Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Engaging Students to Think Critically in a Large History Class – Appendix

Mairi Cowan, Tyler Evans-Tokaryk, Elaine Goettler, Jeffrey Graham, Christopher Landon, Simone Laughton, Sharon Marjadsingh, Caspian Sawczak and Alison Weir
University of Toronto Mississauga

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Cowan, M., Evans-Tokaryk, T., Goettler, E., Graham, J., Landon, C., Laughton, S., Marjadsingh, S., Sawczak, C., & Weir, A. (2014). *Engaging Students to Think Critically in a Large History Class – Appendix.* Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

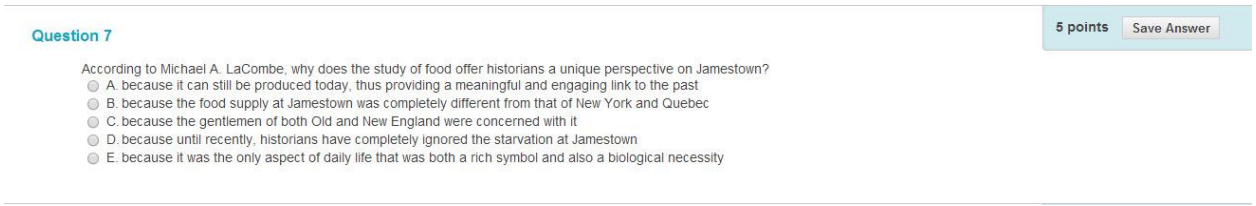
Table of Contents

List of Tables

List of Figures

Appendix 1: Sample Questions from Online Quizzes

Week 3 of the course dealt with North America as its "content" and finding sources as its "skill". The quiz on week 3 material opened immediately after the week 2 lecture and remained open until the beginning of the week 3 lecture. Students were introduced to historical content through questions about the assigned readings. The following is an example of a question about a journal article on colonial Jamestown.



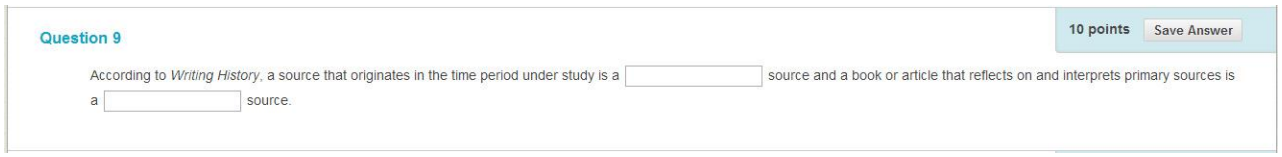
Question 7 — 5 points — Save Answer

According to Michael A. LaCombe, why does the study of food offer historians a unique perspective on Jamestown?

- A. because it can still be produced today, thus providing a meaningful and engaging link to the past
- B. because the food supply at Jamestown was completely different from that of New York and Quebec
- C. because the gentlemen of both Old and New England were concerned with it
- D. because until recently, historians have completely ignored the starvation at Jamestown
- E. because it was the only aspect of daily life that was both a rich symbol and also a biological necessity

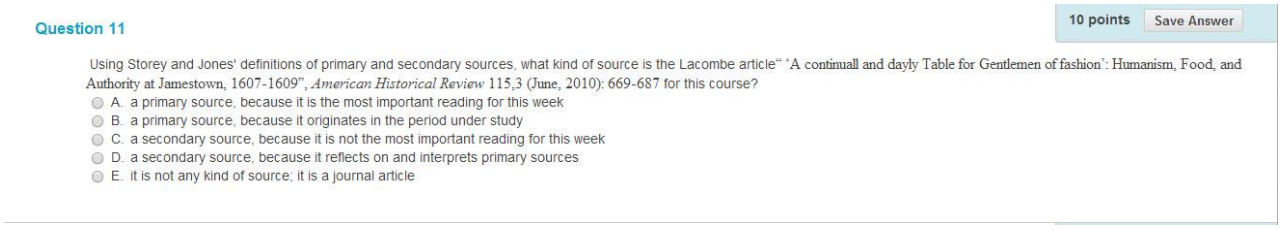
Feedback for the correct answer (e): Good.
Feedback for an incorrect answer: Not quite. Please read the article again, and pay especially close attention to the first few pages.

Students were also introduced to a historical skill through questions about the assigned readings. The following are examples of questions about the basic distinction between primary and secondary sources as discussed in the students' style guide:



Question 9 — 10 points — Save Answer

According to *Writing History*, a source that originates in the time period under study is a [ ] source and a book or article that reflects on and interprets primary sources is a [ ] source.



Question 11 — 10 points — Save Answer

Using Storey and Jones' definitions of primary and secondary sources, what kind of source is the Lacombe article" 'A continuall and dayly Table for Gentlemen of fashion': Humanism, Food, and Authority at Jamestown, 1607-1609", *American Historical Review* 115,3 (June, 2010): 669-687 for this course?

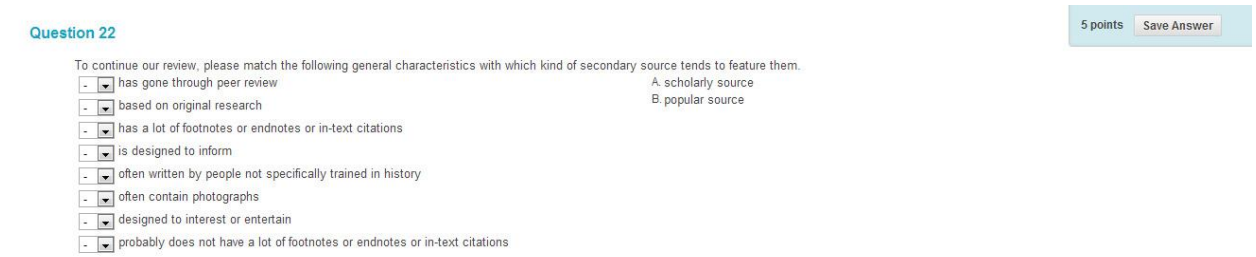
- A. a primary source, because it is the most important reading for this week
- B. a primary source, because it originates in the period under study
- C. a secondary source, because it is not the most important reading for this week
- D. a secondary source, because it reflects on and interprets primary sources
- E. it is not any kind of source; it is a journal article

In week 4, when the topics were "Mesoamerica and South America; Selecting Sources," students continued to develop their skills through the following series of questions that connects information from the previous week – the distinction between scholarly and popular sources – to material from the current week:

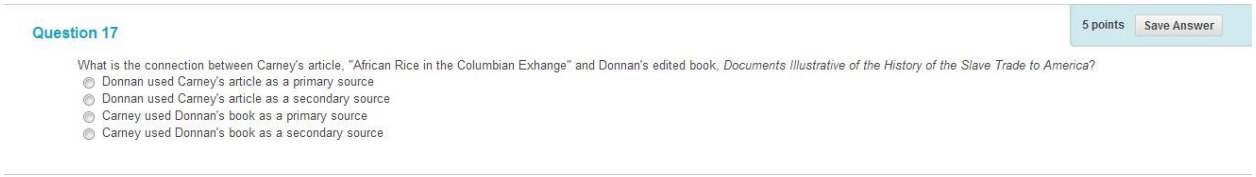
Question 21 — 2 points — Save Answer

The authors of *Writing History* warn on p. 23 that "at first you may find that your sources seem confusing or even contradictory", and they state that "an important part of writing history is deciding how to interpret and assemble material from sources that may raise more questions than they answer". With these insights in mind, let's review what you have been learning about the different kinds of sources historians use in their research. Which of the following is likely a good kind of secondary source for a research paper in a university history course?

- A. a scholarly source
- B. a popular source

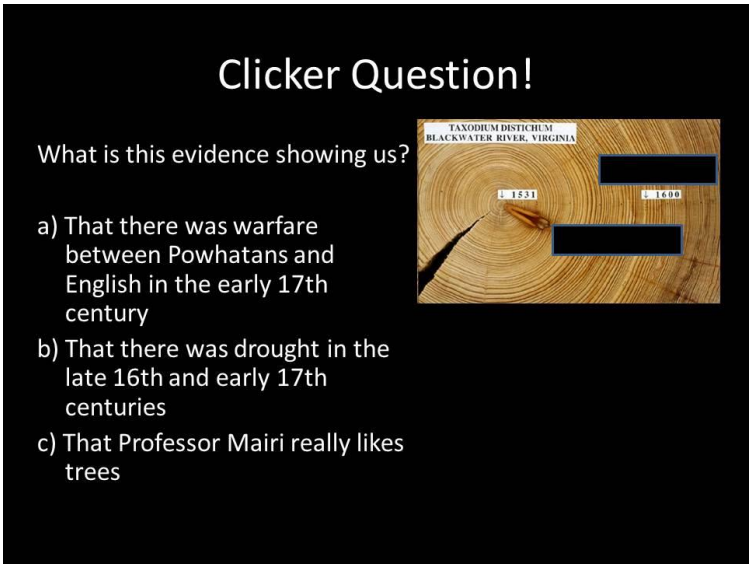


In week 5, when the topics were “The Atlantic World; Using Sources,” students were given questions to reinforce their understanding of using sources. The following is an example of a question that asks students about how an article being read for that week relates to a primary source also being read for that week:

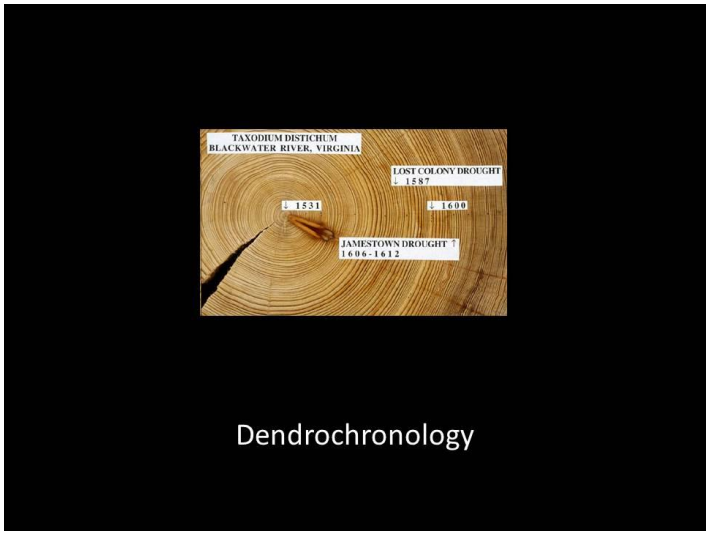


Appendix 2: Sample Clicker Questions

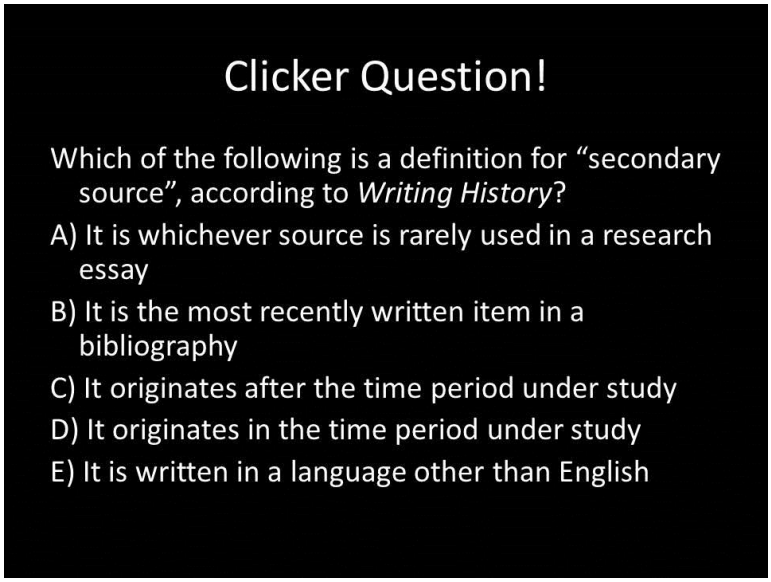
The following slide from week 3 gauged what students already knew about information that was about to be introduced in lecture. (Since they had not done any reading on this topic and had not heard about it at any earlier point in the course, they were not held responsible for knowing the correct answer and were therefore awarded full marks for whatever response they chose).



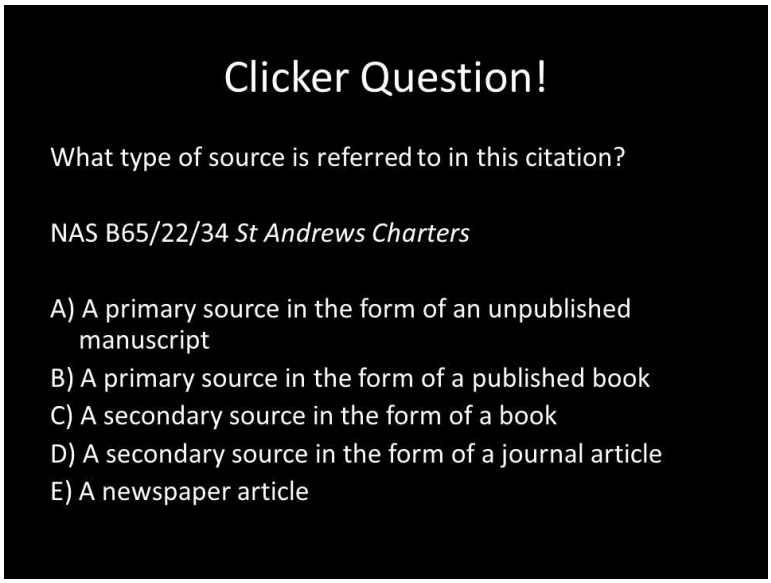
The blacked-out text was revealed on the following slide:



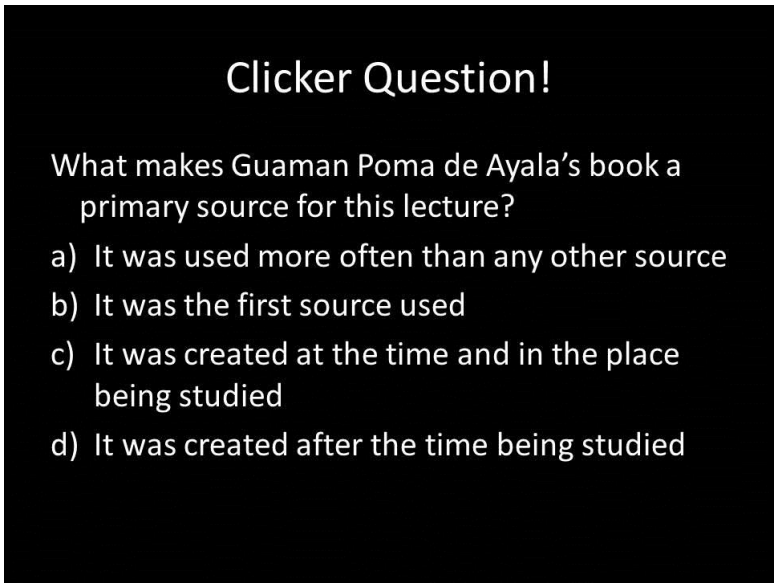
The following slide from week 3 asked students to recall information from one of the readings that should have been completed before the lecture:



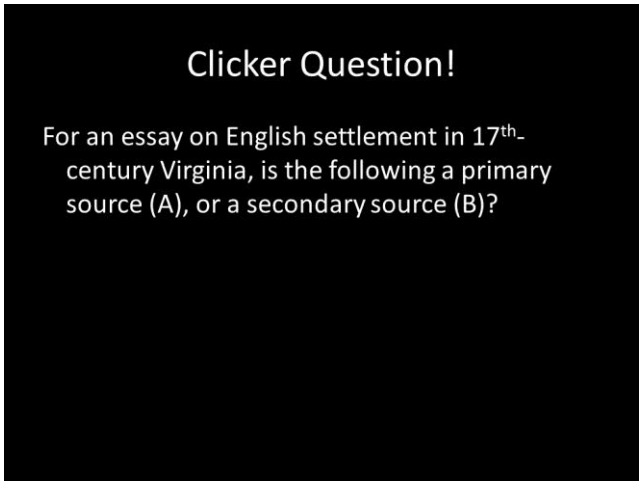
The following slide asked students to identify what kind of source is being indicated by a bibliographic reference:



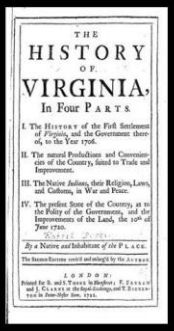
The following slide from week 4 asked students to reflect on one of the sources referred to in the lecture and to determine why it is a primary source. To make this determination, students would have to draw upon both their understanding of what constitutes a primary source generally and also their understanding of what is in this particular source:



Historiographical questions were introduced in the following series of slides from week 3:



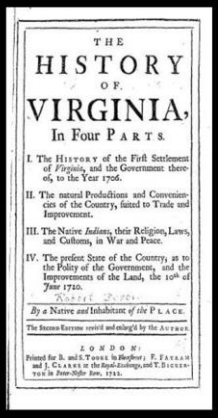
(English settlement in 17th-century Virginia)



A primary source

B secondary source

- For an essay on the 17th-century settlement, it's a secondary source;
- For an essay on 18th-century attitudes about the early settlement, it's a primary source!



Clicker Question!

For a research essay on student behaviour at the University of Moronto in the 15th century, which of the following would be the best to use as a secondary source?

(student behaviour at the University of Moronto in the 15th century)

A. Alison Chan, "The Sword was Mightier than the Pen: University of Moronto Students in the early 20th century", *Journal of the Canadian Historical Association* 13,3 (2010): 35-57.

B. Emmet Cameron, "The Moronto Uprising Again? Echoes of the 15th century among modern students", *The Globe and Mail*, November 13, 2009.

C. Bill Denniston, *The University of Moronto in the Fourteenth and Fifteenth Centuries* (Oxford: Oxford University Press, 1933).

D. Sakib Hassan, *Fighting with Quills: Student conflict at the Universities of Moronto and Stork in the fifteenth century* (Cambridge: Harvard University Press, 2009).

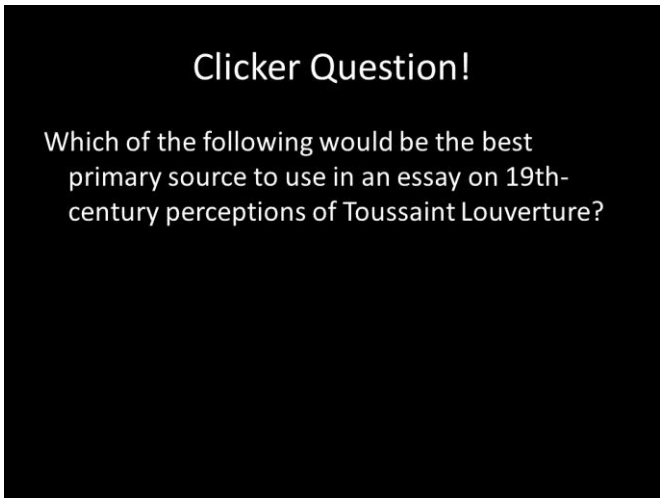
Clicker Question!

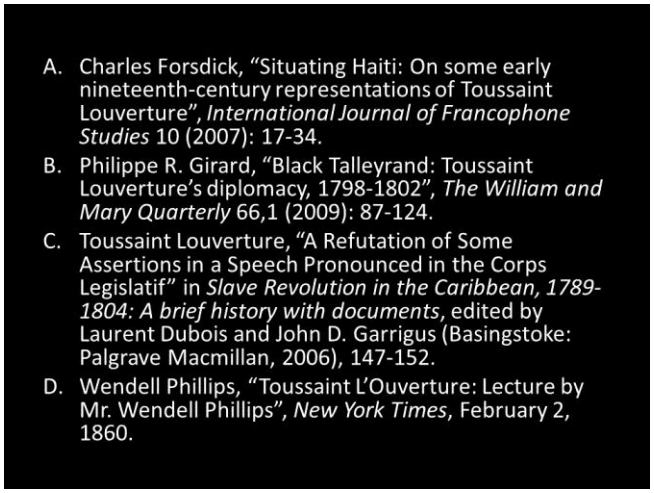
If you really wanted to use Arthur Denniston, *The University of Moronto in the Fourteenth and Fifteenth Centuries* (Oxford: Oxford University Press, 1933), how should you do so?

