The Role of Certainty and Time Delay in Students' Cheating Decisions during Online Testing

Chia-Yuan Chuang (cchuang6@asu.edu)
Arizona State University, Simulation Modeling and Applied Cognitive Science Program
7171 E Sonoran Arroyo Mall
Mesa, AZ 85212, U.S.A

Scotty D. Craig (Scotty.Craig@asu.edu)
Arizona State University, Human Systems Engineering
7171 E Sonoran Arroyo Mall
Mesa, AZ 85212, U.S.A

John Femiani (John.Femiani@asu.edu)
Arizona State University, School of Computing, Informatics, and Decision System Engineering
7171 E Sonoran Arroyo Mall, Peralta Hall, Rm 230U
Mesa, AZ 85212, U.S.A

Abstract

In an attempt to assist proctors to prevent test takers from academic dishonesty in remotely administrated exams, this study investigated the ability of test takers' behaviors during online assessments to predict their cheating decisions. Specifically, this experimental study focused on the role of students' time delay and certainty rating during lab based online testing sessions. The analysis of hierarchical logistic regression indicated that not only time delay but also certainty rating had significantly statistical relation to test takers' cheating decisions. The importance of the two proposed factors during online assessments was discussed and the prospects of the improvements of online proctoring systems were addressed.

Keywords: Cheating; online assessment; online testing; uncertainty

Introduction

Academic institutions are turning to online education in order to expand their reach and provide education to a greater volume and more diverse group of students, while at the same time, using less faculty labor and less physical infrastructure than traditional face-to-face courses. However, the distributed nature of online courses presents a potential risk of increased academic dishonesty, particularly when students are asked to take exams at remote locations without a proctor in the room (Harmon & Lambrinos, 2008; Kennedy, Nowak, Raghuraman, Thomas, & Davis, 2000; Prince, Fulton, & Garsombke, 2011; Watson & Sottile, 2010). In order to ensure the integrity of student work, in 2008, Congress authorized the Higher Education Opportunity Act with the provision that it is the requirement for an institution that offers distance education to verify the identity of online students (Frank, 2010).

There are at least three important reasons for addressing issues of academic dishonesty in distance education. The first is an increasing trend in online education, both in terms of student enrollment (National Science Board, 2012) and corporate market (Adkins, 2008, 2011). Second, surveys of both faculty and students indicate a belief that cheating is more prevalent in online exams when students are not proctored (Kennedy et al., 2000; Watson & Sottile, 2010). Third, empirical studies have demonstrated that given the same online learning materials, scores in un-proctored exams were not only significantly higher than proctored ones (Prince et al., 2011) but also had significantly lower degrees of explanatory power to students' ability (Harmon & Lambrinos, 2008). Therefore, although online education provides opportunities to people who traditionally would not have access to high quality education due to schedule conflicts or physical constraints, these opportunities may be undercut if prospective employers do not trust the diplomas and certificates gained through online courses.

The prevention of academic dishonesty can be addressed to some extent by altering the assessments. Examples of this would include using multiple versions of an exam, randomizing question order, or not using identical exam questions from previous semesters (Harmon, Lambrinos, & Buffolino, 2010). However, a need remains to replace the traditional proctor in the room by another system to ensure the qualification of the online degrees offered by institutions (Frank, 2010; Harmon et al., 2010). A survey of techniques and tools for proctoring remotely administered exams (Frank, 2010) found that the majority of solutions involve recording an exam attempt or streaming a live video to a proctor who will monitor or review the exam sessions from a remote location. However, a naive approach of reviewing a set of recordings of individual exams may take significantly more effort than it would take a single proctor to monitor students in a traditional classroom setting.

The objective of this paper is to explore significant factors which may not only improve the effectiveness of remotely administrated exams but also scale the use of online proctoring. The primary contribution of this paper is to test

the ability of test taking behavior to predict student cheating during online exams. Specifically this paper tests the impact of a student delay time to answer a question and the student certainty rating for the question.

