

MAKING SUMMATIVE ONLINE ASSESSMENT MORE SECURE: INVESTIGATING TIME, CONTENT, AND QUESTION TYPE

Thomas Lee Hench

ABSTRACT

To date, online assessment has been widely used in the formative capacity, wherein the results of the assessment are not graded and the assessment may be retaken a specified number of times. In this capacity, formative online assessment has been used successfully as a diagnostic tool to provide for prompt feedback and interaction on assessments such as self-tests and pre-tests. In these applications, the concern over security is minimal since there is no compelling motivation for students to cheat. While sharing some of these positive features with online formative assessment, summative online assessment does have certain drawbacks, with the problem of student identity and the use of prohibited materials during the assessment among the major concerns. In terms of the latter concern, the time available to complete the assessment is one factor that may impact security; shorter exam times may result in less time to access prohibited materials. Another factor is the type of question type used in the assessment. For example, a multiple choice question may be more readily answered without the need to turn to prohibited materials than a true/false question. In addition, the inclusion of additional content in the exam such as a list of relevant terms again may factor into the need to reference prohibited materials. This paper describes the investigation of the effects of these three factors on performance for an online Astronomy course with the goal of establishing a potentially more secure summative online testing environment while not negatively impacting assessment performance. The results of the investigation are then compared to those obtained for identical exams taken in a proctored environment with the conclusion that the exam question type and content are the critical factors in determining exam performance and may be used to increase the security of an online testing environment.

KEYWORDS

Teaching and learning in science, online security

INTRODUCTION

Online assessment, whether formative or summative, provides distinct advantages over “paper” assessments. In particular, the ability to provide faster feedback at the individual level and to monitor group performance, less time grading, greater opportunities for practice are among the top reasons for using online assessment. On the down side, the need for some level of student technical ability, anxiety about the reliability of the technology employed, and security issues (i.e., Who is actually taking the assessment? Is cheating occurring?) are cited as some reasons to be cautious when using this mode of assessment. Indeed, research by the author (Hench, 2005) has suggested that some level of cheating probably occurs in online assessment while other research suggests that the level of cheating online may be no different from that in proctored situations (Kolitsky, 2008). While recognizing the importance of the disadvantages, the focus of this paper is on investigating what may be done to minimize the opportunity of cheating on online exams. To start, the factors that impact the opportunity to cheating on online assessments are identified. Then, data acquired to determine the effect of these impact factors on assessment performance are presented and the results of the research analyzed to find the optimal conditions for a more secure online assessment environment.

FACTORS IMPACTING CHEATING ON ONLINE SUMMATIVE ASSESSMENTS

One of the prime impact factors for reducing the opportunity for cheating on online is the length of time provided for taking the assessment. Shorter periods of allotted time result in less time available for the use of unauthorized materials. If cheating is reduced by this action, then scores should be lower. However, if the time is reduced too much, the scores may also decrease due to lack of sufficient time to complete the assessment. Thus, if cheating is occurring during online assessments, reducing the length of time would also result, in addition to less cheating, in a reduction in the scores obtained by those taking the assessment and an unclear view of student performance.

The inclusion of a list of relevant terms is another impact factor which limits the opportunity to cheat. These terms possess the potential to act as retrieval cues, which are prompts or hints that can help memory retrieval. Thus, inclusion of a list of relevant terms would also decrease the need to access unauthorized materials. While the inclusion of a terms list may reduce the need for cheating, its omission could potentially penalize students and result in lower scores whether or not cheating is occurring.

A final impact factor is the type of question used in the assessment. Questions such as fill in the blank or true/false may lead to more cheating as a result of the need to access prohibited materials. Multiple choice questions, on the other hand, contain “built-in” retrieval cues, as do a figure labeling questions where visual retrieval cues are provided. Hence, the latter two question types may be of more benefit in reducing cheating. Just as important is the level of learning associated with the question. Certain question types may be able to assess more than lower level cognitive skills, such as recall or comprehension, while others may work for those higher order skills. Thus, the goal of this study is twofold: reduce the opportunity for cheating on online assessments and maintain acceptable levels of student performance (Table 1).