Bill Denniston, *The University of Moronto in the Fourteenth and Fifteenth Centuries* (Oxford: Oxford University Press, 1933).

A) As a primary source for views in the 1930s
B) As a secondary source for views in the 1930s
C) As an elegant doorstop

The following two slides (with one question between them) from week 5 asked students to apply their knowledge of how to identify a source from a citation and their knowledge of what constitutes a primary source in general to a specific historiographic situation (in this case, a situation about how people at one point in time perceived an earlier historical figure):





Appendix 3: Sample Questions from the Pre- and Post-Intervention Test

The first category of question asked students to apply their knowledge of source selection by choosing a primary source for a historical situation that is real but not covered in HIS101 by week 6. For example:

1. Fatima is writing a research essay on the causes of the witchcraft trials at Salem, Massachusetts in 1692 for one of her history courses at university. Which of the following would be the best for her to use as a primary source?

a) Gustav Henningson (ed.), *The Salazar Documents: Inquisitor Alonso de Salazar Frías and others on the Basque witch persecution* (Leiden and Boston: Brill, 2004).
b) Robert Earle Moody and Richard Clive Simmons (eds.), *The Glorious Revolution in Massachusetts: selected documents, 1689-1692* (Boston: Colonial Society of Massachusetts, 1988).
c) Benjamin C. Ray, "The Geography of Witchcraft Accusations in 1692 Salem Village", *The William and Mary Quarterly* 65,3 (July, 2008): 449-478.
d) Bernard Rosenthal (ed.), *Records of the Salem Witch-Hunt* (Cambridge: Cambridge University Press, 2009).
e) Caroline E. Upham, *Salem Witchcraft in Outline*, second edition (Salem, Mass.: Salem Press Publishing and Print Company, 1891).
f) I don't know.

The correct answer to this question would be (d), because it is a collection of primary sources (sources that were written at the time and in the place being studied) relevant to the research topic. The others are incorrect for various reasons: (a) is a collection of primary sources for a different place (the Basque region); (b) is a collection of primary sources for a different topic (politics in Massachusetts); (c) is a secondary source rather than a primary source; and (e) is also a secondary source rather than a primary source (though a less useful one than (c) because it is out of date).

The second category of question asked students to apply their knowledge of source selection by choosing a primary source for a hypothetical (unreal but realistic sounding) historical situation. For example:

3. Joel is writing a research essay on the overseas explorations of fifteenth-century Northlandish sailors during the reign of Queen Marguerite (1404-1453) for one of his history courses at university. Which of the following would be the best for him to use as a primary source?

a) Katherine Allen, *Whence Northland? The voyages of the fifteenth century* (New York: Young Readers Press, 1999).
b) Irene Hamilton (ed.), *Sea Life in the Ancient Greek and Roman World of the Second Century: readings in travel and exploration* (London: Routledge, 2000).
c) Carol Howard (ed.), *Sources of a Sea Empire: Northlandish documents from the fourteenth and fifteenth centuries* (Cambridge: Harvard University Press, 2001).
d) Stephen Jones (ed.), *Northlandish Land and Marriage Contracts: documents of social history* (Princeton, NJ: Princeton University Press, 1993).
e) Errol Singh, *Marguerite's Explorers: imperial ambitions and Northlandish voyages of the fourteenth and fifteenth centuries* (Toronto: University of Toronto Press, 2009).
f) I don't know.

The correct answer for this question would be (c).

The third category of question asked students to apply their knowledge of source selection by choosing a secondary source for a historical situation that is real but not covered in HIS101 by week 6. For example:

10. Tom is writing an essay on the fate of the European economy between the fall of Rome in the fifth century and the formation of the Carolingian empire by Charlemagne in the ninth century for one of his history courses at university. Which of the following would be the best for him to use as a secondary source?

a) Steven Guess, "The End of Empire: as America's global influence wanes, it can either learn from the Roman empire's mistakes or suffer the same fate", *The Guardian* (January 19, 2009).
b) Joachim Henning, "Strong Rulers – Weak Economy? Rome, the Carolingians, and the archaeology of slavery in the first millennium AD," Jennifer R. Davis and Michael McCormick (eds.), *The Long Morning of Medieval Europe: New Directions in Early Medieval Studies* (Aldershot: Ashgate, 2008): 33-53.
c) Gemma C. M. Jansen, Anna Olga Koloski – Ostrow and Eric M. Moormann (eds.), *Roman Toilets: their archaeology and cultural history* (Leuven: Peeters, 2011).
d) P. D. King (ed. and trans.), *Charlemagne: Translated Sources* (Kendal: Published by P.D. King, 1987).
e) J. W. Thompson, "The Commerce of France in the Ninth Century," *Journal of Political Economy* 23, 9 (November 1915): 857-887.
f) I don't know.

The correct answer to this question would be (b) because it is a secondary source (written sometime after the events being described) that fits the criteria given to students as a guide for selecting good secondary sources: it is relevant to the topic, scholarly and recent. The other answers are wrong for various reasons: (a) comes from a popular (not a scholarly) source; (c) is not relevant to the topic of research; (d) is a collection of primary sources; and (e) is not recent.

The fourth category of question asked students to apply their knowledge of source selection by choosing a secondary source for a hypothetical (unreal but realistic sounding) historical situation. For example:

12. Carol is writing a research essay on the economic difficulties facing Nipsan before the Peasants' Revolution of 1783 for one of her history courses at university. Which of the following would be the best for her to use as a secondary source?

a) Janice Chen, "The Crisis in the Counting-House? Financial irresponsibility in eighteenth-century Nipsan reconsidered", *Nipsanese History* 23 (2009): 312-356.
b) Jen Pangwa, *Nipsanese Military Organization, 1720-1800* (New York: Columbia University Press, 2009).
c) Peter Passmore, *Drying Flowers: A novel of love and loss in the Nipsanese Peasants' Revolution* (London: Thatcher & Sons, 1785).
d) Brian Todden, "Nipsan: A Land of History," *Toronto Star*, June 19, 2012.
e) John Wainwright, *Causes of the Nipsanese Peasants' Revolution* (Oxford: Oxford University Press, 1933).
f) I don't know.

For this question, (a) would be the correct answer.

The fifth category of question on the pre- and post-intervention tests asked students to apply their knowledge of source selection in the context of a historiographical investigation (an investigation into the history of history), where they were to choose a primary source for an essay about how people in a specific historical period viewed an earlier historical period. For example:

19. Elizabeth is writing an essay on nineteenth-century perceptions of the Mughal Empire at the height of its power, which was from about 1550 to 1700, for one of her history courses at university. Which of the following would be the best for her to use as a primary source?

a) Muzaffar Alam and Sanjay Subrahmanyam, "Witnesses and Agents of Empire: Eighteenth-century historiography and the World of the Mughal Munshi", *Journal of the Economic and Social History of the Orient* 53,1 (January, 2010): 393-423.
b) Ali M. Athar, *Mughal India: Studies in polity, ideas, society, and culture* (New Delhi, Oxford: Oxford University Press, 2006).
c) Manik Lal Gupta (ed.), *Sources of Mughal History, 1526 to 1740* (New Delhi: Atlantic Publishers, 1989).
d) Edward Singleton Holden, *The Mughal Emperors of Hindustan, A.D. 1398-A.D. 1707* (New York: Scribner, 1895).
e) Reddy C. Srinivasa, "Mughal Historiography", *Social Scientist* 21, 1 (1993): 105-108.
f) I don't know.

The correct answer would be (d) because it is a primary source for nineteenth-century attitudes about the Mughal Empire (since it was written during the nineteenth century and is about the Mughal Empire). The other answers are incorrect for various reasons: (a) is a secondary source about eighteenth-century attitudes towards the Mughal Empire; (b) is a secondary source about the Mughal Empire; (c) is a collection of primary sources from the Mughal Empire; and (e) is a secondary source about the historiography of the Mughal Empire.

The sixth and seventh categories of question asked students to identify what kind of source is listed in a real or hypothetical reference. For example:

21. What kind of source is the following?
Olivia Remie Constable (ed.), *Medieval Iberia: Readings from Christian, Muslim, and Jewish sources* (Philadelphia: University of Pennsylvania Press, 1997).

a) A primary source in the form of an unpublished manuscript
b) A primary source in the form of a published book
c) A secondary source in the form of a book
d) A secondary source in the form of a journal article
e) A newspaper article
f) I don't know.

The correct answer to this question would be (b).

The final category of question on the pre-intervention test was on confidence and metacognition. Students were asked:

How confident are you in your answer to [the previous question]?

a) Very confident
b) Confident
c) Neutral
d) Not very confident
e) Not at all confident

And at the end of the test, students were given this question:

Hooray, it's over! If you had to guess, how well do you think you did?

a) 80% - 100% correct
b) 70% - 79% correct
c) 60% - 69% correct
d) 50% - 59% correct
e) 25% - 49% correct
f) 0% - 24% correct

Appendix 4: Additional Information Related to Statistical Analysis

Data Summary

In the fall semester, 280 students were enrolled in the course by the end of the term; of these, 278 had a final mark greater than 0. 211 students (75.9%) were full participants in this exploratory study (i.e., they had signed the consent form, which covered all portions of the study, written both the pre-intervention test and the post-intervention test, and achieved a mark greater than 0 by the end of the semester). In the winter semester, 218 students were enrolled in the course by the end of the term; of these, 202 had a final mark greater than 0. 86 students (42.5%) were full participants in the study. In both the fall and winter semesters, students were excluded either because they had not signed a consent form, or because they had not written both the pre-intervention test and the post-intervention test, or because they had earned a mark of 0 at the end of the term. In the fall semester, 156 students (56.1%) signed the consent form and provided a response to the question about selecting a secondary source for the RPSS. In the winter semester, 62 students (30.7%) signed the consent form and provided a response to the question about selecting a secondary source for the RPSS. In the fall semester, 203 students (72.5%) signed the consent form and wrote the final examination at the regular time and place (i.e., not deferred or with AccessAbility services). In the winter semester, 83 students (38.1%) signed the consent form and wrote the final examination at the regular time and place. Participation and response rate data can be found in Table 9 below (see also Table 1).