Literature Review

Previous Research in Academic Dishonesty

Crown and Spiller (1998) reviewed a wide range of research on collegiate cheating and categorized factors into two types, individual factors and situational factors. Individual factors represent the sum total of the life experiences and circumstances, including personal attributes, type of education and personality variables. Situational factors represent the situational pressures which come to bear on the individual to encourage or discourage cheating decisions, including honor codes, sanctions, values counseling, and surveillance (Crown & Spiller, 1998).

Based on the research from Crown and Spiller (1998), some researchers formulated structural equation models showing interaction among all possible variables in individual factors and situational factors (Murdock, Hale, & Weber, 2001; Sierra & Hyman, 2006; Smith, Davy, Rosenberg, Haight, & G, 2003). Murdock et al. (2001) categorized individual factors into six categories including grade in school and five academic motivations (academic self-efficacy, personal task goals, personal extrinsic goals, classroom task goals, and classroom extrinsic goals). Situational factors were classified into four categories based on social motivations (participation structure, teacher commitment/ competence, and level of school teacher respect and school belonging). The significant factors included grade in school, academic self-efficacy, extrinsic goal orientation, participation structure, teacher commitment and teacher respect. Smith et al. (2003) grouped individual factors into demographic and attitudinal variables, and organized situational factors into in-class deterrents. Their results indicated that the primary influences on future cheating were in-class deterrents, prior cheating, and the degree of neutralization. Sierra and Hyman (2006) proposed a new model of cheating intentions based on individual factors including cognitive constructs and anticipated emotion constructs. Although prior research (Murdock et al., 2001; Smith et al., 2003) indicated significant relationships between individual factors and situational factors related to cheating intentions, neither the effect of anticipated positive emotions (e.g. elation) nor the simultaneous effect of cognitive factors (e.g., locus of control and personal expertise) in individual factors was considered. Sierra and Hyman (2006) conducted an empirical experiment and revealed that anticipated emotional and personal expertise have significantly positive effect to drive uncertain choice in cheating contexts, while internal locus of control has a significantly negative effect.

Response Times in Testing

One of the factors proposed in this paper is time delay in online testing, which may indicate suspicious behaviors in online exams, such as, cheating. The traditional approach to detecting aberrant behavior is to use person-fit analysis by which aberrant response patterns that defy some expectation can be an index to validate the integrity of test scores (Meijer & Sijtsma, 2001). Although the methods of person-fit analysis have been well developed for the last two decades, the reasons for aberrant response patterns are still largely unknown. However, cheating is one of the reasons for aberrant response patterns (Meijer & Sijtsma, 2001; Petridou & Williams, 2007).

Fortunately, in 2008, Van Der Linden and Guo used response times (RTs) as an additional source of information on the test taker's behavior. They conducted a simulation study on two cheating behaviors: (1) pre-knowledge of some of the items; (2) attempts to take tests only for the purpose of memorizing the items. Under 800 replications of the pre-knowledge simulation, given one item which a test taker had pre-knowledge of and answered within 10 seconds, the satisfactory power of detecting the cheated item was 0.83 with $\alpha = 0.05$. Under 800 replications of memorization simulation, given 5 items which a test taker tried to memorize within a 30 minutes exam, the satisfactory power of detecting that one, two, three, four and five items that test taker memorized were 0.26, 0.35, 0.20, 0.10, and 0.04 respectively, with $\alpha = 0.05$.

Impasses in Learning

The other factor we focus on is students' certainty rating on the scale from one to five, where one indicates a guess and five indicates knowledge with high confidence. The certainty rating as a factor is inspired by the theory of impasses during learning. Impasses are obstructions students encounter in academic settings. They occur when a student gets stuck, detects an error, or does an action correctly but expresses uncertainty about it (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). A cognitive system is in disequilibrium when individuals are confronted with problems or situations that present obstacles to goals, anomalous events, contradictions, discrepancies, and obvious gaps in their knowledge (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). Confusion is exhibited when students hit an impasse and learning is the key process to transition a cognitive system from disequilibrium to equilibrium (Craig, Graesser, Sullins, & Gholson, 2004; Craig, 2012; D'Mello, Lehman, Pekrun, & Graesser, 2014; Lehman, D'Mello, & Graesser, 2012; Lehman, D'Mello, & Strain, 2011).

