DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

ASSIGNING GRADES DURING AN EARTHQUAKE – SHAKEN OR STIRRED?

Stephen Hickson and Stephen Agnew

WORKING PAPER

No. 40/2011

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

ASSIGNING GRADES DURING AN EARTHQUAKE – SHAKEN OR STIRRED?

by

Stephen Hickson and Stephen Agnew*

Abstract

In the event of an unanticipated disruption to normal life, universities tend to shift to an online environment in both delivery and assessment. Course instructors still need to assign grades despite not having the full set of planned assessments. This paper examines how grades are disrupted when an increased reliance is placed on online assessments. We find substantial grade disruption and grade inflation as the weighting on online assessments rises relative to invigilated assessments. Grade inflation can be moderated by scaling to an historical distribution of grades; however such scaling can lead to substantial grade disruption where the quality of the cohort is different than the historical average. We also find evidence that time limited online assessments produce lower grade disruptions as weighting on the online component increases.

JEL Categories: A22

Keywords: Principles of Economics, Online Assessment, Student Grades, Disruption to Assessment, Earthquake.

25 November 2011

*Stephen Hickson and Stephen Agnew are both Teaching Fellows at the University of Canterbury. Their contact details are: Department of Economics and Finance, University of Canterbury, Private Bag 4800, Christchurch 8042, New Zealand; Email: stephen.hickson@canterbury.ac.nz and steve.agnew@canterbury.ac.nz; Phone: +64 3 364 2847.

I. INTRODUCTION

When all goes well assessments in university courses occur as planned. Students know at the start of a course what the assessment items are and what weight they have. The structure of the assessment influences the behaviour of students as they attempt to balance the work in that particular course, their work in other courses and interests outside study.

The experience of the University of Canterbury during the second semester of 2010 (2010-S2) and the first semester of 2011 (2011-S1) following significant earthquakes has been that unanticipated disruptions to assessment certainly can occur. That the disruption is unanticipated is important as students have made choices (e.g. time allocation) given the notified assessment schedule. Under a different scheme they are likely to have made different choices.

When a substantial disruption occurs, the most likely items of assessment to be cancelled completely are invigilated tests and exams. Online assessments are more likely to continue as students can complete these remotely. However course lecturers are still required to assign grades and must use the assessment that does take place to do so. As it transpired, online assessments proved to be invaluable in an earthquake disrupted semester. 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). It also reduced the need to use markers who themselves were earthquake disrupted.

According to the limited research that exists on this topic, moving to a more online format is common in times of natural disasters. As a result of a norovirus outbreak

which closed the Hope College campus for four days, Benton (2009) describes the college now being able to “come close to sustaining the experience of the traditional classroom from dispersed locations” (p. 4). Further examples of a natural event leading to a greater online focus include George Washington University, which uses e-learning as a central part of their plan for educational continuity, prompted by the H1N1 flu. According to George Washington’s Yordanos Baharu, “Part of what we’re doing in training is getting faculty to think about plan B. With this plan, we’re confident that we can mitigate potential disruptions and provide students and faculty the support they need to continue teaching with Blackboard’s system” (ecampus news, p. 1). When Hurricane Ike shut down the Clear Creek Independent school district for nearly two weeks, “teachers and students leveraged online learning to avoid missing academic targets for the year thanks to the ability to communicate and complete assignments even while school buildings were closed” (ecampusnews, p. 1). In response to SARS and Avian flu outbreaks in 2005, Singapore’s Nanyang Technological University (NTU) launched a preparedness program which closes segments of its campus for one week at a time. Students are able to receive lesson plans, watch lecture videos, and complete assignments and tests online. According to Daniel Tan from NTU, “What we’re trying to achieve is learning continuity. Our plan allows university officials to close the campus with a high level of confidence that education operations can continue successfully online”. (ecampus news p. 2). Meyer & Wilson (2011) mention that the use of online learning as a result to disasters is new, and there are few studies looking into this issue. They do however cite Omar, Liu & Koong, (2008) describing The Southern University at New Orleans becoming an online learning campus after being struck by hurricanes. In particular they

discuss continuous education being provided to displaced students through the use of mobile devices. Foster & Young (2005) are also referenced describing the University of New Orleans offering a significant number of additional online courses than in previous semesters. Meyer & Wilson (2011) also refer to Hartman and Lundberg (2009) who promoted online education as the "vehicle for meeting both sets of needs" (p. 593) when referring to the need to support individuals through a disaster but to also "sustain academic work" (p. 3). Finally Danielson (2009) is mentioned as reporting online classes and skype being used at the University of South Florida in the event of an emergency. SchWeber (2008) traces "distance learning as a solution in the face of danger" (p. 38) back to the second world war, where correspondence courses were implemented in France, "which would follow the same program, methods used in the schools and with the same instructors" (p. 38). SchWeber (2008) refers to war again in 2006, where educational institutions in Israel and Lebanon provided online courses for impacted students. SchWeber (2008) also references the *Sloan Semester* Project, an electronic partnership between Louisiana and Mississippi Colleges which allowed their students to continue their studies online subsequent to hurricanes Katrina and Rita.

The above literature focuses on a shift of predominantly teaching resources to an online environment in response to a natural event, rather than assessment becoming predominantly online. This paper seeks to fill this void by examining how effective different forms of assessment are in assigning grades in an earthquake or otherwise affected semester.

