choice questions. The 2005 class was given a comparable multiple choice final in a traditional, proctored environment. An OLS regression model was utilized to explain the final exam performance of the twenty-four students taking the course in 2004 and the thirty-eight students taking the course in 2005. Although the original regression models contained variables measuring GPA (as a gauge of the student's overall ability), class year, age and whether the student subject was an economics major, the final specification for the two final exam regressions (2004 and 2005) utilized only the GPA explanatory variable. The R^2 statistics for the 2004 and 2005 groups were 0.08 percent and 49.72 percent, respectively. The Goldfeld-Quandt test was utilized in order to test for the equality of error variance between the two classes. A statistically significant difference was determined. This led Harmon and Lambrinos (2008) to infer that the two sets of regression results were different from one another. That, in turn, produced the inference that cheating in the unproctored online final for the 2004 class could be interpreted to be, in effect, an implicit, omitted variable

in the 2004 equation which accounts for the statistically significant difference in the pattern of error variances.

Hollister and Berenson (2009) test for the presence of greater cheating in the use of online exams by comparing the examination scores of two sections of students taking the same course in the same semester with the same instructor. One section took their exams in a proctored, face-to-face setting and the other section took the same exams online. Hollister and Berenson (2009) took steps to establish that the students in these two sections had comparable abilities and other key performance related characteristics. With such precautions in place, any statistically significant difference in exam performance would be attributed to cheating. As there was no statistically significant difference in the exam performance of the students in the two sections, Hollister and Berenson (2009) inferred that the use of online exams does not contribute to higher levels of student cheating.

Yates and Beaudrie (2009) tested for cheating by simply comparing the means of the proctored and unproctored groups with a simple t-test and failed to find a statistically significant difference between the two groups. Similarly, Peng (2007) compared the sixth quiz grade (taken online) to each of the five earlier quiz grades (based on different course material and taken in a proctored, in-class environment). Peng (2007) inferred that there is no evidence of cheating based on the lack of statistical significance in comparisons of the two situations.

One important distinction between (a) the Harmon and Lambrinos (2008) and Hollister and Berenson (2009) methods and (b) the methods utilized by Yates and Beaudrie (2009) and Peng (2007) is that the former studies do take explicit account of student grade point average (and potentially other student-specific variables which may be proxies for student ability) in explaining variations in exam performance. The latter studies do not make an explicit allowance for the possible relationship between cheating behavior and ability. The concern with the confounding effect of GPA on efforts to detect cheating in different cheating environments is based on a number of earlier studies (see Crown and Spiller (1998) for a review of such studies) that determined that less academically able students are more inclined to cheat in the performance of their coursework.

An Explicit Latent Variable Model

The view expressed by Harmon and Lambrinos (2008) is that when cheating is present, it is not directly observable and manifests itself as an omitted or latent variable affecting the error variances of the fitted regression. However, despite their observation that cheating may be viewed as a latent variable, Harmon and Lambrinos (2008) did not utilize latent variable methodology in their analysis. The present study explicitly explores the implications of cheating as a latent variable. As a result of this investigation, it is suggested that the use of the Goldfeld-Quandt test only examines part of the effect of cheating. Using an explicit latent variable model can be more revealing.

A classroom example using SEM

An introductory statistics course was offered to 52 undergraduate students at the suburban campus of a small, private university in the Northeast. While this sample of students was technically a "convenience" sample, there was no reason to believe that this sample was in any way not representative of the student body at the university. The final exam consisted of two parallel parts: a proctored in-class exam and an unproctored online exam. Both exams were assumed to be equivalent. The questions from both the proctored and unproctored exams were all taken from the same test bank provided by the publishers of the textbook used in the course. Both exams were based on the students' ability to solve numerical problems (i.e., there were no true-false, multiple choice, essay or other types of questions), were open-book and were subject to a two hour time limit. Great care was taken to assure that the topic coverage, the weighting of the various topics, type of question and the difficulty of the questions were judged to comparable for the online and in-class exams. The test forms could not be guaranteed to be parallel because the item characteristics were not available from the test publisher

Students were allowed two hours for the completion of each exam. Since both tests were taken by each student, there were, therefore, two test grades for each student. It should also be noted that the students took the proctored exam *after* they were given access to the unproctored online exam. The time window for the online exam was three days prior to the in-class proctored exam. It is possible that students may have benefited from a "testing effect," i.e., they

may have memorized or were otherwise more familiar with the type of questions on the proctored exam based on their prior experience with similar questions on the online exam. But this would only help them in the proctored exam and therefore potentially create a measurement bias against the detection of cheating on the online exam.

