

Reliability, Compliance and Security of Web-based Pre/Post-testing

Scott Bonham

*Department of Physics and Astronomy, Western Kentucky University
1906 College Heights Blvd. #11077, Bowling Green, KY, 42101-1077*

Abstract. Pre/post testing is an important tool for improving science education. Standard in-class administration has drawbacks such as ‘lost’ class time and converting data into electronic format. These are not issues for unproctored web-based administration, but there are concerns about assessment validity, compliance rates, and instrument security. A preliminary investigation compared astronomy students taking pre/post tests on paper to those taking the same tests over the web. The assessments included the Epistemological Beliefs Assessment for Physical Science and a conceptual assessment developed for this study. Preliminary results on validity show no significant difference on scores or on most individual questions. Compliance rates were similar between web and paper on the pretest and much better for web on the posttest. Remote monitoring of student activity during the assessments recorded no clear indication of any copying, printing or saving of questions, and no widespread use of the web to search for answers.

Keywords: pretest, posttest, computer assessment.

PACS: 01.40.Fk, 01.50.ht, 01.50.Kw

INTRODUCTION

Pre/post testing has become a very important tool in Physics Education Research. The Force Concept Inventory¹ (FCI) has clearly exposed shortcomings in traditional instruction,² transformed individuals³ and entire departments⁴ and spurred the growth of the Physics Education Research community. It has been used to compare the effectiveness of different curricular approaches,⁵ evaluate the use of technology,⁶ look at gender fairness in instructional methods,⁷ and answer many other research questions. The success of the FCI has inspired the development of a wide range of other assessment tools, covering topics including mechanics,⁸ electricity and magnetism,⁹ DC circuits,¹⁰ astronomy¹¹, problem-solving,¹² uncertainty,¹³ and beliefs about the nature of science.¹⁴

In spite of their utility in revealing the state of our students and the effectiveness of instruction, these assessments are not as widely used as is desirable. An important reason is the time and effort involved in administering them during class and processing the forms afterwards. An alternative is to have students complete the assessments outside of class using a web-based system. This saves instructional time, reduces errors and puts the data directly into electronic format. With the appropriate tools built in, it can quickly and

easily make feedback from the assessment available to instructors. While there are clear advantages, there are also important questions about the validity, compliance and security in this approach. The purpose of this study is to address those.

Research Questions

1. Will administration in an unproctored web environment affect how students respond? Changing administration modes could affect the reliability of items, and students could use resources not available in the classroom.
2. Will administration in an unproctored web environment affect compliance rates? Those who show up to class on the given day and those who comply with instructions to do something outside of class are not necessarily similar groups.
3. Will administration in an unproctored web environment compromise the security of the test? If an assessment is on the web, there is no way to stop students from keeping copies of questions in some format, which would compromise the validity of the instrument.

These research questions are listed in order of increasing seriousness and decreasing amounts of information in scholarly literature. Research on computer testing shows that in most cases, computer-

administered tests are equivalent or even superior to paper-administered ones¹⁵. Most studies show that scores are usually equivalent and that people tend to finish computerized tests quicker and find the experience more enjoyable (or at least less disagreeable). However, most of these studies have compared proctored paper-and-pencil tests with proctored computer assessments, where the students come into a computer testing center and work under supervision. In the study most similar to this one¹⁶ half of a class took the FCI in class and the Maryland Physics Expectation survey (MPEX)¹⁷ on line, with the mode reversed for the other half of the class. No difference was found in the scores on the FCI for the two groups (MPEX scores were not reported), though the other questions above were not addressed. The web administration system used was rudimentary, in that students were not required to log in and all questions were displayed on a single page, requiring students to scroll up and down. No post-testing and no follow-up studies were reported.

SETTING

A preliminary study was carried out using in two pairs of introductory astronomy class sections taught at Western Kentucky University the spring of 2006, one pair in the Astronomy of the Solar System course and in the Astronomy of the Stellar Systems course. Table 1 summarizes the classes in this study. Instructors A and B are highly experienced instructors, and Instructor C is a newer instructor. Classes 1 and 2 were taught in a lecture hall, while classes 3 and 4 were in one of the department's SCALE-UP style classrooms¹⁸ and included hands-on activities along with lecture.