In particular the contribution that this paper will make is in examining how well online assessments perform in the task of assigning grades in our Principles of

Microeconomics (ECON104) and Principles of Macroeconomics (ECON105) courses when the assessment schedule faces a substantial unanticipated disruption. We focus particularly on 2011-S1 due to the extent of the disruption that occurred. We model a similar disruption in our historical data (pre 2010-S2) with similar assessment items and replicate the disruption to 2011-S1 with the 2011-S2 student data where the assessment items are identical.

We define the term "true grade" to mean the grade that a student receives when all assessment items occur as planned. Determining what assessment is "optimal" in some sense is beyond the scope of this paper. We simply take as given the assessment regime that is in place at any particular time. We are then able to compare this "true grade" to an "alternative grade". We define the "alternative grade" as the one that is assigned when the notified assessment schedule is disrupted due to an unanticipated shock. By converting grades to the University of Canterbury GPA[1] scale we are then able to quantify the extent of the change that would occur. Of course when a disruption occurs it is not possible to observe the "true grade" but we are able to take advantage of the data we have from other semesters to model such disruptions.

Section 2 describes the assessments, data and methods, section 3 is the discussion of results and section 4 concludes.

[1] GPA is awarded as follows: A+=9, A=8 etc down to E=-1.

II. ASSESSMENTS, DATA AND METHODS

In this section we discuss the assessments used in ECON 104 and 105 over the time period in question, the data used in this study and the methods applied to the data. Results are discussed in the next section.

From 2005-S1 to 2010-S2 both courses have online multiple-choice (MC) tests worth 10%, an assignment worth 10%, a term test and a final exam. The structure of the term test and final exam remained unchanged though the weight applied to each was changed slightly during this period. From 2007-S1 a minimum mark was required in the final exam in order to receive a continuing pass (C or better) in the course.[2]

There were 10 online multiple choice tests each worth 1%. Students could attempt each test as many times as they wished and their highest mark was the one that counted. Questions were drawn randomly from a large bank of questions sourced from the text book supplier. For most of the period these tests were open for the entire semester rather than being time limited. However, this was changed in ECON 105 for 2009-S2 onwards and the online MC tests were only open for a short window around the time the material was covered in class. Students could still have as many attempts as they wished during this period.

The September 4 quake did cause some disruption to 2010-S2 courses although it was relatively minor compared to 2011-S1. A week of classes was lost in all courses. Plussage was allowed in ECON 104 allowing a student to receive the better of 20/60 or 0/80 in the test and exam respectively. All final exams were reduced from three hours to two.

[2] Students who scored more than 50% in the course overall but did not meet the minimum requirement in the final exam received a restricted (C-) pass meaning they passed the course but could not use it as a pre-requisite for future courses.

At the end of 2010 a number of factors unrelated to the earthquakes prompted the Department to change how the two principles courses were assessed. The traditional tutorial format was widely regarded as being ineffective with poor attendance. The expenditure of scarce funds was not providing good returns. Additionally, an increased focus by the university on student engagement and achievement meant a greater desire to provide earlier indications of non-engagement and a desire to target resources at those students who would benefit the most from tutorial type support.

From 2011-S1 assessment in both ECON 104 and 105 was modified. In ECON 104 the assignment was dropped while tutorial quizzes and a progress test were introduced. The progress test is a 30 question MC test completed in week six, worth five percent. Students have one week to complete the test but once started it must be completed in 90 minutes. The questions for each student are randomly drawn from a test bank. Tutorial quizzes worth 10 percent were introduced where students complete weekly mini tests with the questions being drawn from the same test bank as the progress test. The number of text book based online MC quizzes was reduced to five and became worth five percent in total. The term test and final exam continued as before.

In ECON 105 the assignment and online MC quizzes were retained in their previous format. However, five online tutorials each worth two percent were introduced. Students could make one attempt at each tutorial but the window of opportunity was usually about two weeks. Those who achieved 75 percent or more received the full two percent and did not need to attend a classroom based tutorial though could make use of drop in sessions if they wished. Students who did not make the 75 percent were required to attend a classroom based tutorial in order to receive any credit. The term test and the

final exam continued as before though the weight on the final exam was reduced from 60 percent to 50 percent.

Following the February 22 quake the assignment in ECON 105 was cancelled due to uncertainty over the availability of on-campus computer resources. There was no other disruption to assessment at that point though there was some discussion on whether or not term tests should be held. Ultimately term tests did go ahead. The June 13 quake occurred on the eve of final exams and as a result the ECON 105 exam was cancelled. The ECON 104 exam was cancelled for those students who were given a passing grade based on their other course work. Those who did not receive a passing grade based on their other course work were eligible to sit a special exam.[3]

Hence the grades allocated in 2011-S1 are "alternative grades" rather than "true grades" as we have defined it. While it is not possible to know "true grades" for 2011-S1, it is possible to construct alternative grades for students in other occurrences of the same course. We take advantage of having student data from 2005 to 2010 to examine different scenarios that allow us to understand the impact that removing the final exam would have had historically and thereby calculate "alternative grades". We also replicate the semester 2011-S1 assessment with 2011-S2 students who undertook the full range of assessment identical to 2011-S1.

In the historical data (2005-S1 to 2010-S1) we exclude 2010-S2 as it was impacted by the September 4 quake. For the whole sample we delete students who did

[3] Of the 155 students eligible to sit the special exam, only 103 actually did sit. Of those only 21 managed to improve their grade and eight of these were students who had not been able to attend the term test but were clearly good students. For those who had actually sat the term test, the highest grade they could be awarded was a C.

not attempt any assessment at all and those students who received "aegrotat" grades[4]. That gives a large dataset of 8752 observations. For each student we know their assignment, online multiple-choice test, term test and final exam scores.