Additionally, mastery related information was gathered on each student. The two mastery related variables used were grade point average (GPA) upon entering the course and measured attendance in the introductory statistics class during the semester. It was hypothesized that both would have a positive effect on the final exam grades. The descriptive data for the class may be found in Table 1.

Table 1. Descriptive statistics

Variable	*Exam Score (Both Tests)*	*Exam Score (Proctored Test)*	*Exam Score (Unproctored Test)*	*GPA*	*Attend*
Mean	69.05	65.14	72.96	3.14	21.06
Standard Error	1.77	2.52	2.38	0.07	0.31
Median	67.87	67.08	70.68	3.18	21
Sample Variance	324.39	330.30	293.73	0.23	5.15
Minimum	16.67	16.67	39.12	2.04	9
Maximum	100.00	97.50	100.00	3.97	23
Sample Size	104	52	52	52	52

A path diagram for the example is shown in Figure 1. The use of SEM requires software availability and some training in the methodology. However, it offers a simple and often a graphical depiction of the processes, thus making the model structure readily comprehensible. Here the manifest variables of GPA, attendance, grades from the proctored test and grades from the unproctored test are depicted in rectangular boxes. The latent cheating variable is depicted in a circular box, as are the error components. The numbers associated with bidirectional arrows are covariances and the numbers associated with unidirectional arrows are the regression coefficients.

This model was estimated in SAS (Proc Tcalis) using a maximum likelihood discrepancy function and with the covariance matrix of the manifest variables as input. Note that the paths marked with an asterisk are statistically significant ($p<.05$). The path analysis approach offers a simple depiction of the cheating process. The measured mastery variables (GPA and attendance) both have direct paths to both the proctored and unproctored exam scores and are statistically significant and positive. Thus, not surprisingly, the mastery variables have a positive effect on test grades.

Perhaps more interesting is that both GPA and attendance have a negative effect on the unmeasured cheating variable, with the attendance effect statistically significant. The interpretation of this result is that students with high mastery variables are less likely to cheat than students with low mastery variables. Thus, in some sense, cheating can be seen as a substitute for mastery, affecting test scores.

Cheating is then modeled to have a direct effect on unproctored test grade scores, but not proctored test scores. Note that the correlation between the error terms is not statistically significant, in spite of the fact that the same students are in the proctored and unproctored groups. Thus the model is successfully removing any individual effect from the errors.

Harmon and Lambrinos (2008) argued that an omitted cheating variable would manifest itself as a difference in the variances of the errors. However, a Goldfeld-Quandt test on the error terms e_p and e_u in Figure 1 did not indicate a significant difference in the variances of the error terms ($p>.13$). So for this example, had the Harmon and Lambrinos (2008) approach been taken, it would have been concluded that cheating did not occur. The model presented in Figure 1 suggests that cheating did occur, but only as it relates to mastery variables.[2]

[2] The "regression" coefficient of 1.0 between cheating and the unproctored grades simply indicates that the latent cheating variable is scaled to directly and completely impact on the unproctored grades.

