



## Using Online Assessment to Replace Invigilated Assessment in Times of Natural Disaster: Are Some Online Assessment Conditions Better Than Others?

Steve Agnew, University of Canterbury

Stephen Hickson, University of Canterbury

### Abstract

As a result of the Canterbury earthquake on 4 September 2010, and associated aftershocks on 22 February 2011 and 13 June 2011, final examinations in the two first-year Economics papers at Canterbury University were cancelled at short notice in Semester 1, 2011. The final examination weightings were spread over the remaining assessments to obtain a final grade for students. This paper attempts to establish how different online assessment conditions affected final grade distributions when online assessments were substituted for an invigilated final examination. Pearson correlation coefficients and Spearman rank-order correlation coefficients were used to show that there was a greater correlation between online quizzes and invigilated assessments when those quizzes were only available for a restricted period of time compared with the whole semester. We found that online quizzes were more closely correlated with invigilated assessments when the first attempt at a quiz was recorded, as opposed to the higher of two attempts. We also found that using the first attempt leads to less grade disruption when compared with a 'normal' semester that includes a final examination. Finally, the actual effect on student grades when online quizzes are substituted for a final examination is discussed.

**Keywords:** Principles of economics; online assessment; student grades; disruption to assessment; earthquake; online assessment.

### Introduction

On 4 September 2010 an earthquake measuring 7.1 on the Richter scale struck Canterbury. As a result, the region was subjected to thousands of aftershocks, the most significant of which struck on 22 February 2011 and 13 June 2011, both measuring 6.3 on the Richter scale. The September earthquake and February aftershock each resulted in closures of 2 weeks for the University of Canterbury campus. The June aftershock struck just prior to the Semester 1 examination period, and resulted in the cancellation of final examinations for many papers at short notice.

The cancellation of final examinations at short notice raises questions about how best to calculate final grades. This paper will focus on assessment of the first-year economics papers (Introduction to Macroeconomics [ECON 105] and Introduction to Microeconomics [ECON 104]) at the University of Canterbury, and how well they allocate appropriate grades to students in an earthquake-affected semester. Specifically, this paper will examine the effect of different conditions under which students take online multiple-choice quizzes, such as limiting the time they are available for students to complete, or limiting the number of attempts a student has at a particular quiz. The first part of this

paper will examine the effect of limiting the availability of online quizzes on the correlation between those quizzes and invigilated assessments, such as term tests and final examinations. The second part of this paper will examine the effect of using the quiz “first attempt” rather than the “higher of two attempts” on correlations with invigilated assessments, and the impact on grade allocations. The correlations and grade distributions should provide an insight into how substitutable online quizzes and invigilated assessments are. In simple terms, the paper addresses this question: Can we gain some insight into the conditions under which online assessments provide the most robust grade allocations, compared with an uninterrupted semester, when a final exam or similar is cancelled?

## Literature review

There is a body of literature that focuses on the shift of teaching delivery to an online environment in the event of a crisis. Hickson and Agnew (2011), Benton (2009), Meyer and Wilson (2011), *eCampus News* (2009), Omar, Liu, and Koong (2008), Foster and Young (2005), Danielson (2009), and SchWeber (2008) all discuss courses moving to an online format in response to hurricanes, war, or viral outbreaks. However, there is a gap in the literature when it comes to the study of the effect of moving to online assessment in a semester disrupted at short notice by a natural disaster. Specifically, there is no examination of the effect on grade distributions of moving to online assessment.