For the historical data our interest is in how the online multiple-choice performs the task of assigning grades in conjunction with the term test since this most closely replicates what occurred for 2011-S1 where the final exam was cancelled. The more highly correlated the online assessment items are with the final exam, the better they will proxy for the final exam. To identify the level of correlation we calculate both the Pearson and the rank order Spearman correlation co-efficient. The rank order correlation is particularly important when considering the possible use of scaling.

However, correlation coefficients are not sufficient. Even though, as expected, all the assessment items are positively correlated with each other it is not clear what the impact is on *grades* given the correlation is not perfect. Ultimately what students are interested in are grades. We therefore calculate alternative grades based on different sets of weights for the online multiple-choice quizzes and the term test. We are then able to calculate the change in the GPA from the true grade if a student is awarded an alternative grade. A higher alternative grade than the true grade results in a positive difference.

A clear issue that arises is that online assessments typically have higher means than invigilated assessments. Hence the greater the weight that is given to online assessments the more the average GPA for the alternative grades will rise compared to the average GPA for the true grades. This can be avoided by scaling to a pre-determined distribution.

[4] An aegrotat grade is awarded when a student misses a test or exam due to illness or other critical circumstance or when a student's performance in the test or exam is directly impacted by such circumstances. Hence performance in tests and exams is not truly reflective of ability or final grade.

We calculate average historical distributions for each of the 4 different occurrences – 104 S1 and S2 and 105 S1 and S2 – and apply these to the alternative rank ordering of students. The reason that each set of occurrences must be handled separately is that the four occurrences within a year actually have different distributions. This is most marked between 104-S1 and 104-S2 where students who have failed the S1 occurrence re-take 104 in S2. Hence a different distribution was applied to each of the four different types of occurrences.

III. DISCUSSION OF RESULTS

2005 to 2010 Semester 1

The mean score for the online multiple-choice tests (75.8) and assignment (64.4) are higher than the invigilated term test (50.9) and final exam (53.7).

Tables 1(a) to 1(c) show the ranges of the Pearson correlation coefficients and the Spearman rank order correlation coefficients for the four assessment items across all courses. For ECON 104 (Microeconomics) all the courses are included in one table as the assessment was the same across all the years. ECON 105 (Macroeconomics) is split into two tables. A crucial change was made from 2009-S2 onwards where the online multiple-choice questions were opened for a limited window around the time the material was covered rather than being open for the whole semester.

[TABLE 1 HERE]

The correlation coefficients for 2005-S1 to 2010-S1 (table 1) are all positive and significant at the 0.01 level. However they do show some degree of variability. The online multiple-choice tests are more strongly correlated with the term test and the final exam for ECON 104 than for ECON 105. For 104 the Pearson coefficient ranges from 0.61 to 0.81 for the final exam while for 105 the range is 0.48 to 0.66. This may reflect the more “textbook” nature of microeconomics compared to macroeconomics particularly when it comes to invigilated tests and exams.

For 105 the online multiple-choice tests are more highly correlated with the term test and final exam when they are time limited (see table 1(c) compared to 1(b)). For courses prior to 2009-S2 the Spearman rank order coefficients are all less than the values for 2009-S2 and 2010-S1.

However, as previously mentioned, correlation measures are not sufficient. What matters to students is the impact on grades when an "alternative grade" is required (as compared to a "true grade" when all assessment takes place).

Table 2 shows the percentage of students that would experience a particular change in GPA over the whole dataset when grades are recalculated using different weighting schemes for the online MC tests and the term test. Table 3 shows the same impact on GPA when the same weightings are applied but the resulting grade distributions are scaled to the historical average.

[TABLE 2 HERE]

[TABLE 3 HERE]

There are a small number of extreme grade changes where students have differences of more than 4 GPA points (either plus or minus) between their true grade and their alternative grade. In general the big negative changes (where the alternative grade is much lower than the true grade) are caused by the student not doing the online tests but doing well in the term test and final exam. Some students who are confident in their own

ability and are in fact relatively good students may not consider the online tests worth doing given their low weighting.

The large positive changes (where the alternative grade is much higher than the true grade) are caused by the student doing relatively well in the online MC tests and poorly in the exam compared to the term test. A grade based on the online MC and term test is thus much higher. There are a number of students in the sample who do well in the online tests and the invigilated term test and then for reasons unknown simply do not even sit the final exam.

Tables 2 and 3 also show the percentage of students that would experience a GPA disruption of +/-1 or +/-2 GPA points. For most of the score to grade mapping, 1 GPA point corresponds to 5 percentage points (e.g. a B grade is 65 – 69.9 percent and a B+ is 70 to 74.9 percent). Two GPA points therefore corresponds to 10 percentage points in a raw score.

While there will be some students whose grade changes by more than +/-2 GPA points, what is desirable is a minimisation of this disruption to grades. What is clear from tables 2 and 3 is that the disruption to grades becomes more extreme as the weight on the online multiple choice tests rises. When the weight applied to the online multiple choice tests is 0.8 then 48 percent of students lie in the +/-2 GPA point range when using raw scores (table 2) although this rises to 77.8 percent if scaling is used (table 3). In contrast when the weight is 0.2 the values are 91.5 and 92.6 percent respectively.

We can see the two weighting schemes of (0.2, 0.8) and (0.4, 0.6) as two different approaches to a disruption to assessment. For the first this is similar to loading the weight of the missed final exam onto the term test (which is how the disruption was

handled in 2011-S1 for ECON 104). For the second this is similar to spreading the weight of the missed final exam across the online MC tests and term test (which is how the disruption was handled in 2011-S1 for ECON 105). Both of these approaches produce similar disruptions to grades with at least 85 percent of students experiencing no more than a +/-2 GPA point change to their true grade. Both of these weighting schemes are better than not including the online tests at all which is not surprising since the true grade includes the online tests.