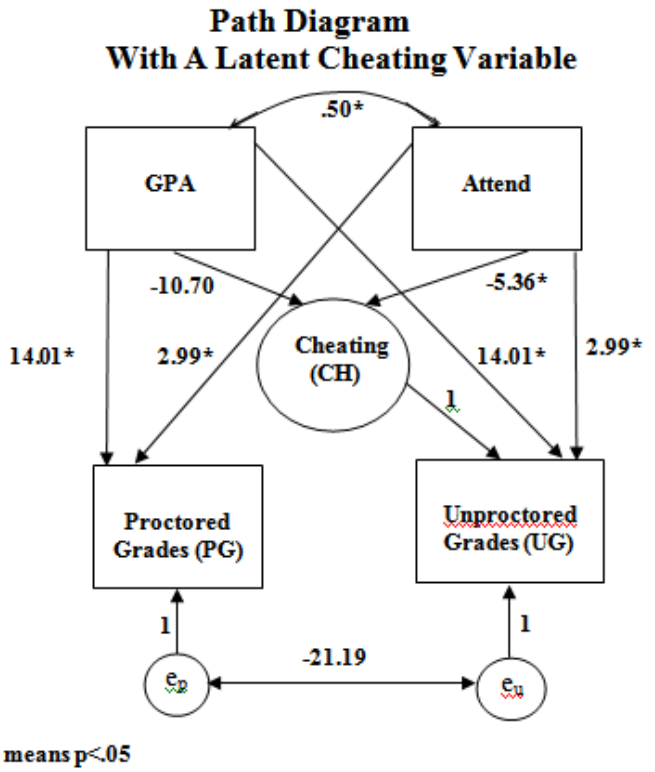

Figure 1. SEM path diagram results related to a latent cheating variable. The latent cheating variable is depicted as a circle, as are the error components. The numbers associated with bidirectional arrows are covariances and the numbers associated with unidirectional arrows are the regression coefficients.

The Same Example Using Stacked Regressions

The structural equation modeling approach, while easily interpretable, can be improved upon in this case. Greater utility can be gained by utilizing a stacked regression approach. Specifically, when the classical regression assumptions apply, the regression approach, unlike the SEM approach:

1) Can be implemented in any standard regression package, even MS Excel.
2) Easily allows power analysis and joint confidence intervals to be constructed.
3) Yields small sample results, rather than asymptotic results. These regression results are also BLUE (Best Linear Unbiased Estimates).

This approach is now described in simple scalar terms. Suppose that the basic model explaining student performance in a *proctored* environment is

$$Y_p = \beta_0 + \beta_1 X + \varepsilon_p, \quad (1a)$$

while the model for explaining student performance in an unproctored environment is

$$Y_u = \beta_0 + \beta_1 X + Z + \varepsilon_u \quad (1b)$$

Here Y_p and Y_u are test grades for proctored and unproctored exams respectively, X is a student mastery variable, the b's are model's coefficients and ε_p and ε_u the errors. Z is the unobserved cheating variable. In the present study, individual student grade point average (GPA) and student class attendance are designated as mastery related variables.[3] However, for the sake of clarity of presentation, the model in (1) contains just one student mastery variable, but it is easily extended to a multiple regression with several mastery variables. Assume the error terms of (1a) and (1b) ε_p , $\varepsilon_u \sim N(0, \sigma^2)$ and are independent of X. Finally, decompose the unobserved variable into two orthogonal components. That is,

$$Z = \delta_0 + \delta_1 X + v \quad (2a)$$

where the δ's are the coefficients relating the mastery variables to cheating and $v \sim N(0, \sigma_v^2)$ is the error, independent of X and of ε_u. Substituting (2a) into (1b) yields

$$Y_u = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)X + (v + \varepsilon_u). \quad (2b)$$

This formulation separates the effect of cheating into two orthogonal components, one which affects test grades through X and the other which acts upon test grades with the effect of X held constant (that is, v). Observe that in the application of this model in the present study, the proctored and unproctored groups consist of the same individuals who have been given equivalent final exams. However, the current methodology can easily be modified for the case where the proctored and unproctored exams are given to different comparable groups, provided the exams themselves are comparable.