In the general literature on online delivery methods, there is some discussion on the strengths and weaknesses of the online assessment of material. Graff (2003) stated, “There are many potential advantages of online assessment to learners. For example, tests are available on demand and at any time. Furthermore, computerised assessment systems give immediate feedback to the user; therefore users learn by taking the test”. (p. 22). Robles and Braathen (2002) find that “many different online components and assessment criteria and tools are needed to accurately and thoroughly assess student learning” (p. 47). They conclude by stating that “assessment should be ongoing and carried out through each chapter throughout the semester, to allow students to determine their own learning outcomes through self-testing” (p. 47). This is less costly to carry out in an online environment compared with running periodic physical tests. Running periodic online tests also provides a greater level of control for the instructor, and a greater incentive for student completion than leaving students to be responsible for their own independent self-testing. Gaytan and McEwen (2007) found that the frequency of assessment and the immediacy and meaningfulness of feedback were important. This concurs with earlier research from Corcoran, Dershimer, and Tichenor (2004), and Stiggins and Chappuis (2005), who stated that monitoring student learning and enhancing teaching were two main purposes of assessment. Gaytan and McEwen (2007) state that several researchers, such as Bartlett, Reynolds, and Alexander (2000), and Farmer (2005), have found that an online environment assessment fosters a student-centered learning environment and allows for more accurate measurement of learning. Gaytan and McEwen also reference Russell, Elton, Swinglehurst and Greenhalgh (2006), who stress the importance of continuous assessment, as it allows instructors to monitor and be familiar with students’ understanding.

Perrin and Mayhew (2000) raised some concerns about the validity of online testing based on the ability of students to cheat, giving the example of students printing online tests and sharing them. Robles and Braathen (2002) explain that this undesirable behaviour can be mitigated by restricting the number of times the student can sit an online quiz, how long they have to complete the quiz, and hiding the questions once the quiz has been submitted. They suggest distinguishing between online quizzes that are summative assessment items, and having formative online assessment items that have fewer restrictions on access and can be used as learning tools rather than assessment tools. Olt (2002) researched strategies for minimising academic dishonesty in online assessment. She cited a survey of American high school students that reported that 80% admitted cheating on an exam (Bushweller, 1999). Kleiner and Lord (1999), using the same survey data, found that 50% of the students did not believe cheating was necessarily wrong. They also found that 95% of those who said they cheated had never been caught. Heberling (2002) suggested that it may actually be easier to detect cheating online. However, Olt (2002) does suggest some disadvantages to online assessment. These include “an

instructor's inability to control a student's unauthorized use of resources in completing an assessment" (p. 3). She does go on to suggest using open-book assessment as a possible solution. Another disadvantage Olt identified is students collaborating with each other on an assessment. One possible solution she suggests is to have a question pool from which questions are randomly selected, so that the chances of two students receiving the same assessment are minimal. Collusion can also be reduced by having restrictions on assessment availability, and setting time limits. Rowe (2004) suggests it is often easier to cheat online, and asks the question, "When a student scores well for an online assessment, does that mean they know the material?" (p. 1). He cites Bork (2001) when stating that it is less cost-effective for students to cheat when assessment is continuous. If the concerns raised in the literature on student cheating in online assessments are legitimate, we must question the reliability of student grades because a greater weighting is applied to online assessments. Hickson and Agnew (2011) certainly found a greater level of grade disruption from a non-earthquake-affected semester when a greater weighting was placed on online quizzes, despite computer packages for online assessments becoming more sophisticated to limit the possibility of cheating.

The contribution this paper aims to make is not to debate the relative merits of online assessment. Rather, the paper aims to establish the ability of online assessment to deliver a grade distribution consistent with a 'normal' semester, in the event of the cancellation of a final exam at short notice.

## **Assessments, data and methods**

During the earthquake-interrupted semesters at the University of Canterbury, online assessments proved to be invaluable. They could be completed by students without the need to come onto campus (although those with no internet access at home could still use the computer labs on campus), and removed the need to use markers (who were themselves disrupted by the earthquake) to meet for moderation meetings, mark assignments and so on.

From 2005 to 2010 assessment in both ECON 104 and ECON 105 consisted of an invigilated 3-hour final examination (60%),<sup>1</sup> an invigilated 90-minute term test (20%), a take-home assignment (10%), and online quizzes (10%). Each online quiz consisted of 10 multiple-choice questions, drawn randomly from a test bank of hundreds of questions provided by the publisher of the textbook. There were 10 quizzes in total, each worth 1%, with the highest mark from an unlimited number of attempts being mark that counts. The questions were delivered with the quiz tool on Moodle. The term test and final exam included both multiple-choice and constructed response-type questions.