Table 1. Impacts and instructional goals

Impact Factor	Goal: Reduction in cheating and
Assessment time	Maintenance of overall performance
List of terms	Maintenance of overall performance
Question type	Accommodate student preference

Since the three identified impact factors possibly share an interdependency, their combined optimization is necessary if an effective online assessment tool is to be produced.

EFFECT OF IMPACT FACTOR OPTIMIZATION ON ASSESSMENT PERFORMANCE

In the optimization that follows, the effect of impact factors on performance was carried out in online sections of an Introduction to Astronomy course taught by the author. The first factor investigated was assessment time. In particular, two different time limits were used, namely 35 minutes and 120 minutes, for assessments in which terms were included or not included. The results of this part of the optimization are shown in Table 2, with the corresponding statistical significances presented in Table 3.

Table 2. Length of assessment time and term list inclusion matrix

		35 min limit		120 min limit			
Exam #		Score	Time (min)		Score	Time (min)	
1	Terms	18.8 ± 2.3	12.9 ± 6.2	Terms	18.5 ± 2.9	14.3 ± 6.6	
2	N = 30	20.1 ± 2.9	20.3 ± 7.1	N = 41	20.8 ± 2.8	22.0 ± 9.5	
3		13.8 ± 2.7	30.2 ± 10.8		15.0 ± 3.0	42.6 ± 20.8	
4		16.8 ± 2.3	27.9 ± 8.3		*	16.8 ± 3.1	35.5 ± 13.2
1		No Terms	14.9 ± 5.8		11.6 ± 5.6	No Terms	13.8 ± 6.4
2	N = 42	17.8 ± 3.2	24.1 ± 7.1	N = 38	17.5 ± 4.7	30.8 ± 14.4	
3		13.8 ± 3.2	28.3 ± 8.6		15.3 ± 3.7	48.3 ± 24.3	
4		15.1 ± 3.7	27.5 ± 7.6		16.8 ± 2.5	41.9 ± 23.0	

*N = 25 for Terms/120 min Exam#4

Table 3. Statistical analysis of length of assessment time and scores

Score Terms/35 min vs. Terms/120 min			Score No Terms/35 min vs. No Terms/120 min		
Exam#	t-test value	Significance p = 0.05	Exam#	t-test value	Significance p = 0.05
1	0.47	N	1	0.81	N
2	-1.03	N	2	-0.34	N
3	-1.74	N	3	-1.94	N
4	0	N	4	-2.38	Y

As seen in Table 3, changing the time limit for the assessment did not significantly impact scores between the 35 minute and 120 minute sections where the terms list was included in the assessment. In addition, there was no significant difference in scores between sections where a terms list was not included. However, when controlling for the time limit, a significant difference is noted in 3 of the four assessments with a 35 minute limit and two of four assessments with the 120 minute limit (Table 4).

Table 4. Statistical terms list inclusion and scores

Score No Terms/35 min vs. Terms/35 min			Score No Terms/120 min vs. Terms/120 min		
Exam#	t-test value	Significance p = 0.05	Exam#	t-test value	Significance p = 0.05
1	3.49	Y	1	4.26	Y
2	3.12	Y	2	3.82	Y
3	0	N	3	-0.40	N
4	2.23	Y	4	0	N

Furthermore, there was little significant difference in the actual time students used to complete the assessments regardless of time or term inclusion. (Table 5).

Table 5. Actual lengths of time comparison

Time Terms/35 min vs. No Terms/35 min			Time Terms/120 min vs. No Terms/120 min		
Exam#	t-test value	Significance p = 0.05	Exam#	t-test value	Significance p = 0.05
1	0.93	N	1	-1.99	N
2	-2.24	Y	2	0.44	N
3	0.83	N	3	-1.12	N
4	0.21	N	4	0.81	N

Thus, for the range of times investigated, the main factor impacting performance when attempting to reduce the opportunity for cheating was found to be the omission or inclusion of a terms list. In a survey administered to the students at the mid semester point, the importance of the inclusion of a terms list was also indicated. As shown in Table 6, a large majority of students preferred the inclusion of a terms list for use in taking the assessment and as a way of reducing the opportunity to cheat.