Table 1: Summary of courses involved in the study. N is number of students on the roster at the end of the semester.

Class	Course	Instructor	N	Paper	Web
1	Solar	A	71	ASTRO	EBAPS
2	Solar	A	64	EBAPS	ASTRO
3	Stellar	B	40	ASTRO	EBAPS
4	Stellar	C	39	EBAPS	ASTRO

Two assessments were used with each class. One is a twenty-four item multiple choice astronomy concept assessment developed for this study,¹⁹ which will be referred to as the ASTRO. This was a quickly developed assessment with no effort to check the reliability or validity that was used as a "stand-in" to avoid the risk of compromising a well-designed astronomy concept assessment. The other assessment was the Epistemological Beliefs Assessment for Physical Science (EBAPS).²⁰ This is a thirty item assessment with five subscales: *Structure of scientific*

knowledge, Nature of knowing and learning, Real-life applicability, Evolving knowledge and Source of ability to learn. Because of the 'opinion' nature of the questions it was felt that security was less of an issue with this assessment, and permission was given by the author for web use.²¹

In each of the pair of classes, one section took the WKU-ASTRO assessment in class using paper and pencil while the other section took the EBAPS assessment in class. Students were then asked to take the other assessment on-line outside of class during a specified period (a little over a week) and class credit was given to those who complied. Both assessments were given pre/post in classes 1 and 2, while the paper assessment was not given at the end of the semester in classes 3 and 4.

SYSTEM

The web-based assessment system was developed by the author using PHP and MySQL. Students log onto the system using their school username and student identification number. They then receive an instruction page, where they are asked to complete the exam at one sitting, not use other resources, and not print, save or copy material. They are then presented with the currently available assessment for their class. Questions are delivered one at a time and by default the system automatically advances to the next question once the student selects one of the choices. There are controls to go back to previous questions, skip ahead to following questions, and to mark a particular question to return to latter. Once the student has gone through all of the questions, the system will return to any skipped questions and marked questions before finishing the assessment.

On the posttest, a script was included to monitor any behavior that could potentially compromise the integrity of the test, such as searching for answers on the web, copying material to another application, saving a page and printing. JavaScript and Java were used to capture events (actions) that could signify such activities and send information about them to be recorded on the server. Potential copying and searching for answers on the web was identified using the focus gained and focus lost events of the browser, which happen every time the browser becomes and ceases to be the active window on the screen. On Microsoft Internet Explorer (MSIE) the contents of the clipboard were also recorded to compare with question text. Printing was monitored on MSIE by events signaling the start and ending of printing, and on non-MSIE browsers by a small, virtually invisible embedded Java Applet. Finally, loading of pages was monitored to detect any that had not been recently sent

by the server, which would indicate a saved page being reloaded. It should be noted that there is innocent activity that could trigger many of these events and that these methods are not entirely fool-proof. It is possible for individual events to escape detection, but unlikely that wide-spread violations could.

RESULTS

Validity

First the scores on the two exams were compared between those completed in the traditional format and those on the web. Table 2 summarizes the two-tailed *t*-tests results and shows no significant differences on the ASTRO and all five EBAPS scales. A question-by-question comparison using the chi-squared test for independence between groups identified two questions (9 and 19) that demonstrated differences significant at the $\alpha=0.05$ level for all courses combined but not comparing class 1 with class 2 and class 3 with class 4. One would expect that on a 30 item inventory, on average between one and two items will be by random chance found significant at the $\alpha=0.05$ level. On the ASTRO, there was no difference in the over-all scores, but an item-by-item comparison identified four questions with significant differences at $\alpha=0.02$ or less dealing with brightness of stars, nature of shooting stars and weightlessness on the space station. This set included a question with an awkwardly worded correct option and one where the first distracter was mostly correct.