In ECON 105 in 2011, 10% was also introduced for some online tutorial work, with the weighting on the final examination being reduced from 60% to 50%. The effect of the earthquake disruption in Semester 1, 2011, was the cancellation of the final examination and take-home assignment, with the weighting redistributed across other assessments.

A crucial change was made to the online quizzes in ECON 105 for Semester 2, 2009, onwards. From 2005 to 2009 Semester 1, the quizzes were all available for the duration of the semester and could be completed multiple times. The highest mark was the counting mark for each quiz. The quiz availability was changed from Semester 2, 2009, when each of the 10 quizzes was open only for a short window around the time the topic was being covered, rather than for the whole semester. With the cancellation of assessments in Semester 1, 2011, a greater weighting was placed on the online quizzes when final grades were calculated. We are able to use this break in the online quiz conditions to examine how correlations between online quizzes and invigilated assessments are affected. There are 2440 observations for the pre-2009 Semester 2 period and 618 for 2009 Semester 2 onwards.

Assessment in ECON 104 also changed in 2011. In that year, Semester 2 assessment in ECON 104 consisted of an invigilated 3-hour final examination (55%), an invigilated 90-minute term test (25%),

---

<sup>1</sup> Before 2007 the term test was worth 35 percent and the final exam was worth 45 percent.

an online progress test (5%), online multiple-choice quizzes provided by the publisher (5%) and weekly online tutorial quizzes tailored to the course (10%). The 10 weekly online tutorial quizzes consisted of 15 multiple-choice questions drawn from a test bank of questions used in previous term tests and examinations. Each quiz was worth 1%, and was open for an average of 1 week. Students were allowed two attempts at each of the tutorial quizzes, with their highest score recorded. For the purpose of this paper, the tutorial quizzes were also re-marked, taking the students' first attempt rather than the highest mark of their two attempts. To establish the impact on the level of correlation between the invigilated term test and final examination, and the tutorial quizzes under the two marking scenarios, both the Pearson correlation coefficient and the Spearman rank-order correlation coefficient were calculated. Hypothetical final grades for Semester 2 ECON 104 2011 were also calculated under each tutorial quiz-marking scenario. In the scenario, the final examination is cancelled and the weighting it carries is redistributed across the term test and tutorial quizzes. The sample size for these correlations was 320.

## Results

In the four semesters since the ECON 105 quizzes went to limited windows of availability (Semester 2, 2009; Semesters 1 and 2, 2010; and Semester 2, 2011), the Pearson correlation coefficients for the online multiple-choice quizzes and final exam have been at the top, or exceeding the top, of the range of Pearson correlation coefficients for the period when the quizzes did not have limited windows of availability (Semester 1, 2005, to Semester 1, 2009). The Pearson correlation coefficients for the online multiple-choice quizzes and term test for the five semesters since the ECON 105 quizzes went to limited windows of availability (Semester 2, 2009; Semesters 1 and 2, 2010; and Semesters 1 and 2, 2011), have all exceeded the Pearson correlation coefficients for the period when the quizzes did not have limited windows of availability (Semester 1, 2005, to Semester 1, 2009). The Spearman rank-order correlation coefficients for the online quizzes and both the exam and the term test for the period Semester 2, 2009, to Semester 2, 2011, have all exceeded the Spearman rank-order correlation coefficients for the period Semester 1, 2005, to Semester 2, 2009, when the quizzes did not have limited windows of availability. This is shown in Table 1 below (for full results see Appendix A). All coefficients are significant at the 1% level of significance.