Table 9: Number of Participants and Response Rates

Semester	Total # of students	# with final mark > 0	# who signed the consent form and wrote both pre- and post-intervention tests	Response rate for pre- and post-intervention tests	# who signed the consent form and provided relevant data on RPSS assignment	Response rate for those who signed the consent form and submitted RPSS	# who signed the consent form and wrote the final exam	Response rate for those who signed the consent form and submitted final exam
Fall	280	278	211	75.9%	156	56.1%	210	75.5%
Winter	218	202	86	42.5%	62	30.7%	83	41.1%
Total	498	480	297	61.9%	218	43.8%	293	58.8%

The number of participants and the response rate in the winter semester, without the intervention, were lower than the number of participants and the response rate in the fall semester with the intervention. Below we investigate any potential implications of these differences and determine that these different response rates are not a confounding factor in our statistical analysis. (The Ethics Review Board has approved our use of final marks from students who did not sign a consent form for the purposes of investigating potential implications of different response rates.)

Figure 4: Histograms of HIS101 Final Marks by Term and Respondent Status

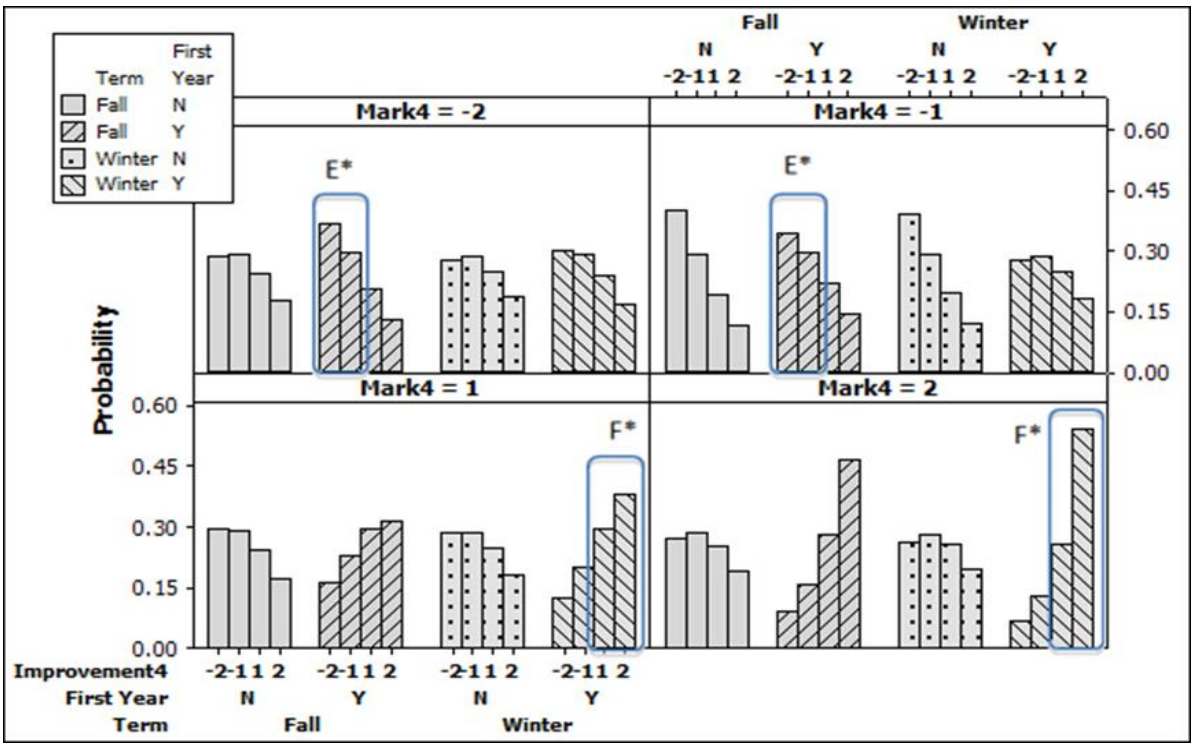
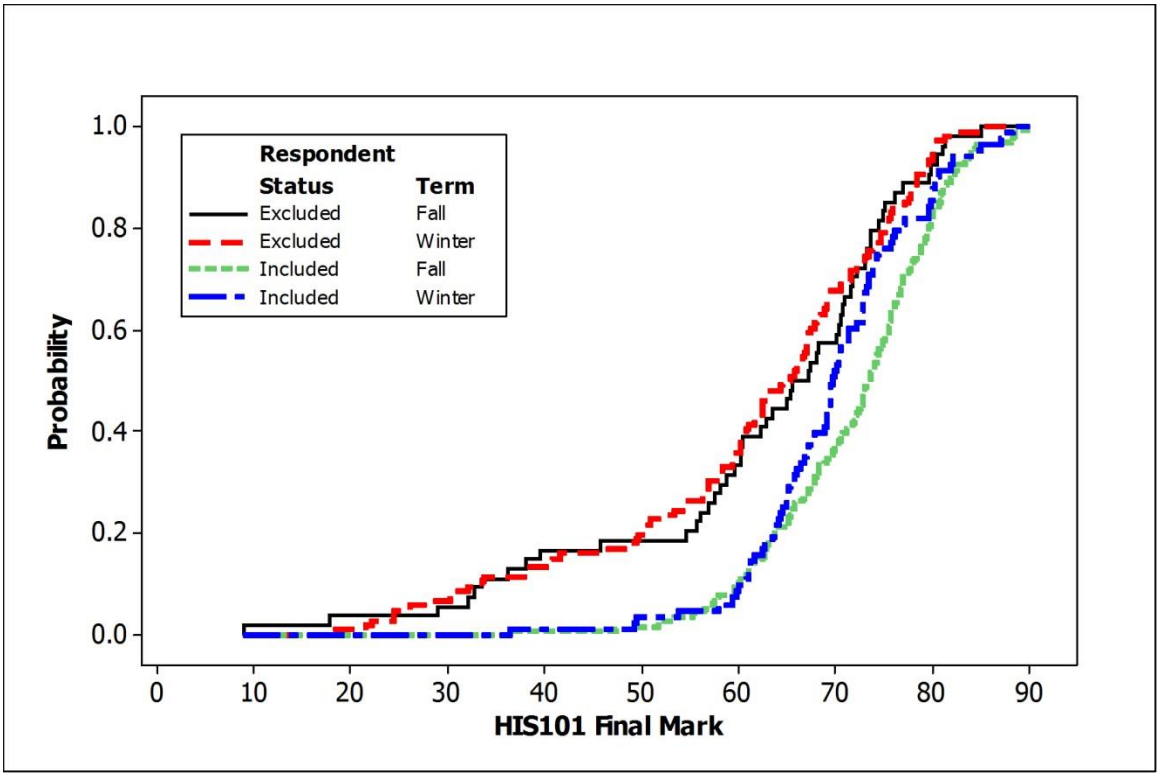


Figure 5: Empirical Distributions of HIS101 Final Marks in Each Semester Split by Respondent Status (included in or excluded from our analysis)



We also compared the distributions of HIS101 final examination marks across the two semesters and the two respondent status groups. We believe that final examination marks are a good indication of students' mastery of course material since these examinations were cumulative, included material taken from readings and lectures, and featured a variety of question types (multiple choice, short answer and essay format).

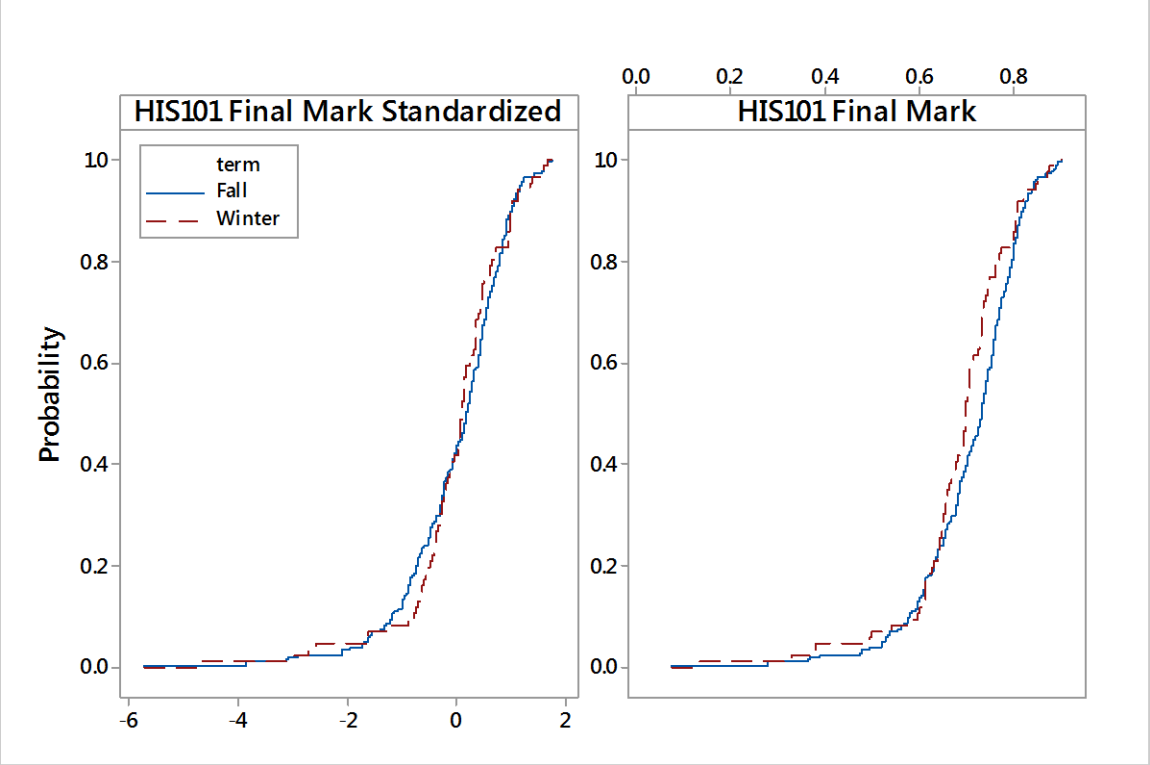
The unique characteristics of the mark distributions shown in Figures 4 and 5 imply that hypothesis tests for equality of summary parameters (means, medians, and standard deviations) are too coarse to assess adequately the equality of the distributions; the normality assumption required for these tests is clearly violated (Elliott & Woodward, 2007).