Table 6. Student preference for terms list inclusion and impact on cheating

	For use in taking the assessment (%)	For use in reducing cheating (%)
Terms included	89	88
Terms omitted	0	12
No opinion	11	X

In addition, 54% of the students surveyed indicated that the omission of a terms list from an assessment would have a negative or very negative effect on their scores (Table 7).

Table 7. Student responses on effect of terms list inclusion and impact of question type

	Effect on assessment score if terms list is omitted (%)	Effect on assessment score on question type (%)
very negative effect	4	0
negative effect	50	13
no effect	39	32
positive effect	4	49
very positive effect	3	7

The data included in Table 7 also suggests that students feel that the appropriate choice of question type is an important factor in determining the resulting score. As concerns what type of questions students prefer, half chose the multiple choice format, as shown in Table 8. This result is also consistent with fact that the “built-in” retrieval cues of multiple choice questions provides students with a greater opportunity to produce the correct answer (Marsh, 2003). Furthermore, 47% of students indicated that this form of question also would be the best type to reduce the opportunity to cheat, a result suggesting a possible awareness on behalf of the students of the retrieval cues. As the online assessments described here were used to determine the level of recall and comprehension, multiple choice questions were well suited. In addition, this type of question may also be used to address higher order skills (CTL, 1990).

Table 8. Student question type preference and impact on cheating

	Preferred question type (%)	Question type and reducing cheating (%)
True/False	21	23
Multiple choice	50	47
Fill in the blank	5	13
Figure labeling	24	17

DISCUSSION AND CONCLUSIONS

The following observations were obtained from the optimization of the factors impacting the opportunity to cheat on an online assessment –

- I. The length of time allotted for completion of the assessment without negatively impacting scores is independent of the scores for the range of times investigated. However, there obviously exists a time threshold after which scores will decrease due to lack of sufficient time. In practice, this threshold must be determined and the length of assessment time put as close as possible to insure an acceptable level of performance and minimize the opportunity to cheat.
- II. The inclusion of a list of terms in the assessment showed a significant improvement in scores and as previously discussed minimizes the need to access unauthorized material. Thus, the inclusion of the terms list and the retrieval cues it provides should be used along with the time threshold determination to establish the optimum conditions for reducing cheating.
- III. The type of question used in an optimization of online security for assessments is determined here as multiple choice. Not only was this question type preferred by most students, multiple choice question answers also provide retrieval cues which have the potential for both maintaining performance while limiting the need to use unauthorized material.

FUTURE WORK

Using the results of this study as a baseline, the author intends to further investigate the use of multiple choice questions to address higher order thinking skills as part of summative online assessments. These assessments will then be subjected to the optimization procedure described here to determine the parameters necessary to achieve the highest possible level of security.

REFERENCES

- CTL, (1990) Improving Multiple Choice Questions, For Your Consideration: Suggestions and Reflections on Teaching and Learning, Number 8, November 1990, Center for Teaching and Learning, University of North Carolina at Chapel Hill
- Hench, T. (2005) Online Versus Proctored Testing: How Do They Compare?, Proceedings of the 7th International Conference on Computer-Based Learning in Science, Zilina, Slovakia, July 2 – 6, 2005.
- Kolitsky, M. (2008). Analysis of Non-proctored Anti-cheating and Formative Assessment Strategies, *e-mentor*, Number 4, 2008. This paper is available online at http://www.ementor.edu.pl/_xml/wydania/26/582.pdf
- Marsh, E. (2003). Myths, Mysteries, and Realities, in J. Guthrie (Ed.), *Encyclopedia of Education* 2nd Edition (pp. 1605-1609), New York: MacMillan.

Thomas Lee Hensch
Professor of Physics and Astronomy
Delaware County Community College
Media, PA 19063 USA
Email: thensch@dccc.edu

**INFORMATION &
COMMUNICATION TECHNOLOGY**