**Table 1** Range of correlation coefficients for assessment items (Semester 1, 2005, to Semester 1, 2009)

Pearson correlation coefficients

	Term test	Exam
<b>2005-S1 to 2009-S1 ECON 105 MC Quiz</b>	<b>0.27-0.47</b>	<b>0.48-0.66</b>
2009 S2 ECON 105 MC Quiz	0.55	0.65
2010 S1 ECON 105 MC Quiz	0.55	0.66
2010 S2 ECON 105 MC Quiz	0.49	0.67
2011 S1 ECON 105 MC Quiz	0.57	N/A
2011 S2 ECON 105 MC Quiz	0.66	0.77

Spearman rank-order correlation coefficients

	Term test	Exam
<b>2005-S1 to 2009-S1 ECON 105 MC Quiz</b>	<b>0.24-0.41</b>	<b>0.35-0.50</b>
2009 S2 ECON 105 MC Quiz	0.51	0.52
2010 S1 ECON 105 MC Quiz	0.55	0.61
2010 S2 ECON 105 MC Quiz	0.45	0.55
2011 S1 ECON 105 MC Quiz	0.46	N/A
2011 S2 ECON 105 MC Quiz	0.63	0.74

Table 2 shows that for the period Semester 2, 2009, to Semester 2, 2011 (excluding Semester 1, 2011, when an examination was not held), both the Pearson correlation coefficients and the Spearman rank-order coefficients for the term test and the examination have shown no discernible change between the semesters with unlimited windows of availability for the quizzes, and the semesters of limited windows of availability. This suggests that the standard of invigilated assessment is staying relatively constant, as it is unlikely both would change by the same amount. We can therefore be confident that the improved correlation coefficients for the multiple-choice quizzes is due to their becoming more closely correlated to the invigilated assessments, rather than the invigilated assessments becoming more closely correlated to the multiple-choice quizzes. Note that the both the Pearson correlation coefficients and the Spearman rank-order coefficients are higher for the term test and the examination, compared with the quizzes and the term test, and the quizzes and the examination. This is an intuitively obvious result, given that the invigilated term test and exam are more closely correlated to each other than to the non-invigilated multiple-choice quizzes.

**Table 2** Range of correlation coefficients for assessment items (Semester 2, 2009, to Semester 2, 2011)

Pearson correlation coefficients

	Exam
<b>2005-S1 to 2009-S1 ECON 105 Term test</b>	<b>0.73-0.81</b>
2009 S2 ECON 105 Term test	0.79
2010 S1 ECON 105 Term test	0.80
2010 S2 ECON 105 Term test	0.69
2011 S1 ECON 105 Term test	N/A
2011 S2 ECON 105 Term test	0.83

Spearman rank-order correlation coefficients

	Exam
<b>2005-S1 to 2009-S1 ECON 105 Term test</b>	<b>0.76-0.83</b>
2009 S2 ECON 105 Term Test	0.80
2010 S1 ECON 105 Term Test	0.82
2010 S2 ECON 105 Term Test	0.73
2011 S1 ECON 105 Term Test	N/A
2011 S2 ECON 105 Term Test	0.84

These results suggest that in the event of assessment disruption the quizzes that have limited windows of availability are more closely correlated to, and thus potentially better predictors of, both the term test and the examination, and therefore produce more robust grades.

To examine the impact of different online assessment conditions on student grades, ECON 104 data from Semester 2, 2011, is used. As mentioned in the methods section, students had two attempts at the online tutorial quizzes, with their highest mark recorded. Their responses were then re-marked using their first attempt only. Table 3 below shows the Pearson correlation coefficients and Spearman rank-order correlation coefficients for the online tutorial quizzes correlated against the term test and exam, for both the highest scoring attempt and the first attempt.

**Table 3** ECON 104 (Microeconomics) (Semester 2, 2011)

Online tutorial quiz—first attempt recorded

	Tutorial quiz	Term test	Final exam
		<b>Pearson correlation coefficients</b>	
Tutorial quiz	1.00	0.65	0.68
Term test	0.62	1.00	0.83
Final exam	0.65	0.85	1.00
	<b>Spearman rank-order correlation coefficients</b>		

Online tutorial quiz—highest mark of two attempts recorded

	Tutorial quiz	Term test	Final exam
		<b>Pearson correlation coefficients</b>	
Tutorial quiz	1.00	0.63	0.64
Term test	0.59	1.00	0.83
Final exam	0.60	0.85	1.00
	<b>Spearman rank-order correlation coefficients</b>		

In ECON 104 Semester 2, 2011, the Pearson correlation coefficients for the online tutorial quizzes compared with both the term test and examination were slightly higher when the first of two attempts was recorded as the student's quiz mark, compared with when the highest mark of their two attempts were recorded. The Spearman rank-order correlation coefficients also found a higher correlation on the first attempt. Table 4 shows the impact on student grades of using the highest scoring attempt compared with the first attempt. The shaded cells represent the number of students who received the same grade as they would have in an uninterrupted semester that included a final exam.