It is may be shown that by ignoring the fact that Z is unobserved and just running the regression on the

[3] These 'mastery' variables were determined through the use of a step-wise regression procedure which initially considered a larger assortment of variables (student procrastination on homework assignments, performance on a quantitative pre-test given on the first day of the semester, SAT scores, etc.) which were hypothesized to be related to exam performance and correlated with students' inclination to cheat on exams.

data for the observed, unproctored observations (which do not include Z), would cause the estimates b_0 and b_1 to be biased with bias equal to δ_0 and δ_1, respectively (Maddala 1977, p. 156). It is seen by a similar argument that b_0 and b_1 are not just biased, but also inconsistent; that is, the bias persists even for large samples. The importance of that bias cannot be overemphasized. If the estimates b_0 and b_1 are substantially biased, it means that, for all intents and purposes, they do not measure what they purport to measure. The estimates are therefore useless.

The bias can be avoided by estimating Equations (1) and (2) by stacking equations (1a) and (1b) (and using appropriate dummy coding as explained by Draper and Smith (1981, p. 248)) which then yields

$$Y = \beta_0 + \beta_1 X + \delta_0 D + \delta_1 XD + \varepsilon_w. \quad (3)$$

For the proctored class, D is a dummy variable equal to 0 (and $\varepsilon_w = \varepsilon_p$) and for the unproctorerd class D equals to 1 (and $\varepsilon_w = v + \varepsilon_u$). Thus $Y_p = \beta_0 + \beta_1 X + \varepsilon_p$ for the proctored class. For the unproctored class, $Y_u = \beta_0 + \delta_0 + (\beta_1+\delta_1)X + (v + \varepsilon_u)$.

This procedure not only correctly estimates the impact (β_1) of the student mastery related variable (X) on the grades, but, interestingly, also the impact (δ_1) of the component of the cheating variable as it manifests itself through X onto the grades. Said another way, if β_1 is positive and significant, it suggests that a higher mastery variable tends to yield a higher final exam grade. Additionally, and importantly, if δ_1 is negative and significant, it suggests that someone with a higher mastery is less likely to cheat. Thus the stacked regression procedure estimates the impact of the mastery variables on both grades and on cheating. The modification of the methodology when the classical regression assumptions are not met is easily implemented.

Using standard regression techniques (Draper and Smith 1981, p. 248), a number of hypotheses can be tested:

a) In Equation (3), testing H_0: $\beta_1 = 0$ versus the alternate hypothesis, H_1: $\beta_1 \neq 0$ with a simple t-test, corresponds to a test of the statistical significance of the effect of the student mastery related variable on grades.

b) A simple t-test may also be used to test the null hypothesis, H_0: $\delta_1 = 0$, versus the alternate hypothesis, H_1: $\delta_1 \neq 0$.

That is, this test examines whether the student mastery related component of the latent cheating variable has a statistically significant impact on test grades. For (a) and (b), the effect of multiple mastery variables may be tested with partial F tests.

c) The Goldfeld-Quandt test can be used to test the hypotheses H_0: $\sigma^2_v + \sigma^2 = \sigma^2$ vs. H_1: $\sigma^2_v + \sigma^2 \neq \sigma^2$ or, equivalently, H_0: $\sigma^2_v = 0$ vs. H_1: $\sigma^2_v \neq 0$; that is, whether there is a statistically significant, non-mastery related component of the latent cheating variable which affects test grades.

Summarizing, the explicit latent variable model can test 1) whether and which mastery related variables affect test grades, 2) whether and which mastery related variables influence cheating and 3) whether there is a non-mastery related component to cheating. Thus, it may be observed that that the Goldfeld-Quandt test suggested by Harmon and Lambrinos (2008) only considers the non-mastery related component and ignores the other cheating components. Similarly, the t-test proposed by Yates and Beaudrie (2009) and the methodological approach adopted by Peng (2007) does not explicitly consider the mastery related components of cheating. It is quite plausible and even expected that the tendency to cheat will be enhanced by poor student mastery related characteristics. For example, Crown and Spiller (1998) reviewed fifteen studies which analyzed the relationship between past student academic performance and the inclination to cheat and found an inverse relationship in thirteen of those studies and no significant relationship in the remaining two studies. The current approach can detect such tendencies while the previously published approaches cannot. The results of the stacked regression, including confidence limits on the parameters are seen in Table 2. Here b_0, b_1, b_2, d_0, d_1, and d_2 estimate β_0, β_1, β_2, δ_0, δ_1, and δ_2 respectively. Note that these results are identical to those produced by the SEM model above. Additionally, a partial F test was performed to see if the set of student mastery related components of cheating is statistically significant. This set of mastery related components of cheating proved to be highly significant ($p < .00002$). That is, as a group, the mastery variables are related to cheating.