Table 2: Summary of two-tailed *t*-tests comparing scores on the ASTRO assessment and five subscales on the EBAPS.

	M_{web}	M_{paper}	t	p
ASTRO	8.50	8.57	-0.12	0.90
Structure sci. knowledge	2.11	2.11	0.64	0.95
Nature knowing/learning	2.50	2.49	0.16	0.87
Real-life applicability	2.58	2.48	1.2	0.23
Evolving knowledge	2.50	2.48	0.12	0.91
Source of ability	3.02	2.88	1.4	0.16

Compliance

The compliance rates for the four classes are summarized in Table 3. This table lists the enrollment fraction of students on the roster at the end of the semester that took each of the assessments. The pretest compliance rate for the web is nearly as high as the paper test for all except class 4. In that class the instructor encouraged students to participate but did not give any course credit on the pretest. Extra credit

was given on the post test, leading to an increase in compliance. It may also be observed that the posttest compliance rates on the web are nearly as high as the pretest rates, but that the posttest on paper is much lower than the pretest for classes 1 and 2. The paper posttests were given during the last class session, late afternoon on the last class day of the semester. The difference in compliance was mainly among freshman, suggesting that the web and paper groups are not equivalent.

Table 3: Compliance rates for each class and assessment. Total posttest paper is for classes 1 and 2 only.

	Class 1	Class 2	Class 3	Class 4	Total
Enrollment	72	63	40	39	214
pretest paper	90%	84%	88%	92%	88%
pretest web	86%	83%	83%	54%	79%
posttest paper	58%	60%	N/A	N/A	59%
posttest web	81%	71%	75%	69%	75%

Security

During the post-test, on-line student activity was monitored as previously described, recording events that could indicate printing, saving, copying or other activities. The start print event in MSIE browsers and Java print() in others would indicate a student printing a page, but no events were recorded out of the 160 students who participated. Saving of pages was monitored by listening for page load events not paired to a page serve event. Again, no events were recorded up to and past the time of the student final exams.

Possible copying and searching for answers were monitored by focus gained/lost events, and on MSIE the contents of the clipboard were also recorded. No question text or any other clipboard contents were recorded. (78% of the students used MSIE.) Time during which the browser is not the active window can also be a signal of copying or searching for answers, as well as innocent activities like checking email or closing pop-up ads. Many thousands of focus lost events were generated, but most of these were caused by a page being submitted. Eliminating all events between submission and page load, those within a second of page load, and those without a paired focus gained event left 86 focus lost events that appear to be actual instances where some other window or menu became the active window for a period of time. Figure 1 provides a histogram of the time gap between the lost focus and corresponding gained focus event for these instances, except for a single one on the EBAPS with gap of 737 seconds. Excluding that event, the median gap on the EBAPS and WKU-ASTRO were 16 and 10 seconds, respectively. Of the 33 students where an instance was recorded, almost half involved

only a single instance, and over 80% had less than three instances. Nine out of twelve of the instances from the student with the most occurred on the first question, and the next highest, with 11 events, involved only 8 different questions.

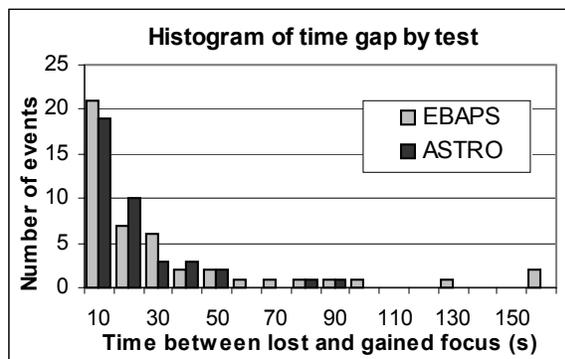


FIGURE 1. Histogram of elapsed time between focus lost and focus gained events. Median gap is 10 s for ASTRO and 16 s for EBAPS.