**Table 4** Number of students—using raw scores, highest tutorial quiz mark  
Weighting (term test 80%)

		Alternative grade										
		A+	A	A-	B+	B	B-	C+	C	C-	D	E
True grade	A+	39	1									
	A	16	5	3								
	A-	9	11	7	1							
	B+		11	7	3	1						
	B	1	1	8	7	9	3					
	B-		2	5	2	8	3	3				
	C+		1	2	4	5	8	6	1	1		
	C				2	2	3	7	10	2	1	
	C-					1	3	8	2	6	2	
	D						3	3	1	5	2	4
E					2	1	2	2	9	4	39	

Table 5 shows the grade distribution if the final exam had been cancelled at short notice and the 55% weighting from the final exam was placed onto the term test. The calculation of the students' grades under both scenarios includes 10% on the online tutorial quizzes, using the higher of two attempts as the student mark. Of the 320 students in the course, 70.3% received a grade within  $\pm 1$  GPA<sup>2</sup> number of their grade including a final exam, and 88.4% received a grade within  $\pm 2$  GPA numbers. These percentages reflect the high correlation between the term test and the examination, both of which were invigilated.

Appendix B shows the grade distribution if the exam hadn't run, and the 55% weighting was put on to the online tutorial quizzes rather than the term test, using the students' highest scoring of two attempts. Using this approach, only 47.2% of students received a grade within  $\pm 1$  GPA number of their grade if a final exam was included, with 65.3% receiving a grade within  $\pm 2$  GPA numbers.

<sup>2</sup> Grade point average (GPA) is awarded as follows: A+ = 9, A = 8 and so on down to E = 1.

**Table 5** Number of students—using raw scores, first tutorial quiz mark  
Weighting (term test 80%)

		Alternative grade										
		A+	A	A-	B+	B	B-	C+	C	C-	D	E
True grade	A+	32	1									
	A	17	6	3								
	A-	9	9	4	1							
	B+	2	11	5	8	1						
	B		1	6	7	5	2					
	B-		2	3	6	9	4	3	1			
	C+		2	2	3	7	5	6	1			
	C				3	3	6	9	5	2	1	1
	C-						21	5	6	6	2	
	D						4	1	4	5	3	3
E					1	2	3	2	7	10	40	

Table 5 shows the grade distribution if the exam hadn't run and the 55% weighting was put onto the term test, with 10% weighting still on the online tutorial quizzes, using the students' first attempt at the tutorial quizzes as their mark. This approach yielded similar results to those in Table 4, with 68.8% of students receiving a grade within +/- 1 GPA number of their grade if a final exam were included, and 88.4% receiving a grade within +/- 2 GPA numbers.

Appendix B shows the grade distribution if the exam hadn't run, and the 55% weighting was put onto the online tutorial quizzes rather than the term test, using the students' first of two attempts. Using this approach, 63.4% of students receive a grade within +/- 1 GPA number of their grade if a final exam was included, with 79.1% receiving a grade within +/- 2 GPA numbers. This is an improvement on when the higher of two attempts was used, where the corresponding percentages were 47.2% and 65.3% respectively. Table 6 shows that in the absence of a final exam, as more weighting is put on to online tutorial quizzes, there is more grade disruption when the highest scoring of two tutorial quiz attempts is used relative to when the first of two tutorial quiz attempts is used. This confirms the earlier results in Table 3, which showed that online tutorial quiz marks are more closely correlated to both the term test and final exam, when the first of two attempts is used rather than the highest scoring of two attempts. Table 6 summarises the level of grade inflation under each of the different assessment scenarios.