While, as stated above, there are some advantages to using stacked regression to estimate simple SEM models, in practice, either approach is quite acceptable. It is worth noting that the stacked regression approach is easily extended to simple SEM models with multiple latent variables.

In summary, the mastery related variables, GPA and attendance, are both positive and statistically significant, as might be expected. The student mastery related effects of the "cheating" variable are negative for both GPA (DGPA) and attendance (DAttend) and are, as a group, statistically significant. Thus the mastery related variables do affect the students' test grades. The mastery related "cheating" variables are negatively related to grades, and in the case of attendance, this negative relationship is statistically significant. In other words, attendance in class and, to a lesser degree, GPA become less important predictors of grades when students can more easily cheat on an online exam than they are when students' cheating opportunities are limited by a proctored exam

Table 2. OLS regression results explaining student exam scores (dependent variable)

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept (b0)	-41.89	21.15	-1.98	0.05	-83.86	0.08
GPA (b1)	14.01	5.09	2.75	0.01	3.90	24.12
Attend (b2)	2.99	1.08	2.77	0.01	0.85	5.13
Dmod (d0)	154.35	29.91	5.16	1.3E-06	94.99	213.71
DGPA (d1)	-10.70	7.20	-1.49	0.14	-25.00	3.59
DAttend (d2)	-5.36	1.53	-3.51	6.8E-04	-8.39	-2.33
Model:	R^2= 0.29	Adj. R^2= 0.25		Model F= 7.88	P-value= 2.8E-06	

Since the Goldfeld-Quandt test failed, this indicates that there is no statistically significant non-mastery components of cheating. Regarding the Partial F test (item (b) above), following Jamshidian *et al.*, (2007), the 95% upper and lower simultaneous confidence bands associated with the partial F-test are presented below in Figure 2. These are joint confidence bands for the effects of GPA and attendance on cheating (through their δ's). Note that for most levels of attendance and GPA the bands are above the zero plane. However, for high levels of attendance and/or GPA, the zero plane is included between the confidence bands. Applying the cheating interpretation to the latent variable suggests that except for the most serious students, cheating tends to substitute for attendance and GPA as a determinant of grades.

Finally, the stacked regression approach also allows a power analysis to be easily performed. The definition of the effect size follows Cohen (1988). Here the sample size is 104, the alpha level is .05, and the base R^2 is .06263. Figure 3 shows the ability of the partial F-test to detect the observed difference in R^2 of .22419 which results from the inclusion of the online variables is nearly 100%. Thus even with this relatively small sample size, the partial F test is sufficiently powered. The second graph below, Figure 4, shows the effect of sample size on power for the effect size exhibited in this example (.31435).

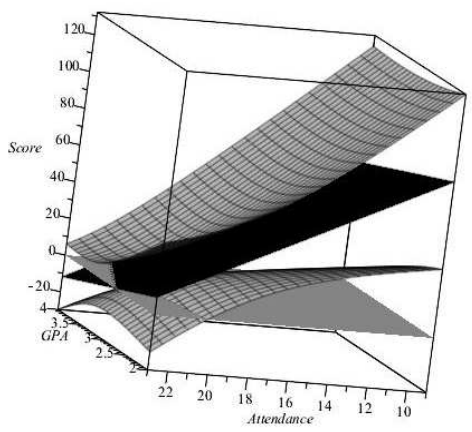

Figure 2. Expected test scores with lower and upper simultaneous confidence limits. The fitted regression plane is represented in black. The upper and lower confidence bands are rectangularly hashed. The zero plane is represented in solid gray. Attendance is measured against a maximum score of twenty-three class meetings and GPA is measured against a maximum value of 4.0.

environment. It may be inferred that cheating can substitute for mastery as a determinant of grades. Had only the Harmon and Lambrinos (2008) analysis been performed on the current sample, it would have concluded that cheating was not present.