As shown in Figures 4 and 5, the distributions of the final marks of students in our exploratory study are not the same in both semesters (Kolmogorov-Smirnoff test, p=0.029). There are significant differences between the distribution of final marks for respondents in the fall semester and the distribution of final marks for respondents in the winter semester: there is a higher probability of C and low-B marks in the winter semester than in the fall semester, and a higher probability of mid-to-high B marks in the fall semester than in the winter semester (Kolmogorov-Smirnoff test p = 0.0141). These differences may be due in part to the technological interventions that were applied in the fall semester but not used in the winter semester.

Below in Figure 6 we see a transformation of final marks that removes the heterogeneity. Recognizing that the study respondents from either semester are not representative of all HIS101 students, the transformation process enables the creation of two comparable subsets of HIS101 students that serve as a basis for further analysis and investigation

After centering each semester's marks at zero and scaling each to have standard deviation one as noted in Figure 6 below, there are no significant differences between the two marks distributions (Kolmogorov-Smirnoff test, p=0.4551).

Figure 6: HIS101 Empirical Distributions of Final Marks by Term and Respondent Status, Raw and Standardized



Non-Normal and Non-Homogeneous Sample

Therefore, as noted above, the data indicate that the final marks of students who participated in this exploratory study (Fall, Included and Winter, Included) are non-Gaussian within each semester and form a non-homogeneous sample across the two semesters.

Figures 4 and 5 demonstrate that the final marks are not normally distributed and that all four distributions are skewed to the left. The vertical axis on Figure 4 is the proportion of students who earned a final mark less than or equal to the marks on the horizontal axis. This figure shows that the two distributions for respondents ("included") are shifted to the right of (generally higher than) the two corresponding distributions for non-respondents ("excluded").

Technological Interventions and Semester Cannot be Distinguished

Minimization of confounded effects is accomplished by randomly assigning each student in a single class to one of two groups: one group to receive the intervention and the other group not to receive the intervention. Thus, prior to the intervention, there should be no systematic differences in the composition of the two groups.

Initial Analysis: Linear Models

Prior to the standardization process that is described above, numerous other approaches were implemented and reviewed; and in the end were not used, as they were not suitable. An overview is provided below of the different approaches that were explored.

Two Independent Sample t-Test

A simple assessment of our research hypothesis is a t-test for equality of the two semesters' mean differences between pre-intervention test and post-intervention test marks (see Figure 7 below, which is also duplicated in the main body of the report as Figure 1). Our data provide no evidence in support of the fundamental research hypothesis (p = 0.667).

Figure 7: Post-Intervention Test Mark minus Pre-Intervention Test Mark, Split by Semester (Fall semester with technological interventions and Winter semester without technological interventions)

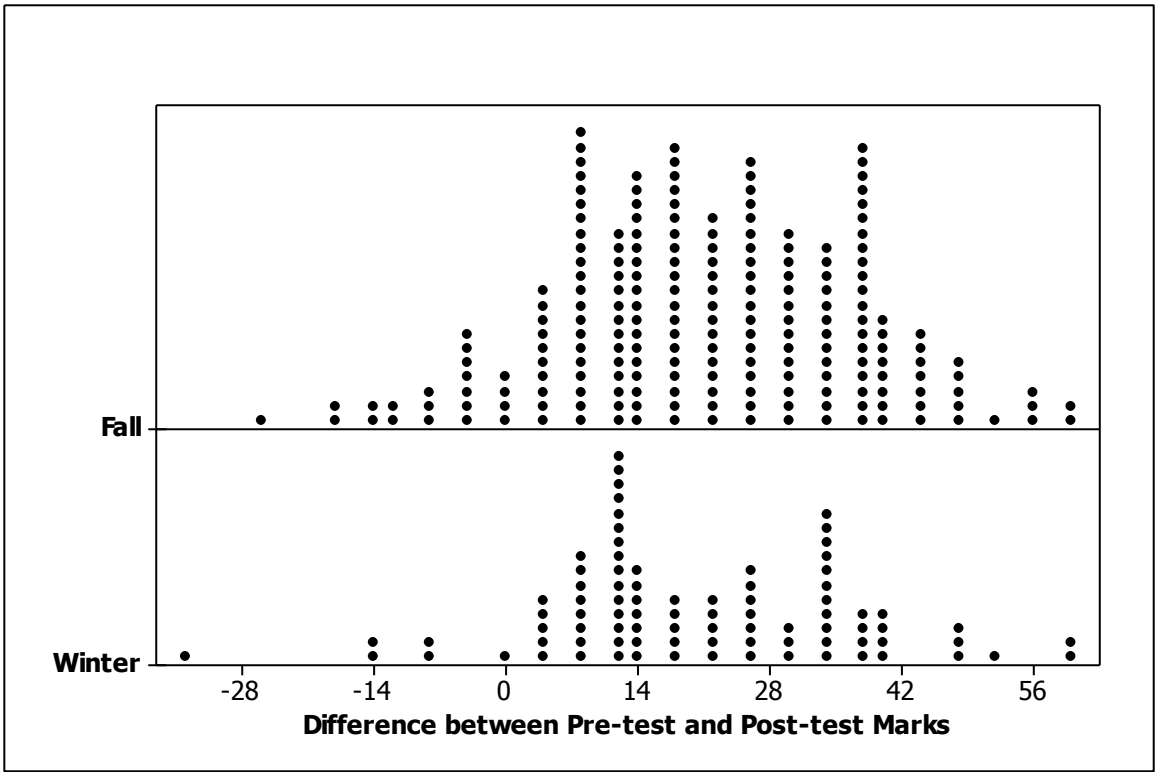


Figure 7 supports the conclusion of the t-test; there is no significant difference between the locations of the two distributions. However, Figure 7 also shows differences between the shapes of the two distributions: the winter semester looks bimodal whereas the fall semester looks to have high dispersion. This suggests that

additional variables may need to be investigated to improve focus on the differences between the two semesters. We also notice a potential outlier in the winter semester.

We chose not to proceed with this analysis, although it is statistically valid:

- We have graphical suggestions that a more complex model is appropriate.
- It does not lead to rejection of the null hypothesis in favour of the research hypothesis. We have no evidence that students exposed to our technological interventions exhibit greater improvement in marks on average than those not exposed to the technological interventions.

Linear Regression Models

More textured than the t-test are linear regression models.

One linear model that may appear promising is:

$Post.test = \beta_o + \beta_1 Intervention + \beta_2 Pre.test + \beta_3 Intervention \times Pre.test + error$, where

- *Pre.test* and *Post.test* are a student's numeric marks.
- *Intervention* indicates whether the student received the technological interventions or not.
- *error* represents everything, other than pre-intervention test mark and intervention status, that influences the student's post-intervention test mark.
- β_0, β_1, β_2, and β_3 weight the influence of the pre-intervention test mark and the intervention status (either they had the intervention or they did not) on the post-intervention test mark. These weights are constant for students.

Using our data we estimated this model and obtained these results:

- In the class that received the technological interventions – a post-intervention test mark is predicted as 32.58 + 0.6072*Pre.test*; for any pair of students whose pre-intervention test marks differ by 1%, we expect a difference of about 0.6072% in post-intervention test mark.
- In the class that did not receive the technological interventions – a post-intervention test mark is predicted as 33.878 + 0.6173*Pre.test*; for any pair of students whose pre-intervention test marks differ by 1%, we expect a difference of about 0.673% in post-intervention test mark.
- The difference between the additive constants is not statistically significant (p = 0.809).
- The difference between the weights of pre-intervention test marks is not statistically significant (p = 0.945).
- This model accounts for 22.7% of the differences (total variation) between the post-intervention test marks. Determinants of the remaining 77.3% of the variation must be attributed to other, as yet unspecified, variables.

Figure 8: Linear Regression of Pre-Intervention Test Mark and Intervention Status on Post-Intervention Test Mark