**Table 6** Number of students who received a grade higher, lower, or the same as their grade when an examination runs

Highest scoring attempt tutorial quiz mark	Higher	Same	Lower
80% Term test	52.5%	40.3%	7.2%
Tutorial quizzes 65%	62.5%	29.7%	7.8%
First attempt tutorial quiz mark			
80% Term test	55.9%	37.2%	6.9%
Tutorial quizzes 65%	39.1%	35.3%	25.6%



All approaches result in grade inflation, which reflects the fact that the mean for the final exam in ECON 104 Semester 2, 2011, was lower than all of the other assessments. It is clear from the table that using the first of two attempts on the online tutorial quizzes results in less grade inflation than using the highest scoring of two attempts, as more weighting is put onto the online tutorial quizzes.

## **Conclusions**

The data from ECON 105 suggests that online quizzes will provide a greater correlation to invigilated assessments if they are available for a shorter time period of around a week rather than a whole semester. The result will be less grade disruption if weightings are increased on online quizzes in the event of a cancelled examination or term test.

In ECON 104, online quizzes that record a student's first attempt rather than their highest scoring of two attempts also show a closer correlation to invigilated assessments, and result in less grade disruption when weightings on online quizzes are increased. Had the entire examination weighting of 55% been placed onto the online quizzes using the highest scoring of two attempts, only 47% of students would have received a grade within plus or minus one GPA point of the grade they would receive with a final examination. This compares with a corresponding figure of 63% for the first attempt when the quizzes are marked using the first attempt. The percentages of students that get the same grade as they would have in a normal semester with all assessment completed are 29.7% and 37.2% respectively. It is important to note that this isn't just a case of students getting higher than usual grades if online quizzes are substituted for invigilated assessments. The Spearman coefficient measures the rank order of students under different assessment regimes. In ECON 104 Semester 2, 2011, the final examination was harder than the other assessments, so grades were inflated when the alternative assessments were used as substitutes for the final examination. This may not be the case in every semester, however. In ECON 104, the term test often has a lower mean than the final examination.

Online quizzes that were marked using the first attempt yielded a better correlation to invigilated assessments, and therefore less grade disruption, when compared with a normal semester of assessment. Interestingly, if the weighting from a cancelled final exam had been put solely onto the online quizzes, 25.6% of students would have received a lower grade than in a normal semester, compared with only 7.8% if the highest scoring of two quiz attempts had been recorded. Using the first attempt results in less disruption to grades, and maintains the rank ordering of students more effectively, but any grade disruption that does exist is more likely to result in a lower grade for some students compared with using the highest scoring of two attempts.

We cannot state the reasons for the differing correlations when the window of availability or the number of attempts is varied. However, as mentioned in the introduction, Robles and Braathen (2002), Olt (2002), and Bork (2001) suggest that the undesirable behaviour of cheating can be mitigated by restricting the number of times the student can sit an online quiz, and how long they have to complete the quiz.

The results presented in this paper raise questions about how more traditional methods of assessment such as invigilated examinations could be replaced by the use of online assessment. We believe that this is an area worthy of future study. However, the goal of this paper is not to suggest that one type of assessment is 'better' than another. It does find that online assessments are more substitutable for invigilated assessments in the event of invigilated assessments having to be cancelled at short notice, if certain restrictions are placed on them such as period of availability and number of attempts. This should inform the decision making of course instructors when assessment is being designed for a course, especially if there is an increased risk of some future disruption to assessment items at short notice.