What tables 2 and 3 also show is that a complete reliance on online tests is likely to produce substantial disruptions to grades. The weighting scheme (1.0, 0.0) is the case where all invigilated assessment is cancelled and the grade is assigned purely on the basis of in course work. In this case only 39.1 percent of students are in the +/-2 GPA point difference range when using raw scores though this does rise to 69.9 percent if scaling is used. Some students would experience extreme grade changes. These students will have done well in the online tests and very poorly in the term test and final exam.

Noticeable from these results is the grade inflation that occurs as the weight on the online tests rises. At the extreme end where all the weight is placed on the online tests, the average rise in GPA is 3.4 points. However, as table 3 shows, overall grade inflation can be removed via scaling but the extent of the disruption to individual students is still substantial with the range of disruption extending from -8 to +9 GPA points.

Summarising to changes in GPA is useful and instructive but does not reveal how grades are impacted for different grade bands. Some readers may find this information useful. The tables in appendix 1 show true grades vs. alternative grades for each grade band using two example weight schemes, (0.2, 0.8) and (0.8, 0.2).

Time limiting of online tests appears to make a difference. Tables 4 and 5 compare results for all courses (2005-S1 – 2010-S1) to where the online tests are only open for a limited time (105-S2 2009 and 105-S1 2010). The use of time limited online tests leads to less disruption to grades compared to when tests are open for the whole semester, particularly when the weight applied to the online assessment rises. The weighting schemes (0.4, 0.6) and (0.8, 0.2) show more students in the zero change in grade group and higher percentages in the +/-2 GPA point change groups.

[TABLE 4 HERE]

[TABLE 5 HERE]

2011 Semester 2

We then apply the 2011-S1 weighting schemes (where the final exam was cancelled) to the 2011-S2 data where all assessment items took place as planned. We are then able to compare the theoretical alternative grades to the known true grades. Table 6 shows means for the assessment items in 2011 semesters 1 and 2. Table 7 shows the correlation coefficients (Pearson and Spearman) for semester 2 ECON 104 and ECON 105. Table 8 shows the disruption to grades using both raw scores and scaling to an historical distribution. The tables in appendix 2 show the disruption to each grade band.

[TABLE 6 HERE]

[TABLE 7 HERE]

[TABLE 8 HERE]

Both sets of weightings performed slightly better than what might have been expected given the historical simulations. In this simulation 87 percent of students would experience a +/-2 change in GPA in ECON 104 and the 95% in ECON 105. These are the two highest values when compared with the historical simulations (table 2). The overall changes to GPA were in line with the historical simulations (+0.9 and -0.3 for 104 and 105 respectively).

In ECON 104, scaling would have been very disruptive. The reason for this is a change in the cohort. The historical group of students on which the historical grade distribution is based is markedly different from the current cohort. For example, changes in the wider university mean that engineering students now tend to take ECON 104 in S2 rather than S1 with engineering students tending to be strong students.

Recall that in ECON 104 semester 1 the weighting for the cancelled final exam was allocated entirely to the term test while in ECON 105 the weighting for the final exam was distributed equally across all the available assessment items. Does our data shed any light on which approach might be better in general? Since the approach used in ECON 105 (macroeconomics) gave a 94.5 percent value for the +/-2 GPA change range and the approach used in ECON 104 (microeconomics) gave 87.3 percent it would appear

at first look that spreading evenly is better. If that hypothesis is correct then we should see a lower grade distribution in ECON 104 if we apply the ECON 105 approach (spread evenly). Further we should greater grade disruption in ECON 105 if we apply the theoretically inferior ECON 104 approach (i.e. all on the term test). Table 9 shows the 2011-S2 simulations with the weighting approaches reversed. The interesting outcome is that both courses actually perform worse with the reverse weighting schemes.[5]

[TABLE 9 HERE]

[5] The authors would like to note that they do not take any credit for this brilliant piece of wisdom of applying differing weight schemes in their respective courses – the benefits of which are only seen in hindsight.

IV. CONCLUSION

Improvements in technology have resulted in a shift to more online assessment particularly in large first year courses. In the event of an unanticipated disruption to normal life, universities tend to further shift to an online environment in both delivery and assessment. Course instructors still, however, need to assign grades despite not having the full set of planned assessments.

Online assessments are correlated with other items of assessment such as invigilated tests and exams but not perfectly so. As a result we find substantial grade disruption as the weighting on online assessments increases relative to invigilated assessments as evidenced by the fall in the percentage of students in the +/-2 GPA change range.

We find grade inflation occurs as the weighting applied to online tests increases. Grade inflation can be moderated by scaling to an historical distribution of grades; however such scaling can lead to substantial grade disruption where the quality of the cohort is different than the historical average. This implies that instructors should use raw data rather than employ scaling unless it is certain there is no change in the cohort quality compared to previous years.

In the disrupted semester 1 of 2011 two slightly different approaches were used in Microeconomics (104) and Macroeconomics (105) to re-distribute the final exam weighting, viz. (i) allocating all the final exam weight to the term test; and (ii) distributing the final exam weight evenly. We do not find evidence that supports one or other of these approaches is better in a general sense. The choice of which approach to use can be left to the instructor's judgment and will depend on factors particular to

individual courses and institutions. We do find that at least one piece of invigilated assessment is crucial so planning two pieces of invigilated assessments is a sound risk management strategy.