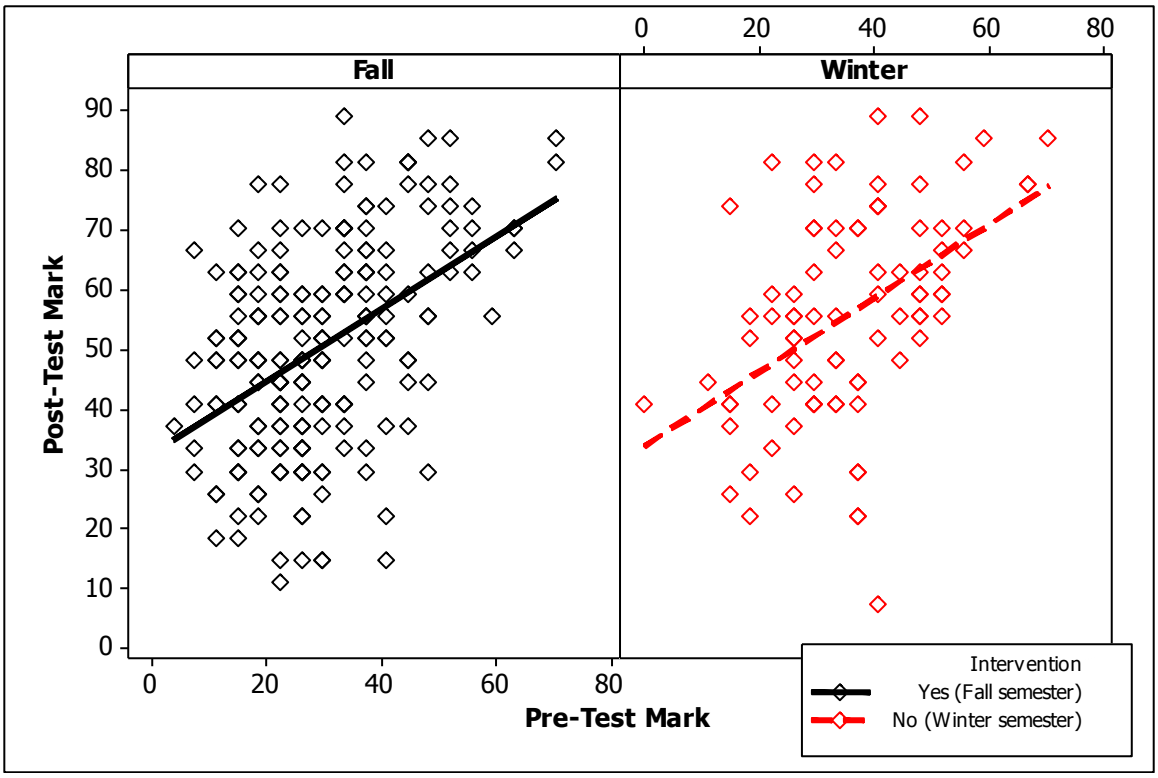


Figure 8 shows that the relationship between pre-intervention test marks and post-intervention test marks is not a straight line. In each semester the relationship appears concave (down), possibly quadratic or logarithmic.

Figure 8 shows that the percentage of pre-intervention test marks that are low is smaller in the Winter semester than in the Fall semester. This implies that, to some degree, an indicator for our technological interventions is equivalent to an indicator for low pre-intervention test marks. The mechanics of regression model estimation are not particularly robust to the inclusion of highly correlated, or (somewhat) equivalent, predictors. The result is poor statistical power.

In addition to this model, we fit many other models. Details pertaining to some of these other models can be found below.

We chose not to proceed with this analysis:

- None of the models we fit led to rejection of the fundamental null hypothesis in favour of the research hypothesis. We have no evidence that students exposed to the technological interventions exhibit greater improvement in marks on average than those not exposed to the technological interventions.
- We have graphical suggestions that a more complex, likely concave, model is appropriate.
- We have concerns about statistical power. These concerns are rooted in the fact that pre-intervention test scores are generally higher in the winter semester than in the fall semester.

- The models of this section are all predictors of post-intervention test mark and so investigation of student improvement (learning) must be conducted indirectly by conditioning on pre-pest mark. This approach is necessary to accommodate censored data that arise with perfect post-intervention test marks. However, we have no censored data since no one earned 100% on the post-intervention test. This is one of the reasons we chose to model improvement directly.

Formally the t-test model is $Post.test = \beta_o + \beta_1 Intervention + Pre.test + error$, where *Intervention* indicates the presence or absence of the technological interventions used in our exploratory quasi-experiment.

Two potential linear models are:

$$Post.test = \beta_o + \beta_1 Intervention + \beta_2 Pre.test + error$$

Note that this is an extension of the t-test model in which a weighted difference between pre-intervention test and post-intervention test marks is introduced.

$$Post.test = \beta_o + \beta_1 Intervention + \beta_2 Pre.test + \beta_3 Intervention \times Pre.test + error$$

Note that this is an extension of the previous model in which the weighted difference between pre-intervention test and post-intervention test marks is allowed to differ in the two semesters.

Table 10: Coefficients for Six Different Regressions on Post-Intervention Test Mark

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	33.878**	32.129**	42.429**	41.400**	43.820**	43.211**
pre.test	0.617**	0.610**	0.078	0.114		
Intervention	-1.298	-1.640	-2.948	-1.737	-2.637	-1.805
pre.test*intervention	-0.010		0.036			
$pre.test^2$			0.007*	0.007*	0.008**	0.009**
$pre.test^2$*intervention					0.001	
R^2-adjusted	22.50%	22.70%	23.20%	23.40%	23.40%	23.60%
** indicates significance at α = 0.05, and* indicates significance at □ = 0.10.						

Table 10 shows that the relationship between pre-intervention test and post-intervention test marks is more quadratic than it is a straight line. Note that intervention is not significant in any of these models.

Since pre-intervention test mark and intervention status are equivalent to some extent, putting both predictors in the same model will likely result in unstable parameter estimates. This instability of model estimates decreases statistical power, making it harder to correctly detect significant effects. The following example demonstrates the model instability. We changed one student's pre-intervention test mark from 59% to 29%. The student was in the non-intervention group; the student was not an outlier (or close to one) under either mark scenario. But the estimated model changed from $\widehat{Post.test}$ = 32.58 + 0.6072 Pre.test (with interventions) and $\widehat{Post.test}$ =33.878 + 0.6173 Pre.test (without interventions) to $\widehat{Post.test}$ = 32.6 + 0.6072 Pre.test (with interventions) and $\widehat{Post.test}$ =35.44 + 0.5801 Pre.test (without interventions). This change is not statistically significant, and may not be practically significant, but it is indicative of the magnitude of the problem.

Measuring Student Learning

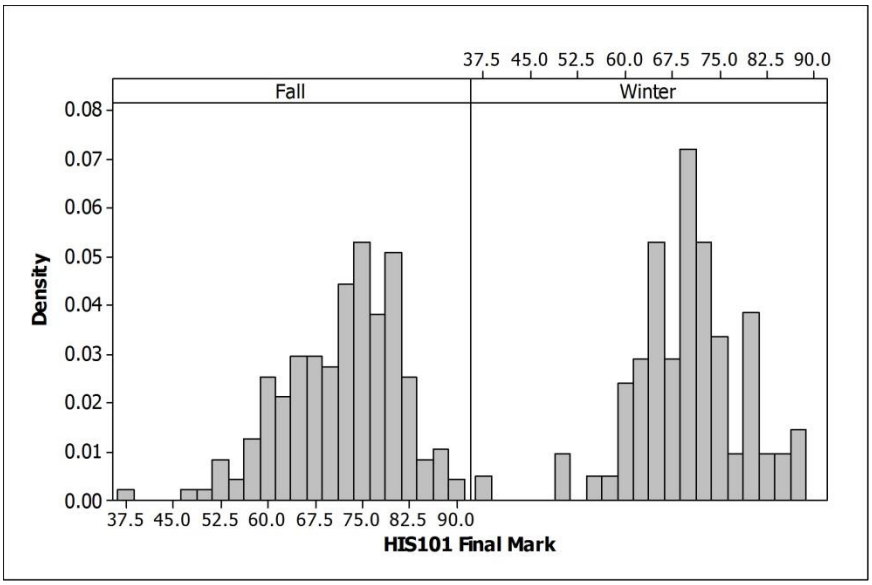
Table 11: Improvement Categories and Associated Partitions

	Range of (Post-Intervention Test Mark)% minus (Pre-Intervention Test Mark)% Included in Each *Improvement* Group	
***Improvement* Group**	**From**	**To**
-2 (None)	-33.33%	7.41%
-1 (Slight)	7.41%	14.81%
0 (Some)	14.81%	25.93%
1 (Good)	25.93%	37.04%
2 (Very Good)	37.04%	59.26%

Predictors of Student Learning

As noted above, there was a statistically significant difference in the distributions of final marks for the two semesters: looking only at higher marks, the winter semester is shifted to the right of the fall semester (Kolmogorov-Smirnoff test $p = 0.01407$; see Figure 9). There are many reasons why these mark distributions could differ, including but not limited to the following: the intervention and its contingent adjusted weightings in course marking schemes; a completed semester of university for the students in first year; timetabling issues forcing stronger students into the winter semester. To remove the effect of any of these hidden, possibly influential variables, final marks were converted into standardized scores within each semester. There were no significant differences between the distributions (and means) of final marks, after conversion to standardized scores, for the intervention group in the fall semester and the non-intervention group in the winter semester (Kolmogorov-Smirnoff test $p = 0.3446$).

Figure 9: Probability Density Distribution of Final Mark in HIS101 by Term



After being converted to standardized scores, the final marks were partitioned into five groups defined by their categories (see Table 12 below and Table 4). One set of categories was calculated using all $n = 297$ students from both semesters of this exploratory study. This categorization produced five groups, labeled -2 (Lowest), -1 (Low), 0 (Middle), 1 (High), and 2 (Highest). We have called this variable Mark; it is an ordinal measure of performance in HIS101.

Table 12: Final Mark Quintiles and Associated Partitions

	Range of *Mark* Group Fall Semester (Intervention) (N=211)		**Range of *Mark* Group Winter Semester (No Intervention) (N=86)**	
***Mark* Group**	**From**	**To**	**From**	**To**
-2 (Lowest)	7.5%	63.6%	1.0%	60.9%
-1 (Low)	63.6%	69.8%	60.9%	67.5%
0 (Middle)	69.8%	74.4%	67.5%	72.5%
1 (High)	74.4%	79.3%	72.5%	77.7%
2 (Highest)	79.3%	90.1%	77.7%	89.2%

There are very few students in some combinations of Improvement and Intervention, particularly in the non-Intervention (winter) semester. Table 13 shows the number and percentage of students falling into each cell in a Mark x Improvement matrix for each semester. When this is done, it becomes clear that there are unequal proportions of students across the cells, particularly in the non-intervention (winter) semester. This data sparseness limits the complexity of our modeling ability and indicates that we need to exercise caution when making conclusions about a fitted model. However, our total sample size is sufficient for the model we estimate (Hsieh, Bloch & Larsen, 1998).