Current study

Rather than focusing on personal or situational factors, this paper examined test taking behaviors. Specifically it investigated if students' delay time to answer a question and their certainty rating for the question impacted their decision to cheat. This study implemented a common metric to

define cheating behaviors used by a majority of proctoring systems: a misuse of forbidden resources, such as a smart phones or cheat sheets (Frank, 2010).

Based on Van der Linden and Guo (2008) model that used response times as a potentially significant factor for cheating, it is hypothesized that the time a test taker spends on a single question plays a significant role in predicting students' decision to consult a forbidden resource. It is expected that test takers spend a greater amount of time to search for the answer than as opposed to answering the question honestly.

Additionally, the student's confidence in their ability to answer the question correctly could impact their cheating behavior. When a student encounters a question that they cannot answer or have difficulty answering, their level of certainty will decrease. This inability to answer the question is the equivalent to an impasse (VanLehn et al., 2003) in a learning setting. During testing, there are no learning opportunities remaining to work pass the impasse. If the student is sufficiently motivated to provide a correct answer and resources are readily available with limited monitoring within the online setting, as uncertainty level increase, students are more likely to cheat because it is the only way to resolve the impasse and move forward.

This paper seeks to answer the following two research questions: (1) Can time delay be an indication reliably predict cheating decisions during online exams? (2) Can student's certainty rating be an indication to reliably predict cheating decisions during online exams? In order to answer the two research questions, there are two null hypotheses for each question:

$H1_0$ A test taker's time delay on each question has no statistically significant relation to cheating decisions during online exams.

$H2_0$ A test taker's certainty rating on each question has no statistically significant relation to cheating decisions during online exams

Methods

Participants

Forty-two students (28 male, 14 female) took part in the study. They were between the ages of 18 and 36 (M = 20.93, SD = 3.90). Participants were undergraduate students enrolled in an Introductory Psychology course. They were offered partial course credit in return for their participation in the study.

Materials

Learning materials. The learning materials were a 12 minute video covering the basics of the Python computer programming language. Two pages of printed summary along with the video lecture were provided to participants during the learning phase. In addition, one piece of blank paper was provided to participants. Participants were allowed to take their own notes either on the two pages of summary or on the blank page while watching the video. Python program was selected as the domain because most participants would not have been exposed to it before the study. So, it was less likely they would already know the answers to the test.

The Python lecture started by the introduction of Python interpreter, such as entering and leaving Python interpreter through terminal, different types of variables in Python, declaration of variables and assigning new values to the declared variables. After that, the lecture went over some default operators and functions in Python, for example, modulo and comparison operators and a length function. Finally, participants were taught how to declare and execute their own functions in a Python file.

Testing materials. Two ten-item multiple choice tests were presented to participants. Both of these tests covered the material presented on the Python programming language, but each test had unique questions and covered different concepts. The first test session was implemented within a typical online exam setting. The second session was implemented in a cheating inducing environment in which participants were encouraged to answer questions by all means even if cheating. The second session was used to ensure that some cheating behaviors were observed.

Cheating materials. The notes and documents provided in the learning phase were returned to participants as cheating materials. The cheating materials included one page of self-written notes and two pages of summary of the video lecture. Participants were allowed to put the cheating materials at any place they liked.

Interview materials. At the end of each testing phases, participants were interviewed by the experimenter. This interview required participants to provide a self-reported certainty rating, prior knowledge, self-reported cheating, methods for cheating, preparation of cheating, and demographic survey for each question. In this paper, we only focused on answering certainty, prior knowledge, and self-reported cheating. The participant's certainty rating consisted of a scale from one to five, where one indicates a guess and five indicates knowledge with high confidence Prior knowledge was assessed by self-reports with a scale from zero to five, while zero indicates no knowledge on computer programming and five indicates mastering the topic before the experiment. For the cheating measure, the experimenter stepped one by one through each assessment item and asked if the participant cheated. If they cheated they received a follow up questions of how they cheated on the question. Finally, participants reported to the experiment whether they had a thorough plan to cheat or had an impulsive cheating if they cheated in online exams.