We find evidence that time restricted online tests lead to less disruption to grades compared to when tests are open for the whole semester particularly as the weighting applied to the online tests increases. This is an issue for further examination.

REFERENCE LIST

Benton, T. H. (2009, November 30). Teaching in the plague year. *The Chronicle of Higher Education*. Retrieved November 1, 2011 from http://chronicle.com/article/Teaching-in-the- Plague-Year/49275.

Danielson, R. (2009, August 24). If widespread swine flu closes USF campus, classes will go on. Retrieved November 1, 2011 from http://www.tampabay.com/.

eCampus News. (2009, September 11). Web classes could help in H1N1 outbreaks.

Foster, A.L., & Young, J.R. (2005, September 16). The Internet as emergency tool. *The Chronicle of Higher Education*. Retrieved November 1, 2011 from http://chronicle.com/article/The-Internet-as-Emergency-Tool/15881.

Hartman, S. J., & Lundberg, O. H. (2009). The experiences of New Orleans college administrators, faculty and students during and following Hurricane Katrina. *International Journal of Innovation and Learning, 6*(6), 593-605.

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*.

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* [On-line serial], *3*(4). Retrieved November 1, 2011, from http://www.westga.edu/~distance/ojdla/winter34/mayhew34.html .

SchWeber, C. (2008). Determined to learn: Accessing education despite life threatening disasters. *Journal of Asynchronous Learning Networks*, 12 (1), 37-43.

TABLE 1
Range of Correlation Coefficients for Assessment Items (2005-S1 to 2010-S1)

Table 1(a) ECON 104 (Microeconomics) All Courses

	Assignment	*Online MC*	*Term Test*	*Final Exam*
		Pearson Correlation Coefficients		
Assignment	1.00	0.48 – 0.69	0.48 – 0.74	0.52 – 0.72
Online MC	0.42 – 0.57	1.00	0.49 – 0.77	0.61 – 0.81
Term Test	0.51 – 0.69	0.44 – 0.65	1.00	0.65 – 0.85
Final Exam	0.51 – 0.68	0.59 – 0.73	0.68 – 0.84	1.00
	Spearman Rank Order Correlation Coefficients			
	Assignment	*Online MC*	*Term Test*	*Final Exam*

Table 1(b) ECON 105 (Macroeconomics) 2005-S1 to 2009-S1

	Assignment	*Online MC*	*Term Test*	*Final Exam*
		Pearson Correlation Coefficients		
Assignment	1.00	0.36 – 0.53 (1)	0.41 – 0.57	0.49 – 0.61
Online MC	0.21 – 0.44	1.00	0.27 – 0.47	0.48 – 0.66
Term Test	0.42 – 0.58	0.24 – 0.41	1.00	0.73 – 0.81
Final Exam	0.46 – 0.60	0.35 – 0.50	0.76 – 0.83	1.00
	Spearman Rank Order Correlation Coefficients			
	Assignment	*Online MC*	*Term Test*	*Final Exam*

(1) the value for 2008-S1 of 0.16 was excluded being a highly unusual outlier.

Table 1(c) ECON 105 (Macroeconomics) 2009-S2 and 2010-S1

	Assignment	*Online MC*	*Term Test*	*Final Exam*
		Pearson Correlation Coefficients		
Assignment	1.00	0.53 – 0.53	0.53 – 0.59	0.58 – 0.61
Online MC	0.41 – 0.52	1.00	0.55 – 0.55	0.65 – 0.66
Term Test	0.48 – 0.60	0.51 – 0.55	1.00	0.79 – 0.80
Final Exam	0.52 – 0.66	0.52 – 0.61	0.80 – 0.82	1.00
	Spearman Rank Order Correlation Coefficients			
	Assignment	*Online MC*	*Term Test*	*Final Exam*

All coefficients are significant at the 1% level of significance.

TABLE 2
Percent of Students and the Change in GPA Under Different Weighting Schemes Using Raw Scores.

The first number shown in the brackets is the weight applied to the online multiple-choice and the second is the weight for the term test.

	Weighting Scheme					
Change in GPA	**(0.0, 1.0)**	**(0.2, 0.8)**	**(0.4, 0.6)**	**(0.6, 0.4)**	**(0.8, 0.2)**	**(1.0, 0.0)**
+10					0.2	1.6
+9					0.9	2.0
+8					2.4	4.5
+7				0.7	4.2	6.6
+6	0.1	0.1	0.1	2.5	6.3	8.6
+5	0.2	0.2	0.9	5.8	9.5	11.4
+4	0.6	0.7	3.2	10.3	13.0	12.6
+3	1.5	2.6	8.8	16.0	14.7	12.6
+2	4.2	8.1	18.0	19.0	14.9	11.8
+1	10.0	17.9	24.3	18.2	12.4	9.0
0	31.9	37.6	32.2	22.5	18.0	16.1
-1	19.1	18.2	8.4	3.4	2.0	1.5
-2	15.7	9.7	2.8	1.0	0.7	0.7
-3	9.7	3.8	0.7	0.4	0.4	0.4
-4	4.8	1.0	0.1	0.2	0.2	0.2
-5	1.6	0.2	0.1	0.1	0.1	0.1
-6	0.4			0.1	0.1	0.1
-7	0.1					
-8						
-9						
-10						
Overall Change in GPA	-0.8	0.0	+0.9	+1.9	+2.8	+3.4
Percent of students in range (-1,+1)	61.0	73.7	64.9	44.1	32.4	26.6
Percent of students in range (-2,+2)	80.9	91.5	85.7	64.1	48.0	39.1
N	8752	8752	8752	8752	8752	8752

TABLE 3
Percent of Students and the Change in GPA Under Different Weighting Schemes Using the Historical Grade Distributions.