## References

- Bartlett, J., Reynolds, K., & Alexander, M. (2000). A tool for online learning. *Journal of Online Learning, 11*(3–4), 22–24.
- Benton, T. H. (2009, November 30). Teaching in the plague year. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Teaching-in-the-Plague-Year/49275>
- Bork, A. (2001). What is needed for effective teaching on the internet? *Educational Technology and Society, 4*(3), 139–143.
- Bushweller, K. (1999). Generation of cheaters. *The American School Board Journal, 186*(4), 24–32.
- Corcoran, C., Dershimer, E., & Tichenor, M. (2004). A teacher's guide to alternative assessment: Taking the first steps. *The Clearing House, 77*(5), 213–216.
- Danielson, R. (2009, August 25). If widespread swine flu closes USF campus, classes will go on. *Tampa Bay Times*, A1.
- eCampus News*. (2009, September 15). Web classes could help in H1N1 outbreaks. *LBCC eLearning*. Retrieved from <http://lbcclearning.blogspot.co.nz/2009/09/web-classes-could-help-in-h1n1.html>
- Farmer, L. (2005). Using technology to facilitate assessment of library education. *Teacher Librarian, 32*(3), 12–15.
- Foster, A. L., & Young, J. R. (2005, September 16). The internet as emergency tool. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/The-Internet-as-Emergency-Tool/15881>
- Gaytan, J., & McEwen, B. (2007). Effective online instructional and assessment strategies. *The American Journal of Distance Education, 21*(3), 117–132.
- Graff, M. (2003). Cognitive style and attitudes towards using online learning and assessment methods. *Electronic Journal of e-Learning, 1*(1) 21–28.
- Heberling, M. (2002). Maintaining academic integrity in on-line education. *Online Journal of Distance Learning Administration, 5*(2). Retrieved from <http://www.westga.edu/~distance/ojdl/spring51/spring51.html>
- Hickson, S., and Agnew, S. (2011). Assigning grades during an earthquake – Shaken or stirred? Department of Economics and Finance, College of Business and Economics, University of Canterbury. Retrieved from [http://www.econ.canterbury.ac.nz/research/working\\_papers.shtml](http://www.econ.canterbury.ac.nz/research/working_papers.shtml)
- Kleiner, C., & Lord, M. (1999). The cheating game: Cross-national exploration of business students' attitudes, perceptions, and tendencies toward academic dishonesty. *Journal of Education for Business, 74*(4), 38–42.
- Meyer, K., & Wilson, J. (2011). The role of online learning in the emergency plans of flagship institutions. *Online Journal of Distance Learning Administration, 14*, 1.
- Olt, M. (2002). Ethics and distance education: Strategies for minimizing academic dishonesty in online assessment. *Online Journal of Distance Learning Administration, V*(III). Retrieved from <http://www.westga.edu/~distance/ojdl/fall53/olt53.html>
- Omar, A., Liu, L. C., & Koong, K. S. (2008). From disaster recovery to mobile learning: A case study. *International Journal of Mobile Learning and Organisation, 2*(1), 4–7.

- Perrin, K. M., & Mayhew, D. (2000). The reality of designing and implementing an internet-based course. *Online Journal of Distance Learning Administration*, 3(4). Retrieved from <http://www.westga.edu/~distance/ojdla/winter34/mayhew34.html>
- Robles, M., & Braathen, S. (2002). Online assessment techniques. *The Delta Pi Epsilon Journal*, XLIV(1), 39–49.
- Rowe, N. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration*, VII(II). Retrieved from <http://www.westga.edu/~distance/ojdla/summer72/rowe72.html>
- Russell, J., Elton, L., Swinglehurst, D., & Greenhalgh, T. (2006). Using the online environment in assessment for learning: A case-study of a web-based course in primary care. *Assessment and Evaluation in Higher Education*, 31(4), 465–478.
- SchWeber, C. (2008). Determined to learn: Accessing education despite life threatening disasters. *Journal of Asynchronous Learning Networks*, 12(1), 37–43.
- Stiggins, R., & J. Chappuis. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44(1), 11–18.

## Biographical notes

### Steve Agnew

[steve.agnew@canterbury.ac.nz](mailto:steve.agnew@canterbury.ac.nz)

Steve Agnew is a Senior Teaching Fellow in the Department of Economics and Finance, College of Business and Economics at the University of Canterbury, New Zealand. Steve has research interests in economics education and financial literacy at secondary school and tertiary level.