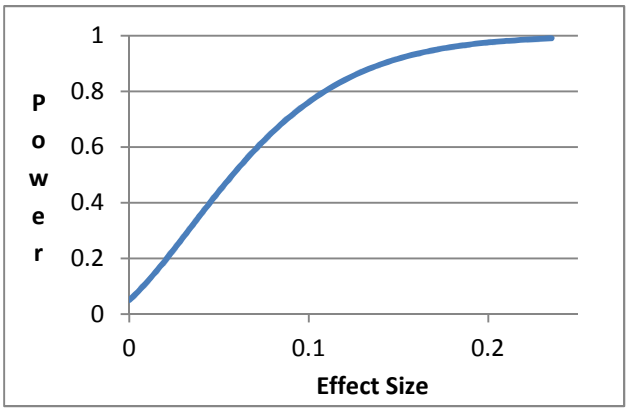

Figure 3. Power by Effect Size

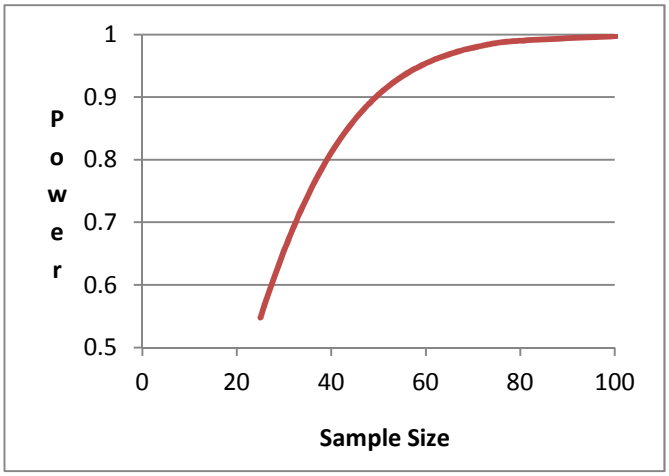

Figure 4. Power by Sample Size

Limitations of This Study

Sample size

Although the sample size of fifty-two students is relatively small, a smaller sample size generally decreases the investigator's ability to find statistical significance. However, the sample size was adequate to yield statistical significance regarding the central issue addressed in this study--the relationship between online testing and student cheating. Thus the approach that is offered in this study to examine whether online exams facilitate a greater presence of cheating among students should be useful to detect such cheating in either small or large classes.

This study was performed in a small class environment. However, it could be argued that these results may not generalize well to a large class environment, The reasons for this may be myriad. For example, in a large class, for the in-class test, it may be more difficult for the instructors to monitor cheating. This would affect the baseline and make it more difficult to detect cheating in the online exam, even though the sample size would be larger. Thus it would be wise to investigate the effect of class size on the effectiveness of the current approach.

The interpretation of the latent variable

Regarding the interpretation of latent variables, it is important to understand that whenever latent variables are estimated, whether here or in many other contexts, meaning can rarely be ascribed unambiguously to the latent variable. Thus, this study has been designed in such a way that the latent variable is intended to correspond to cheating.

While it appears that cheating is the most reasonable interpretation of the latent variable, another interpretation should be considered. Specifically, the proctored exam was given in a traditional classroom setting, while the unproctored exam was administered online, with the students in their dorm or home environment. It may be hypothesized that students are more at ease in their dorm or home environments than they are in a more structured classroom environment in taking an exam. This hypothesized greater comfort level in the dormitory or home setting could reduce the anxiety associated with examinations or otherwise lead to a higher level of performance for students taking the online exam. In such a case, a measured higher level of performance in an unproctored environment could be attributed, at least partially, to the physical environment rather than to a hypothesized increased level of student cheating on an unproctored exam.