The first number shown in the brackets is the weight applied to the online multiple-choice and the second is the weight for the term test.

Change in GPA	**(0.0, 1.0)**	**(0.2, 0.8)**	**(0.4, 0.6)**	**(0.6, 0.4)**	**(0.8, 0.2)**	**(1.0, 0.0)**
+10						
+9						0.1
+8						0.3
+7					0.1	0.8
+6	0.2	0.1			0.5	1.2
+5	0.4	0.1	0.1	0.4	1.5	2.4
+4	1.1	0.7	0.7	1.6	3.1	4.1
+3	3.9	2.8	2.9	4.6	6.0	6.1
+2	9.4	8.9	8.9	9.5	9.5	8.9
+1	17.6	19.1	19.1	17.2	14.0	11.8
0	34.7	37.8	38.1	34.4	29.5	26.1
-1	17.4	18.0	17.3	16.1	14.4	12.1
-2	10.0	8.8	8.5	9.6	10.4	11.0
-3	4.0	3.1	3.3	4.6	6.8	7.8
-4	1.0	0.7	0.6	1.4	2.7	4.4
-5	0.2	0.1	0.1	0.3	1.0	2.0
-6			0.1	0.1	0.3	0.7
-7					0.1	0.1
-8						0.1
-9						
-10						
Overall Change in GPA	0.0	0.0	0.0	0.0	0.0	0.0
Percent of students in range (-1,+1)	69.7	74.9	74.5	67.7	57.9	50.0
Percent of students in range (-2,+2)	89.1	92.6	91.9	86.8	77.8	69.9
N	8752	8752	8752	8752	8752	8752

TABLE 4
Percent of Students and the Change in GPA Under Different Weighting Schemes Using Raw Scores - Comparison of all Courses and ECON 105-S2 2009 Onwards.

The first number shown in the brackets is the weight applied to the online multiple-choice and the second is the weight for the term test.

Change in GPA	**All courses (0.2, 0.8)**	**105-S2 2009 onwards (0.2, 0.8)**	**All courses (0.4, 0.6)**	**105-S2 2009 onwards (0.4, 0.6)**	**All courses (0.8, 0.2)**	**105-S2 2009 onwards (0.8, 0.2)**
+10					0.2	
+9					0.9	0.3
+8					2.4	3.1
+7					4.2	2.9
+6	0.1		0.1		6.3	4.7
+5	0.2		0.9	0.2	9.5	7.3
+4	0.7		3.2	0.1	13.0	11.3
+3	2.6	0.5	8.8	6.6	14.7	14.6
+2	8.1	4.5	18.0	11.2	14.9	16.2
+1	17.9	12.0	24.3	20.4	12.4	13.6
0	37.6	30.3	32.2	35.8	18.0	21.7
-1	18.2	23.5	8.4	15.8	2.0	1.8
-2	9.7	18.3	2.8	6.3	0.7	1.0
-3	3.8	8.6	0.7	2.0	0.4	0.5
-4	1.0	2.1	0.1	0.3	0.2	0.8
-5	0.2	0.2	0.1	0.2	0.1	0.3
-6		0.2		0.2	0.1	
-7						
-8						
-9						
-10						
OVERALL CHANGE IN GPA	-0.1	-0.7	+0.9	+0.3	+2.8	+2.4
Percent of students in range (-1,+1)	73.7	65.8	64.9	72.0	32.4	37.1
Percent of students in range (-2,+2)	91.5	88.6	85.7	89.6	48.0	54.3
N	8752	618	8752	618	8752	618

TABLE 5
Percent of Students and the Change in GPA Under Different Weighting Schemes Using Historical Distribution of Grades - Comparison of all Courses and ECON 105-S2 2009 Onwards.

The first number shown in the brackets is the weight applied to the online multiple-choice and the second is the weight for the term test.

Change in GPA	**All courses (0.2, 0.8)**	**105-S2 2009 onwards (0.2, 0.8)**	**All courses (0.4, 0.6)**	**105-S2 2009 onwards (0.4, 0.6)**	**All courses (0.8, 0.2)**	**105-S2 2009 onwards (0.8, 0.2)**
+10						
+9						
+8						
+7					0.1	
+6	0.1				0.5	0.2
+5	0.1		0.1		1.5	2.1
+4	0.7	0.3	0.7	0.8	3.1	2.4
+3	2.8	2.6	2.9	2.4	6.0	6.3
+2	8.9	7.8	8.9	8.4	9.5	8.7
+1	19.1	15.7	19.1	15.4	14.0	11.8
0	37.8	35.6	38.1	36.9	29.5	29.0
-1	18.0	22.7	17.3	19.3	14.4	13.3
-2	8.8	10.7	8.5	10.7	10.4	12.0
-3	3.1	4.4	3.3	4.7	6.8	9.2
-4	0.7	0.2	0.6	1.1	2.7	3.4
-5	0.1	0.2	0.1	0.2	1.0	0.8
-6			0.1	0.2	0.3	0.5
-7					0.1	0.3
-8						
-9						
-10						
OVERALL CHANGE IN GPA	0.0	-0.2	0.0	-0.2	0.0	-0.2
Percent of students in range (-1,+1)	74.9	74.0	74.5	71.6	57.9	54.1
Percent of students in range (-2,+2)	92.6	92.5	91.9	90.7	77.8	74.8
N	8752	618	8752	618	8752	618

TABLE 6
Mean Value of Assessment Items by Course (2011)