### Stephen Hickson

[stephen.hickson@canterbury.ac.nz](mailto:stephen.hickson@canterbury.ac.nz)

Stephen Hickson is a Senior Teaching Fellow in the Department of Economics and Finance, College of Business and Economics at the University of Canterbury, New Zealand. Stephen has research interests in economics education as well as law and economics. He teaches economics to undergraduates and MBA level students.

Agnew, S., & Hickson, S. (2012). Using online assessment to replace invigilated assessment in times of natural disaster: Are some online assessment conditions better than others? *Journal of Open, Flexible and Distance Learning*, 16(1), [pp. 1–13].



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

## Appendix A: Spearman rank-order correlation coefficients

Table A1(a) ECON 105 (Macroeconomics) Semester 1, 2005, to Semester 1, 2009

	Online MC	Term test	Final exam
	<b>Pearson correlation coefficients</b>		
Online MC	1.00	0.27-0.47	0.48-0.66
Term test	0.24-0.41	1.00	0.73-0.81
Final exam	0.35-0.50	0.76-0.83	1.00
	<b>Spearman rank-order correlation coefficients</b>		

Table A1(b) ECON 105 (Macroeconomics) Semester 2, 2009, and Semester 1, 2010

	Online MC	Term test	Final exam
	<b>Pearson correlation coefficients</b>		
Online MC	1.00	0.55-0.55	0.65-0.66
Term test	0.51-0.55	1.00	0.79-0.80
Final exam	0.52-0.61	0.80-0.82	1.00
	<b>Spearman rank-order correlation coefficients</b>		

Table A1(c) ECON 105 (Macroeconomics) Semester 2, 2010

	Online MC	Term test	Final exam
	<b>Pearson correlation coefficients</b>		
Online MC	1.00	0.48	0.67
Term test	0.45	1.00	0.69
Final exam	0.55	0.73	1.00
	<b>Spearman rank-order correlation coefficients</b>		

Table A1(d) ECON 105 (Macroeconomics) Semester 1, 2011

	Online MC	Term test	Final exam
	<b>Pearson correlation coefficients</b>		
Online MC	1.00	0.57	N/A
Term test	0.46	1.00	N/A
	<b>Spearman rank-order correlation coefficients</b>		

Table A1(e) ECON 105 (Macroeconomics) Semester 2, 2011

	Online MC	Term test	Final exam
	<b>Pearson correlation coefficients</b>		
Online MC	1.00	0.66	0.77
Term test	0.63	1.00	0.83
Final exam	0.74	0.84	1.00
	<b>Spearman rank-order correlation coefficients</b>		

## Appendix B: Grade distribution if exam hadn't run

Table A2(a) Weighting (tutorial quizzes [highest scoring attempt] 65%)

		Alternative grade										
		A+	A	A-	B+	B	B-	C+	C	C-	D	E
True grade	A+	39	1									
	A	21	3									
	A-	17	5	3	1	2						
	B+	5	4	5	4	2	2					
	B	6	7	8	5	1		1	1			
	B-	4	5	2	5	1	4	1			1	
	C+	5	5	9	1	2	2	1	1	1	1	
	C	2	5	2	4	4	2	1	1		1	5
	C-	1	3	2	1	3	3		1	5		3
	D			3	1	2	1	4	2	3	1	1
E		2	1		2	2	4	2	8	5	33	

Table A2(b) Weighting (tutorial quizzes [first attempt] 65%)

		Alternative grade										
		A+	A	A-	B+	B	B-	C+	C	C-	D	E
True grade	A+	24	7	2								
	A	12	5	6	2	1						
	A-	4	4	6	5	3		1				
	B+	3		6	4	4	7	3				
	B		2	4	5	6	2	2				
	B-			2	2	5	10	3	4	1	1	
	C+			1	5	7	5	4	1		2	1
	C		1	2	5	3	5	2	3	1		8
	C-			1	1	1	3	1	2	3	4	5
	D				2	3	1	2	2	2	2	6
E		1			2		2	1	5	8	46	