Table 13: Mark by Improvement Matrices for Fall and Winter Semesters

Intervention	HIS101 *Mark* Categories	Number and Percent of Observations in Each Improvement Category					
		-2 (None)	-1 (Slight)	0 (Some)	1 (Good)	2 (Very Good)	Total
No (Winter semester)	-2 (Lowest)	1 8.33	4 33.33	3 25	4 33.33	0 0	12 100%
	-1 (Low)	9 39.13	9 39.13	2 8.7	2 8.7	1 4.35	23 100%
	0 (Middle)	4 21.05	4 21.05	5 26.32	4 21.05	2 10.53	19 100%
	1 (High)	5 29.41	4 23.53	3 17.65	2 11.76	3 17.65	17 100%
	2 (Highest)	0 0	1 6.67	4 26.67	6 40	4 26.67	15 100%
Yes (Fall semester)	-2 (Lowest)	16 34.04	9 19.15	14 29.79	5 10.64	3 6.38	47 100%
	-1 (Low)	15 40.54	5 13.51	8 21.62	6 16.22	3 8.11	37 100%
	0 (Middle)	6 15	9 22.5	12 30	7 17.5	6 15	40 100%
	1 (High)	9 20.93	3 6.98	10 23.26	17 39.53	4 9.3	43 100%
	2 (Very High)	6 13.64	6 13.64	10 22.73	12 27.27	10 22.73	44 100%

The reference category for this model (see Table 14) is the application of no interventions (our winter semester) to students whose final mark in HIS101 is in the median final mark group. We observe that Mark has a strong effect on the probabilities associated with each Improvement group when this effect is measured across both intervention and non-intervention and the five ordinal levels of Improvement ($p = 0.002$). We also observe that having or not having the Intervention has no significant effect on the probabilities associated with each Improvement group when this effect is measured across the five Mark categories and the five ordinal levels of Improvement ($p = 0.686$). Finally, we observe that the interaction of Intervention and Mark does not have a significant effect on the probabilities associated with each Improvement group when this effect is measured over the five ordinal levels of Improvement ($p = 0.223$). However, this does not imply that there are no significant differences between all pairs of estimated probabilities.

Table 14: Ordinal Logistic Regression Model Summary: Effects of Intervention and HIS101 Final Mark

Predictor	Odds Ratio	Coefficient	Standard Error	z	p-value
Improvement Group -1		-1.258	0.419	-3.000	0.003
Improvement Group 0		-0.342	0.412	-0.830	0.406
Improvement Group +1		0.749	0.414	1.810	0.070
Improvement Group +2		2.160	0.437	4.940	0.000
Intervention (Fall Semester)	0.82	-0.200	0.495	-0.400	0.686
HIS101 Final Mark					0.002
Mark Group -2	0.92	-0.079	0.654	-0.120	0.903
Mark Group -1	2.95	1.083	0.562	1.930	0.054
Mark Group +1	1.29	0.252	0.593	0.430	0.671
Mark Group +2	0.24	-1.410	0.624	-2.260	0.024
HIS101 Final Mark within Intervention (Fall Semester)					0.223
Mark Group -2	2.39	0.871	0.760	1.150	0.252
Mark Group -1	0.77	-0.258	0.692	-0.370	0.709
Mark Group +1	0.60	-0.516	0.711	-0.730	0.468
Mark Group +2	2.34	0.851	0.732	1.160	0.245

Figure 10: Probability Distributions of Improvement Category by Term and Final Mark

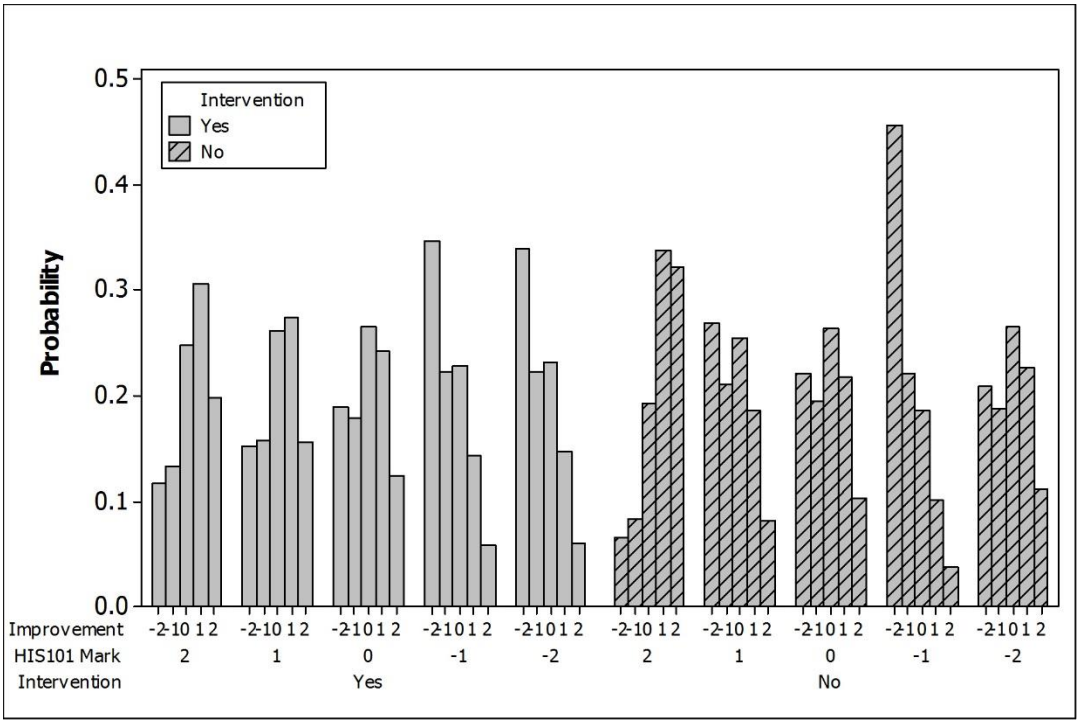


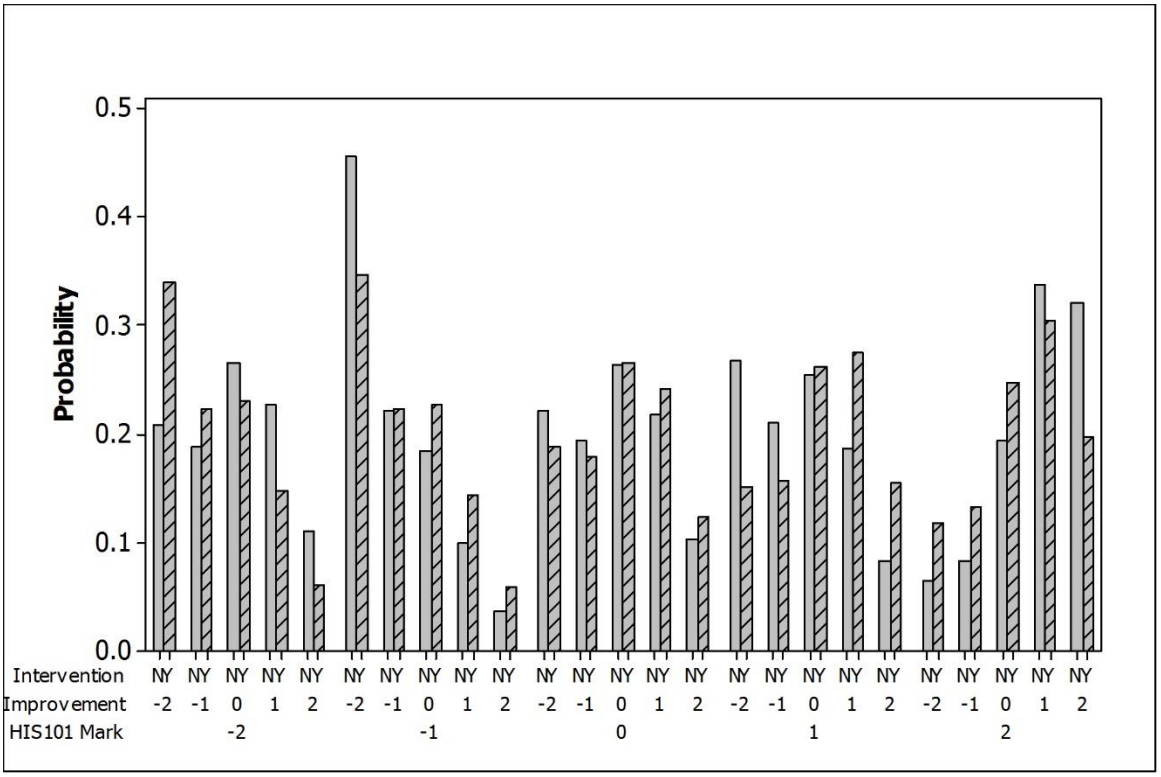
Figure 10 (above) and Table 15 (below) show the visual and the tabular representations of the probability distributions of Improvement category by term and final mark. The differences are made more apparent in Figure 11 below where the final mark categories are separated out by semester.

Table 15: Probability Distributions of Improvement Category by Term and Final Mark

Intervention	HIS101 Mark Group	Improvement Probability Distributions				
		-2 (None)	-1 (Slight)	0 (Some)	1 (Good)	2 (Very Good)
Yes (Fall)	2 (Highest)	0.117	0.132	0.248	0.305	0.198
No (Winter)	2 (Highest)	0.065	0.083	0.193	0.338	0.321
Yes (Fall)	1 (High)	0.152	0.157	0.262	0.274	0.155
No (Winter)	1 (High)	0.268	0.210	0.254	0.186	0.082

Intervention	HIS101 Mark Group	Improvement Probability Distributions				
		-2 (None)	-1 (Slight)	0 (Some)	1 (Good)	2 (Very Good)
Yes (Fall)	0 (Middle)	0.189	0.179	0.266	0.242	0.123
No (Winter)	0 (Middle)	0.221	0.194	0.264	0.218	0.103
Yes (Fall)	-1 (Low)	0.347	0.223	0.228	0.144	0.058
No (Winter)	-1 (Low)	0.456	0.221	0.185	0.100	0.038
Yes (Fall)	-2 (Lowest)	0.339	0.223	0.231	0.147	0.060
No (Winter)	-2 (Lowest)	0.208	0.188	0.265	0.227	0.111

Figure 11: Improvement Categories by Final Mark Categories, Paired by Semester



A non-statistical pairwise comparison of Improvement probabilities in each semester, within each combination of its five levels together with the five Mark categories, yields some interesting observations (see Figure 11). Students in Mark categories -1 (Low), 0 (Middle), or 1 (High) had a higher probability of above average

Improvement (2 or 1) in the fall than in the winter (indicated in Figure 3 as A* and A**). But for students in Mark categories -2 (Lowest) and 2 (Highest), the probability of above average Improvement (2 or 1) is higher in the Winter than in the Fall (indicated in Figure 3 as B* and B**). For students in Mark categories 0 (Middle) and 1 (High), the probability of below average improvement (-2 (None) or -1 (Slight)) is higher in the winter than in the fall for the middle and slightly above middle mark categories (0 (Middle) and 1 (High)). But the probability of below average improvement (-2 or -1) is higher in the fall than in the winter for the extreme Mark categories -2 (Lowest) and 2 (Highest).