Procedure

The experiment was a repeated-treatment design. Figure 1 shows a diagram of the overall process that participant's undertook. After participants arrived at the lab and completed the informed consent procedure, they completed the study which consisted of four phases labeled phase A through D in Figure 1.

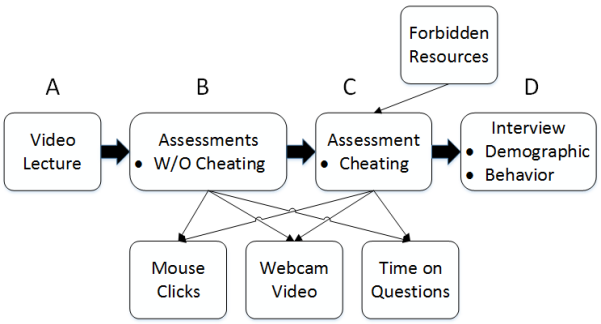

Figure 1: Design and Procedures of Online Exams

Phase *A* was a learning session. After the informed consent process was completed and initial instructions presented to participants, they started the learning phase of the study. In this phase, participants were asked to watch a 12 minute video lecture on the subject of programming in the Python language. One piece of blank paper and two pages of printed summary were provided to participants as notes and learning materials. After finishing the video lecture, the experimenter went through the learning documents to make sure that participants not only understood the content in the video lecture, but also were familiar with the content in the learning documents. Therefore, participants could find the answers in the learning materials while using them as cheating materials.

Phase *B* was a replication of a typical online testing setting. Participants took a 13 minute online exam in which forbidden resources such as smart phones and cheat sheets were not allowed. The computer system was also locked into the testing program, by which participants could not leave the exam window until they finished the exam. Only the experimenter knew the special keystroke combination to leave the testing program. So, no online resources could be accessed during the experiments. The time and positions of mouse clicks, webcam videos, and time on questions were recorded during this experiment. However, analysis of the dataset collected in the phase *B* was beyond the scope of the current study. The goal of this session was to make participants familiar with the testing environment.

Phase *C* was a cheating inducing environment. Participants again took a 13 minute exam and were asked to cheat without being caught by an online proctoring system. The forbidden materials were returned to participants. Participants had 5 minutes to arrange their cheating materials before starting the exams. The time and positions of mouse clicks, webcam videos, and time delay in each question were recorded by a proctoring system. In this study, only time delay in the phase *C* was analyzed.

In phase *D*, the experiment stepped one by one through each assessment item and asked participants' certainty rating of each question and if they cheated. If they cheated, they received a follow up questions of how they cheated on the question. After that, participants were given a demographic survey, including their major, previous experiences in computer programming, gender, age and the year in school. Finally, participants reported to the experiment that whether they had a preparation to cheat or not if they want to cheat in online exams.

Results

A hierarchical logistic regression was conducted using subjects, previous experience, time delay and certainty rating as predictor variables to predict student's cheating behavior (criterion variable). The initial model had two predictors, subject and previous experience. This indicated that these two variables were not significant predictors of cheating, $\chi^2(2) = 1.49$; $p > 0.05$. The first model had an $R^2 = 0.004$. A second model added time delay as a predictor. This addition resulted in a significant model, $\chi^2(1) = 48.91$; $p < 0.001$. The delta R^2 between the first model and the second model are 0.109 and the second model had an $R^2 = 0.113$. Finally, in the third model, certainty rating is added in addition to the previous predictor variables resulting in a significant change, $\chi^2(1) = 7.81$; $p < 0.01$. The delta R^2 between the second model and the third model were 0.016 and the third model had $R^2 = 0.129$. This indicates that student's certainty rating is a significant factor to predict cheating decisions. Time delay has the strongest predictive power with $\exp(B) = 1.00$; $t(1) = 29.01$; $p < 0.01$, which means an increment of one second will increase about one percent of odds ratio of cheating. Certainty was also significant predictor with negative relationship as certainty decreased likelihood of cheating increased, $\exp(B) = 0.757; t(1) = 7.83$; $p < 0.01$. The mean and S.D. of time delay and certainty rating for cheating and non-cheating items is provided in Table 1.