Course	Assign-ment	Online MC tests	Progress test	Tutorials	Term test	Final exam
ECON104S111	n.a	76.2	75.4	58.3	63.1	cancelled
ECON105S111	cancelled	77.6	n.a	60.0	57.5	cancelled
ECON104S211	n.a	64.2	69.4	72.7	63.3	52.6
ECON105S211	68.1	74.6	n.a	62.1	55.5	60.3

TABLE 7
Correlation Coefficients for Assessment Items (2011-S2)

Table 7(a) ECON 104 (Microeconomics)

	Online MC	*Tutorials*	*Progress test*	*Term Test*	*Final Exam*
		Pearson Correlation Coefficients			
Online MC	1.00	0.75	0.61	0.68	0.69
Tutorials	0.70	1.00	0.60	0.63	0.64
Progress test	0.62	0.59	1.00	0.62	0.63
Term Test	0.67	0.59	0.66	1.00	0.83
Final Exam	0.70	0.60	0.67	0.85	1.00
	Spearman Rank Order Correlation Coefficients				
	Online MC	*Tutorials*	*Progress test*	*Term Test*	*Final Exam*

Table 7(b) ECON 105 (Macroeconomics)

	Online MC	*Tutorials*	*Assignment*	*Term Test*	*Final Exam*
		Pearson Correlation Coefficients			
Online MC	1.00	0.73	0.62	0.65	0.76
Tutorials	0.76	1.00	0.54	0.62	0.74
Assignment	0.56	0.55	1.00	0.66	0.68
Term Test	0.62	0.65	0.61	1.00	0.83
Final Exam	0.73	0.75	0.64	0.84	1.00
	Spearman Rank Order Correlation Coefficients				
	Online MC	*Tutorials*	*Assignment*	*Term Test*	*Final Exam*

TABLE 8
Percent of Students and the Change in GPA for 2011-S2 Under 2011-S1 Weighting Schemes (see note below).

	Microeconomics		Macroeconomics	
Change in GPA	**Raw**	**Historical**	**Raw**	**Historical**
+10				
+9				
+8				
+7				
+6	0.6			
+5	0.6			
+4	4.4		0.3	0.3
+3	6.7	0.6	1.4	0.9
+2	14.3	0.6	3.4	3.4
+1	24.4	3.8	12.3	8.0
0	41.3	24.4	47.3	29.8
-1	6.3	18.7	20.3	27.5
-2	1.0	25.7	11.2	22.6
-3	0.3	17.1	3.4	7.2
-4		7.9	0.3	0.3
-5		1.0		
-6				
-7				
-8				
-9				
-10				
Overall Change in GPA	+0.9	-1.5	-0.3	-0.8
Percent of students in range (-1,+1)	72.0	46.9	79.9	65.3
Percent of students in range (-2,+2)	87.3	73.2	94.5	91.3
N	315	315	349	349

Weightings:
Microeconomics: 5% online MC tests, 5% progress test, 10% tutorials, 80% term test.
Macroeconomics: 22% online MC tests, 22% tutorials, 56% term test.

TABLE 9
Percent of Students and the Change in GPA for 2011-S2 using "Reversed" Weighting Schemes (see note below).

	Microeconomics		Macroeconomics	
Change in GPA	**Raw**	**Historical**	**Raw**	**Historical**
+10				
+9				
+8				
+7				
+6	0.6		0.3	0.3
+5	2.5			
+4	5.1		0.6	0.6
+3	11.4		0.3	0.9
+2	17.1	1.0	1.7	3.4
+1	24.1	3.8	6.6	9.5
0	32.7	27.3	35.8	28.7
-1	5.1	14.3	22.6	25.5
-2	1.0	27.6	20.0	22.1
-3	0.3	19.4	7.7	6.6
-4		4.4	4.3	2.6
-5		1.9		
-6		0.3		
-7				
-8				
-9				
-10				
Overall Change in GPA	+1.2	-1.5	-0.9	-0.8
Percent of students in range (-1,+1)	61.9	45.4	65.0	60.7
Percent of students in range (-2,+2)	80	74	86.7	86.2
N	315	315	349	349

This table shows the impact on grades under the alternative assumptions. Here microeconomics replicates the S1 macroeconomics approach and we distribute the weight of the omitted final exam equally over the remaining assessment items (progress test, online MC tests and tutorials).Macroeconomics is calculated using the S1 microeconomics approach where the weight from the omitted final exam is placed entirely on the term test.

APPENDIX 1

True vs. Alternative Grades for Grade Bands – Examples.

These tables are best read across each row. For example, in row 1 of table (a) we see that in the sample period (2005-S1 to 2010-S1) 4.5 percent of students actually received an A+ (the number in the far right column). Under the alternative weighting scheme of (0.2, 0.8) 3.1 percent of students would have continued to receive an A+, 0.9 percent would receive an A, 0.4 percent an A- and 0.1 percent a B+. The column total shows how many students would now receive an A+ being 5.2 percent. Note that in the historically scaled set, row and column totals are constrained to be the same with any minor differences due to rounding.

The disruption to grades with different weighting schemes can also be seen in these tables. Students on the diagonal receive the same alternative grade as their true grade. Compared to the weighting scheme of (0.2, 0.8) the weighting scheme (0.8, 0.2) shows smaller numbers on the diagonal (using either raw or scaled grades) and greater dispersion away from the diagonal.