Alternatively, as pointed out by Hollister and Berenson (2009), the dorm or home environment could be subject to greater distractions, difficulties with computer or network connections and problems that students might have in interpreting test questions. In such cases, the home or dormitory environment could lead to poorer exam scores. It is therefore possible

that even if cheating did not occur, the latent variable reflects this difference in student comfort in a less structured environment rather than cheating. That is, the environment and cheating are confounded in this design. However, an empirical argument against this interpretation is that if the environment were the issue, it would be expected to manifest itself largely in the non-mastery related component. In the current analysis the non-mastery related cheating component was not statistically significant ($p = .130$) while the mastery related cheating component ($p < .00002$) was highly significant. This pattern of results suggests that if the environment is a factor, it most likely has far less of an effect in this sample than the mastery related cheating component.

Directions for Future Research

This paper presents an empirical approach to the detection of cheating. Using the tools developed here, some interesting avenues of future research are possible. In particular, the notion of creating empirically verifiable strategies to impede efforts to cheat, would seem possible. The exact nature of those strategies will depend in part on the software tools which currently exist or will be developed for online test delivery. Non-technical interventions should also be considered, e.g., a classroom honor code could be tested with the methodology developed here.

Another issue for future research is a refinement of the methodology. In particular, for this study, the online test was administered first, thereby allowing for a "teaching effect" to possibly enhance the in-class test scores. An alternative would be to use a counterbalanced design (Hersen & Barlow, 1976) with one class getting the in-class test first and the other class getting the online test first. The advantage of this approach would be to provide a clearer picture of the cheating effect, which may have been understated in the current study. However, when students become aware of the counterbalancing, depending on the student body, a negative emotional reaction may occur, since students are likely to believe that the other class has been given an advantage. This reaction could possibly damage the generalizability of the results. Thus the effect of counterbalancing would need to be explored as well.

Summary and Conclusions

The present article utilizes a latent variable model to measure whether there is a greater level of cheating in an unproctored, online environment relative to the cheating level in a proctored, in-class environment. It was shown that this method has the capacity of separating out (1) the influence of student mastery related variables on exam performance, (2) a determination of which mastery related variables are statistically linked to cheating behavior and (3) a determination of whether there are non-mastery related components that are statistically linked to cheating. Applying this method (either with an SEM approach or a stacked regression) to the data collected from a sample of fifty-two students who took introductory statistics at a private university in the Northeast revealed, first, that an unproctored, online testing environment can facilitate a relatively higher level of cheating and, second, that the mastery related variables were statistically linked to cheating — a pattern of results consistent with the interpretation that some students may view class attendance and cheating as alternate strategies to pass the course.

Of course, the approaches to detect cheating presented in this analysis can be part of a college or university's efforts to maintain the quality and academic integrity (and, therefore, reputation) of its overall distance education activities. However, this approach can be downscaled, i.e., implemented for a particular college within the university, a particular program within a college, a particular course or even for a particular instructor if investigative efforts or other research indicate a greater likelihood of cheating. For example, Crown and Spiller (1998), Whitley (1998) and Day, Hudson, Dobies, & Waris (2011) suggest that there are different inclinations to cheat among various categories of students in different disciplines, at different levels (e.g., introductory courses, upper-level undergraduate courses, graduate or professional courses) and different modes of interaction between students and faculty. The methodologies presented in this paper may be applied to investigate cheating under a wide variety of circumstances.

References

Allen, I., & Seaman, J. (2014). *Grade Change: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group, Babson College.

Angoff, W. (1974). The development of Statistical indices for detecting cheaters. *Journal of American Statistical Association, 69*, 44-49.

Baumeister, R., Vohs, K., & Funder, D. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403.

Bellezza, F., & Bellezza, S. (1989). Detection of cheating on multiple-choice tests by using error- similarity analysis. *Teaching of Psychology, 16*, 151-155.

Bird, C. (1927). The detection of cheating in objective examinations. *School and Society, 25*, 261-262

Bird, C. (1929). An improved method of detection cheating in objective examinations. *Journal of Educational Research, 19*, 341-348.

Brown, B., & McInerney, M. (2008). Changes in academic dishonesty among business students in the United States, 1999-2006. *International Journal of Management*, 5, 621-632.