Table 1: Mean and S.D. for time delay and certainty rating

	Time (sec)		Certainty	
	Mean	SD	Mean	SD
Cheat	58.84	33.28	3.51	1.24
No Cheat	37.14	26.43	4.08	1.06

Discussion

Both null hypotheses, $H1_0$ and $H2_0$, were rejected and it was concluded that not only time delay but also certainty rating of each question were significant predictors of test takers' cheating decisions. Specifically, the probability of cheating was positively related to time delay but negatively related with the participant's certainty rating. The results also indicated that neither subjects themselves nor their experiences in the test materials significantly predicted cheating decisions. The strongly positive relationship between time delay and academic dishonesty matches the expectation that given the opportunity to cheat, cheaters spend more time in consulting forbidden resources than non-cheaters. The significant relationship between uncertainty rating and cheating behaviors is also compatible with impasse theory (VanLehn et al., 2003).

One of the strengths in this study was that the incidents of cheating were based on retrospective reports provided directly after testing instead of questionnaire surveys, which

were not based on actual behavior, but only responses to potential situations (Crown & Spiller, 1998; Murdock et al., 2001; Sierra & Hyman, 2006; Smith et al., 2003). A challenge to the use of surveys to determine cheating intentions is a possible self-report bias which may lead to underreporting (Scheers & Dayton, 1987) or over-reporting the probability of cheating (Nelson & Schaefer, 1986). It can be argued that surveys may depend upon students to admit their guilt which may be to their perceived disadvantage and cause them under-report dishonest behaviors. Conversely, it can be argued that subjects perceive their deviant behaviors which place themselves at odds with others in the classroom and therefore over-report cheating (Crown & Spiller, 1998).

The other strength in this study was that the proposed factors, such as time delay, can be monitored in real time. The factors explored in the previous research were personal/situational constructs, which ignore the dynamic behaviors of test takers during online testing. Moreover, given the personal/situational factors, it seems unlikely for schools to run a mass profiling survey. The proposed factor, time delay, provided an objectively quantitative measurement which can be easily implemented and coped with current online proctoring systems.

Since current proctoring system can record test takers behaviors during online exams, including facial expressions, it is possible that a test taker's certainty rating of each question can be assessed more objectively based on affective states, for example, confusion (Craig, D'Mello, Witherspoon, & Graesser, 2008). Craig et al. (2008) found that the physical exhibition of confusion has a significant relationship to observable human facial action units (Ekman & Friesen, 1978), especially for AU 4 and AU 7. The current findings provide the basis for future work on the automatic analysis of video data. The analysis of confusion and delay time through recorded videos also provides an opportunity for proctoring systems to monitor test takers' certainty rating in real time.

There are several open questions left in this research. The first one, classification accuracy, is currently unknown. Further research is still needed to determine the accuracy for classification but the significant features found in this paper are a start toward such as a model of detecting suspicious behaviors during remote testing. The application of this research could be significant time reduction in remote proctoring.

The second open question is the relationship among time delay, certainty, and cheating. Theoretically, time delay and certainty are both causes of cheating. However, low certainty could be a prerequisite for cheating and time delay could be a consequence of the act of searching for the answer in the materials.

The third open question is the validity of the research in the real world. It is possible that the research could not have the fidelity to transfer from the laboratory setting into a real world setting. Therefore, replication in the real world setting would be beneficial for understanding the generalizability of the finding.

The fourth open question is that there are other potential factors other than time delay and uncertainty useful for detecting cheating behaviors during online exams. It is not recommended that the classification criteria of cheating behaviors are just based on two proposed factors. Proctors should combine more evidence, such as checking the recorded videos and see if test takers actually access forbidden resources. The propose work only indicated that time delay and certainty rate are significant factors which may help proctors to improve the proctoring process in remotely administrated exams.