DISCUSSION

The results from this preliminary study are promising for using unproctored web assessment, but further study is needed. The difference in how the assessment was administered did not change response patterns on most items and there were no significant differences on the over-all scores. Furthermore, the items that showed strong format effects were all on the relatively poorly developed ASTRO assessment. The compliance rates at the beginning of the semester were nearly the same for all but one section. The instructor in that class did not provide any course credit for the web pretest, illustrating the importance of providing at least a minimal amount of class credit to ensure good compliance. Compliance rates were significantly better for the web on the posttest because of the number of students who did not come to class at the end of the semester. It appears that compliance rates on the web can be as good or even better than those using class time when some class credit is provided and the instructor makes expectations clear. This study implemented methods to monitor security of the tests, and found no signals that any of the material was being printed, copied or saved, or that there was extensive searching for answers on the web. While this method is not perfect or foolproof, it is clear that most if not all of the students complied with the request to not copy, print, save or use other resources for the questions. These preliminary results suggest that web-based pre/post assessments may provide an acceptable alternative to paper ones, though validity needs to be established for individual instruments, procedures to encourage participation need to be

attended to, and security needs to be considered. Future work will include repeating the study with a larger population, using additional instruments, and estimating failure rates of the monitoring system to detect suspicious activity.

REFERENCES

- David Hestenes, Malcolm Wells, and Gregg Swackhamer, *Physics Teacher* **30** (3), 141 (1992).
- Ibrahim Abou Halloun and David Hestenes, *American Journal of Physics* **53** (11), 1043 (1985); Richard Hake, *American Journal of Physics* **66** (1), 64 (1998).
- Eric Mazur, *Peer Instruction*. (Prentice-Hall, Upper Saddle River, NJ, 1997); Catherine H. Crouch and Eric Mazur, *American Journal of Physics* **69**, 970 (2001).
- Tim Stelzer and Gary Gladding, in *Forum on Education of The American Physical Society* (2001), Vol. 2002.
- Jeff Saul, University of Maryland, 1998.
- S.W. Bonham, D.L. Deardorff, and R.J. Beichner, *Journal of Research in Science Teaching* **40**, 1050 (2003).
- Mercedes Lorenzo, Catherine H. Crouch, and Eric Mazur, *American Journal of Physics* **74** (2), 118 (2006).
- Ronald K. Thornton and David R. Sokoloff, *American Journal of Physics* **66** (4), 338 (1998).
- Dave Maloney, Alan Van Heuvelen, Tom O'Kuma et al., (Joliet Junior College, Joliet, Illinois, 1999).
- Paula Vetter Engelhardt and Robert J. Beichner, *American Journal of Physics* **72** (1), 98 (2004).
- Beth Hufnagel, *The Astronomy Education Review* **1** (1), 47 (2002).
- David Hestenes and Malcolm Wells, *Physics Teacher* **30** (3), 159 (1992).
- Duane Lee Deardorff, North Carolina State University, 2001.
- Edward F. Redish, Jeffery M. Saul, and Richard N. Steinberg, *American Journal of Physics* **66** (3), 212 (1998); Ibrahim Halloun, *International Conference on Undergraduate Physics Education* (1996); W. K. Adams, K. K. Perkins, N. Podolefsky et al., *Phys. Rev ST: Phys. Educ. Res.* **2** (1), 010101 (2006).
- David Zandvliet and Pierce Farragher, *Journal of Research on Computing in Education* **29** (4), 423 (1997); Allan C. Bugbee, *Journal of Research on Computing in Education* **28** (3), 282 (1996).
- Dan MacIsaac, Rebecca Pollard Cole, David M. Cole et al., *Electronic Journal of Science Education* **6** (3) (2001).
- Edward F. Redish, Richard N. Steinberg, and Jeffery M. Saul, *American Journal of Physics* **66**, 212 (1998).
- R. Beichner, J. Saul, D. Abbott et al., in *PER-Based Reform in University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, in press).
- The effort of Richard Gelderman and Micheal Carini in developing this assessment are gratefully acknowledged.
- Andy Elby, *American Journal of Physics* **69** (7), S54 (2001).
- Andy Elby, personal communication.