Table 16: Ordinal Logistic Regression Model Summary: Effect of First-Year Status

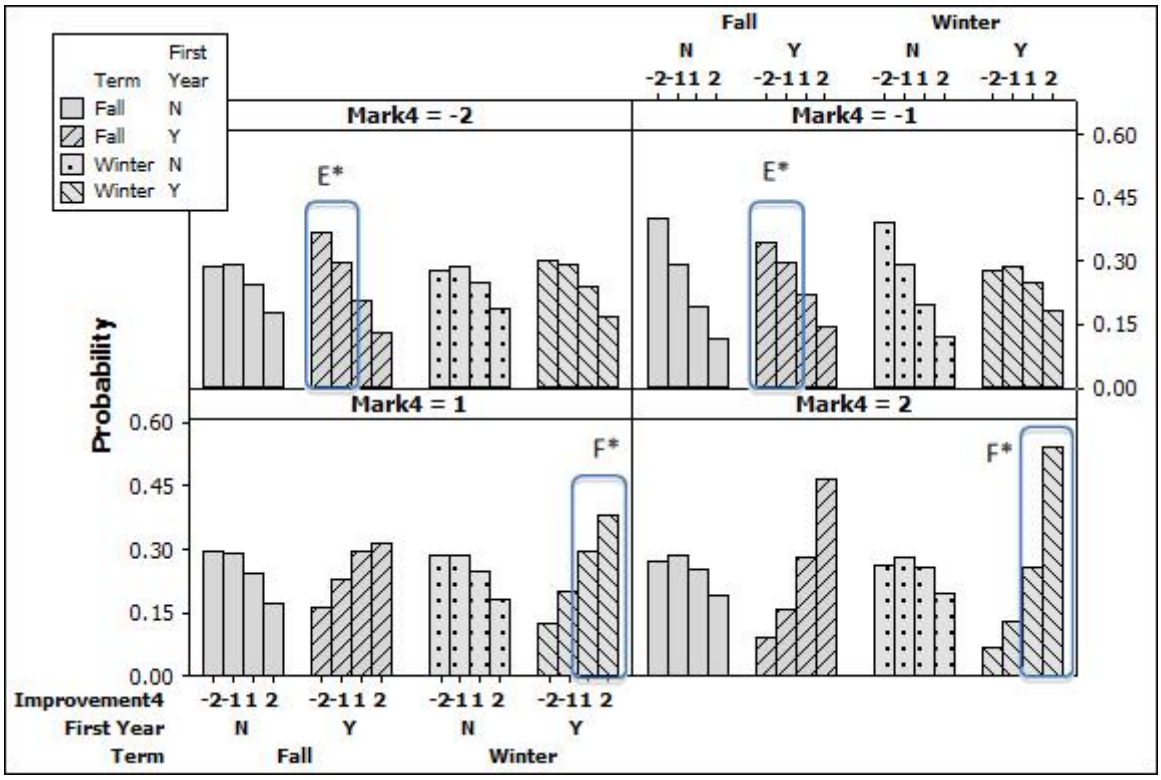
Term	First Year	Mark4	Improvement4 Probability Distributions			
			-2 (None)	-1 (Small)	1 (Intermediate)	2 (Large)
Fall (Intervention)	No (Not First Year)	-2 (Poor)	0.288	0.289	0.246	0.177
		-1 (Fair)	0.403	0.292	0.191	0.114
		1 (Good)	0.295	0.290	0.242	0.172
		2 (Very Good)	0.271	0.285	0.254	0.190
	Yes (First Year)	-2 (Poor)	0.368	0.295	0.207	0.130
		-1 (Fair)	0.344	0.295	0.219	0.143
		1 (Good)	0.160	0.231	0.295	0.313
		2 (Very Good)	0.089	0.159	0.281	0.470
Winter (No Intervention)	No (Not First Year)	-2 (Poor)	0.278	0.287	0.250	0.184
		-1 (Fair)	0.391	0.293	0.197	0.120
		1 (Good)	0.285	0.288	0.247	0.179
		2 (Very Good)	0.261	0.283	0.259	0.198
	Yes (First Year)	-2 (Poor)	0.301	0.291	0.239	0.168
		-1 (Fair)	0.279	0.287	0.250	0.184

Term	First Year	Mark4	Improvement4 Probability Distributions			
			-2 (None)	-1 (Small)	1 (Intermediate)	2 (Large)
		1 (Good)	0.124	0.199	0.296	0.381
		2 (Very Good)	0.068	0.129	0.258	0.545

We observe that in Table 16 and Figure 12, for non-first-year students, within each of the four Mark4 categories, the distribution of improvement in the fall semester with the Intervention is virtually identical to the distribution of improvement in the winter semester without the Intervention. However, this is not the case for first-year students, for whom we observe:

- There is a higher probability of no improvement with the intervention than without the intervention among those students who earned poor to fair final marks.
- There is a higher probability of very good improvement without the intervention than with the intervention among those students who earned good to very good final marks.
- There is a higher probability of good improvement with the intervention than without the intervention among those students who earned very good final marks.

Figure 12: Ordinal Logistic Regression Model Summary: Effect of First-Year Status



Additional Data regarding Participation

Below in Table 17 additional data are provided regarding the participation of students in the online quizzes for the Fall 2012 semester.

Table 17: Participation in Online Quizzes during Fall 2012

	% of Students who Completed at least 1 Quiz	Median # of Quizzes Completed (out of 3)	Median # of Attempts per Quiz (averaged over all 3 quizzes)	Median Total Time Spent (averaged over initial quiz attempts) in Hours	Mean Total Performance on Questions relevant to this Study (averaged initial quiz attempts) as %
Total (n = 211)	96.2	3	2.3	2.8	60.6
True first year					
No (n = 51)	98.0	3	2.3	2.0	57.3
Yes (n = 160)	95.6	3	2.3	3.1	58.0
Gender					
Male (n = 77)	94.8	3	2.3	2.2	57.8
Female (n = 134)	97.0	3	2.4	2.9	57.9
HIS101 mark quintile					
2 (n = 44)	97.7	3	2.3	2.1	71.4
1 (n = 43)	100.0	3	2.7	2.1	66.0
0 (n = 40)	100.0	3	2.7	3.8	56.7
-1 (n = 37)	86.5	3	2.0	7.1	48.8
-2 (n = 47)	95.7	3	2.0	1.8	45.8
Improvement quintile					
2 (n = 26)	100.0	3	2.5	2.6	66.7
1 (n = 47)	93.6	3	2.7	2.0	60.0
0 (n = 54)	100.0	3	2.3	3.8	57.0
-1 (n = 32)	93.8	3	2.2	2.1	61.9
-2 (n = 52)	94.2	3	2.3	2.9	49.9

Student Responses to Course Evaluation Survey Questions

The large p-values in Tables 18, 19, and 20 indicate that there is no evidence of a relationship between improvement and attitude to classroom technology. Therefore, analysis of student responses on the course evaluation survey shows no correlation between performance in the course and attitude toward the intervention. In other words, for this exploratory study it seems that students' belief in the efficacy of classroom technology is unrelated to their academic success.

Table 18: Response to Course Evaluation Survey Question 102

Question 102 text: Course lectures improved my understanding of the course material. (Pearson's χ^2 p = 0.769)				
Improvement Group	HIS101 Mark Group	Positive (%)	Neutral or Negative (%)	Number of Students
Above Median	Above Median	73.1	26.9	26
	Median	66.7	33.3	3
	Below Median	80.0	20.0	5
Median	Above Median	61.5	38.5	13
	Median	50.0	50.0	4
	Below Median	40.0	60.0	10
Below Median	Above Median	57.1	42.9	14
	Median	50.0	50.0	4
	Below Median	56.3	43.8	16
	Total	61.1	38.9	95

Table 19: Response to Course Evaluation Survey Question 110

Question 110 text: The use of clickers in the classroom contributed to my learning of the course material. (Pearson's χ^2 p = 0.487)				
Improvement Group	HIS101 Mark Group	Positive (%)	Neutral or Negative (%)	Number of Students
Above Median	Above Median	73.1	26.9	26
	Median	66.7	33.3	3
	Below Median	100.0	0.0	5
Median	Above Median	84.6	15.4	13
	Median	50.0	50.0	4

Question 110 text: The use of clickers in the classroom contributed to my learning of the course material. (Pearson's χ^2 p = 0.487)				
Improvement Group	HIS101 Mark Group	Positive (%)	Neutral or Negative (%)	Number of Students
	Below Median	50.0	50.0	10
Below Median	Above Median	71.4	28.6	14
	Median	75.0	25.0	4
	Below Median	62.5	37.5	16
	Total	70.5	29.5	95

Table 20: Response to Course Evaluation Survey Question 111

Question 111 text: Educational technology (e.g. Portal system) contributed to my learning of the course material. (Pearson's χ^2 p = 0.655)				
Improvement Group	HIS101 Mark Group	Positive (%)	Neutral or Negative (%)	Number of Students
Above Median	Above Median	73.1	26.9	26
	Median	66.7	33.3	3
	Below Median	80.0	20.0	5
Median	Above Median	76.9	23.1	13
	Median	25.0	75.0	4
	Below Median	60.0	40.0	10
Below Median	Above Median	64.3	35.7	14
	Median	50.0	50.0	4
	Below Median	56.3	43.8	16
	Total	65.3	34.7	95

Higher Education Quality Council of Ontario

An agency of the Government of Ontario