Finally, proctoring has been shown to not only deter cheating in online assessments but also enhance learning performance in online courses. Wellman (Wellman, 2005) showed that online-module delivery paired with proctored quizzes was more effective in promoting learning when compared to un-proctored quizzes. The proctored group practiced more frequently than the un-proctored group, especially students in the bottom half of performers. In spite of the benefits, it can be impractical to supervise all quizzes in large online courses. Typically only high-stakes exams, such as midterms or final exams, are under surveillance (Luecht, 2006). The standard methods of proctoring and human surveillance are extremely resource intensive. This current work provides the first steps toward potential methods to automatically detect cheating during online assessments.

References

Adkins, S. S. (2008). *US Corporate eLearning Market Reached $5.2 Billion in 2007*.

Adkins, S. S. (2011). *The US Corporate Market for Self-Paced eLearning Products and Services: 2010-2015 Forecast*.

Craig, S. D. (2012). Confusion's Impact on Learning. *Encyclopedia of the Sciences of Learning*, 766–767.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). *Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive–affective states during learning* (pp. 777–788). Cognition and Emotion.

Craig, S. D., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241–250.

Crown, D., & Spiller, M. (1998). Learning from the literature on collegiate cheating: A review of empirical research. *Journal of Business Ethics*, *17*(6), 683–700.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.

Frank, A. (2010). *Dependable distributed testing: Can the online proctor be reliably computerized? Proceedings of the 2010 International Conference on IEEE* (pp. 1–10).

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition*, *33*(7), 1235–1247.

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

Harmon, O. R., Lambrinos, J., & Buffolino, J. (2010). Assessment Design and Cheating Risk in Online Instruction. *Online Journal of Distance Learning Administration*, *13*(3).

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

Lehman, B., D'Mello, S., & Graesser, A. (2012). Confusion and complex learning during interactions with computer learning environments. *Internet and Higher Education*, *15*(3), 184–194. doi:10.1016/j.iheduc.2012.01.002

Lehman, B., D'Mello, S., & Strain, A. (2011). Inducing and tracking confusion with contradictions during critical thinking and scientific reasoning. In *Artificial Intelligence in Education* (pp. 171–178).

Luecht, R. M. (2006). *Operational issues in computer-based testing*. *Computer-based testing and the Internet: Issues and advances* (pp. 91–114).

Meijer, R. R., & Sijtsma, K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*, *25*, 107–135. doi:10.1177/01466210122031957

Murdock, T., Hale, N., & Weber, M. (2001). Predictors of cheating among early adolescents: Academic and social motivations. *Contemporary Educational Psychology*, 96–115.

National Science Board. (2012). *Science And Engineering Indicators*.

Nelson, T., & Schaefer, N. (1986). Cheating among college students estimated with the randomized-response technique. *College Student Journal*, 321–325.

Petridou, A., & Williams, J. (2007). Accounting for aberrant test response patterns using multilevel models. *Journal of Educational Measurement*, *44*(3), 227–247. doi:10.1111/j.1745-3984.2007.00036.x

Prince, D., Fulton, R., & Garsombke, T. (2011). Comparisons Of Proctored Versus Non-Proctored Testing Strategies In Graduate Distance Education Curriculum. *Journal of College Teaching & Learning*, 51–62.

Scheers, N., & Dayton, C. M. (1987). Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education*, 61–69.

Sierra, J. J., & Hyman, M. R. (2006). A Dual-Process Model of Cheating Intentions. *Journal of Marketing Education*, *28*, 193–204. doi:10.1177/0273475306291464

Smith, K., Davy, J., Rosenberg, D., Haight, T., & G. (2003). *A structural modeling investigation of the influence of demographic and attitudinal factors and in-class deterrents on cheating behavior among accounting majors* (pp. 45–65). Journal of Accounting Education.

Van Der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, *73*, 365–384. doi:10.1007/s11336-007-9046-8

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). *Why do only some events cause learning during human tutoring? Congnition and Instruction*. 209-249.

Watson, G., & Sottile, J. (2010). *Cheating in the Digital Age: Do Students Cheat More in Online Courses?*.

Wellman, G. S. (2005). Comparing learning style to performance in on-line teaching: Impact of proctored v.s. un-proctored testing. *Journal of Interactive Online Learning*, *4*(1), 20–39.