Charlesworth, P., Charlesworth, D., & Vician, C. (2006). Student's perspectives of the influence of web-enhanced coursework on incidences of Cheating. *Journal of Chemical Education, 83*, 1368-1375.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crown, D., & Spiller, M. (1998). Learning from the literature on collegiate cheating: a review of empirical research. *Journal of Business Ethics*, *17*, 683-700.

Day, N., Hudson, D., Dobies, P., & Waris, R. (2011). Student or situation? Personality and classroom context as predictors of attitudes about business school cheating. *Social Psychology of Education, 14*(2), 261-282.

Draper, N., & Smith, H. (1981). *Applied regression analysis*. New York: Wiley.

Frary, R., Tideman, T., & Watts, T. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics, 6*, 152-165.

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

Hanson, B., Harris, D., & Brennan, R. (1987). A comparison of several statistical methods for examining allegations of copying (ACT Research Report Series No. 87-15). Iowa City, IA: American College Testing

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

Hersen, M., & Barlow, D. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press.

Higher Education Opportunity Act, Public Law No. 110-315, 122 Stat. §§ 3078-3508 (2008).

Holland, P. (1996). Assessing unusual agreement between the incorrect answers of two examinees using the K-index: Statistical theory and empirical support (Research Report RR-94-4). Princeton, NJ: Educational Testing Service.

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

Jamshidian, M., Jennrich, R., & Liu,W. (2007). A study of partial F tests for multiple linear regression models. *Computational Statistics and Data Analysis*, *51*, 6269-6284.

Kennedy, K., Nowak, S., Raghuraman, R., Thomas, J. & Davis, S. (2000). Academic dishonesty and distance learning: student and faculty views. *College Student Journal*, *34*, 309-314.

Khalid, M., Mehmood, Z., & Rehman N. (2011). Statistical methods for answer copying: A brief overview. *British Journal of Arts and Social Sciences, 1,* 49-61.

Lanier, M. (2006). Academic integrity and distance learning. Journal *of Criminal Justice Education*, *17*, 244-261.

Maddala, G. (1977). *Econometrics*. New York: McGraw-Hill.

McCabe, D., Treviño, L., & Butterfield, K. (2001). Cheating in academic institutions: a decade of research. *Ethics & Behavior*, *11*, 219-232.

Miller, A., Shoptaugh, C., & Parkerson, A. (2008). Under reporting of cheating in research using volunteer college students. *College Student Journal, 42*(2A), 326-339.

Peng, Z. (2007). Giving online quizzes in corporate finance and investments for a better use of seat time. *The Journal of Educators Online*, *4*, 1-18.

Porter, S. (2011). Do college student surveys have any validity? *The Review of Higher Education,35,* 45–76.

Randall, D., & Fernandes, M. (1991). The social desirability response bias in Ethics Research. *Journal of Business Ethics, 10*, 805-817.

Rogers, C. (2006). Faculty perceptions about e-cheating during online testing. *Journal of Computing Sciences in Colleges*, *22*, 206-213.

Saupe, J. (1960). An empirical model for the corroboration of suspected cheating on multiple- choice tests. *Educational and Psychological Measurement, 20*, 475-489.

Sotaridona, L., & Meijer, R. (2002). Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement*, *39*, 115-132.

van der Linden, W., & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics, 31*, 283-304.

Whitley, B. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, *39*(3), 235-274.

Wollack, J. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. *Applied Measurement in Education*, 19, 265-288.

Yates, R., & Beaudrie, B. (2009). The impact of online assessment on grades in community college distance education mathematics courses. *The American Journal of Distance Education*, *23*, 62-70.

.

Citation:

Fask, Alan, Englander, Fred, Wang, Zhaobo (2015). On the Integrity of Online Testing for Introductory Statistics Courses: A Latent Variable Approach. *Practical Assessment, Research & Evaluation*, 20(10). Available online: http://pareonline.net/getvn.asp?v=20&n=10

Corresponding Author:

Fred Englander
Department of Economics and Finance
Fairleigh Dickinson University
285 Madison Avenue
Madison, NJ 07940

email: englandr [at] fdu.edu