Percent of Students – Using Raw Scores

Table (a) Weighting: (0.2, 0.8) (n=8752)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	3.1	0.9	0.4	0.1								**4.5**
	A	1.4	1.5	1.3	0.8	0.4	0.1						**5.4**
	A-	0.5	1.6	1.8	1.9	1.0	0.4	0.1	0.1				**7.4**
	B+	0.2	0.6	1.6	2.3	1.7	1.3	0.6	0.1	0.1			**8.4**
	B		0.2	0.7	2.1	2.3	2.5	1.1	0.7	0.2			**9.9**
	B-		0.1	0.3	1.1	2.0	2.8	2.0	1.4	0.5	0.2		**10.5**
	C+			0.1	0.3	1.0	2.2	2.5	2.3	1.5	0.5	0.1	**10.5**
	C				0.1	0.4	1.2	2.1	2.7	2.4	1.3	0.6	**10.7**
	C-				0.1	0.2	0.6	1.3	1.9	2.2	1.6	1.1	**8.9**
	D						0.1	0.2	0.6	1.1	1.5	1.7	**5.3**
	E					0.1		0.1	0.3	1.1	1.9	15.0	**18.5**
Tot.		**5.2**	**4.9**	**6.2**	**8.8**	**9.1**	**11.2**	**10.0**	**10.1**	**9.1**	**7.0**	**18.5**	**100**

Table (b) Weighting: (0.8, 0.2) (n=8752)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	4.3	0.1	0.1									**4.5**
	A	4.4	0.6	0.2	0.1								**5.4**
	A-	5.1	1.3	0.7	0.2								**7.4**
	B+	4.3	1.9	1.2	0.5	0.3	0.1						**8.4**
	B	4.0	2.1	1.7	1.1	0.6	0.2	0.1				0.1	**9.9**
	B-	3.1	2.0	2.3	1.3	1.0	0.5	0.2				0.1	**10.5**
	C+	2.4	1.6	1.9	1.7	1.3	0.7	0.4	0.2			0.1	**10.5**
	C	1.9	1.4	1.7	1.6	1.6	1.1	0.6	0.3	0.2	0.1	0.3	**10.7**
	C-	1.4	1.0	1.2	1.4	1.3	1.1	0.6	0.5	0.1		0.2	**8.9**
	D	0.4	0.5	0.6	0.6	0.6	0.9	0.5	0.4	0.4	0.2	0.4	**5.3**
	E	0.2	0.6	0.5	0.7	0.8	1.1	1.4	1.1	1.4	1.1	9.6	**18.5**
Tot.		**31.5**	**13.1**	**12.1**	**9.2**	**7.5**	**5.7**	**3.8**	**2.5**	**2.1**	**1.4**	**10.8**	**100**

Percent of Students – Using Historical Grade Distribution

Table (c) Weighting: (0.2, 0.8) (n=8752)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	2.8	1.2	0.4	0.1								**4.5**
	A	1.3	1.6	1.3	0.8	0.3	0.1						**5.4**
	A-	0.4	1.7	2.3	1.6	0.9	0.4	0.1					**7.4**
	B+	0.1	0.6	2.0	2.1	2.0	0.9	0.5	0.2				**8.4**
	B		0.1	0.9	2.2	2.6	2.3	1.1	0.6	0.2			**9.9**
	B-		0.1	0.3	1.2	2.2	2.7	2.0	1.4	0.4	0.1		**10.5**
	C+			0.1	0.3	1.3	2.2	2.5	2.5	1.3	0.3	0.1	**10.5**
	C				0.2	0.4	1.2	2.4	2.9	2.3	0.8	0.5	**10.7**
	C-					0.2	0.7	1.4	2.1	2.0	1.3	1.2	**8.9**
	D						0.1	0.3	0.7	1.5	1.2	1.5	**5.3**
	E							0.2	0.5	1.2	1.6	15.0	**18.5**
Tot.		**4.5**	**5.4**	**7.4**	**8.4**	**10.0**	**10.4**	**10.6**	**10.7**	**8.9**	**5.3**	**18.5**	**100**

Table (d) Weighting: (0.8, 0.2) (n=8752)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	2.6	0.9	0.5	0.2	0.1		0.1					**4.5**
	A	1.3	1.3	0.9	0.7	0.5	0.3	0.1	0.1				**5.4**
	A-	0.6	1.5	1.4	1.2	1.1	0.9	0.5	0.2				**7.4**
	B+	0.1	1.0	1.4	1.2	1.4	1.5	1.0	0.4	0.3			**8.4**
	B		0.4	1.4	1.4	1.4	1.7	1.5	1.3	0.5	0.1	0.1	**9.9**
	B-		0.2	0.9	1.1	1.5	1.7	1.8	1.7	1.1	0.3	0.3	**10.5**
	C+			0.4	1.1	1.2	1.5	1.7	2.0	1.3	0.7	0.5	**10.5**
	C			0.1	0.7	1.1	1.3	1.4	2.0	1.9	1.0	1.1	**10.7**
	C-			0.2	0.6	0.8	0.9	1.2	1.5	1.6	1.0	1.1	**8.9**
	D				0.1	0.4	0.3	0.6	0.7	1.0	0.7	1.5	**5.3**
	E				0.1	0.2	0.4	0.6	0.8	1.1	1.5	13.8	**18.5**
Tot.		**4.5**	**5.4**	**7.4**	**8.4**	**10.0**	**10.4**	**10.6**	**10.7**	**8.9**	**5.3**	**18.5**	**100**

APPENDIX 2

True vs. Alternative Grades for Grade Bands 2011-S2 Using 2011-S1 Weightings.

Using Raw Scores

Table (a) Microeconomics (104) 2011-S2 Using 2011-S1 Weights (n=315)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	13.0	0.3	0.3									**13.7**
	A	4.1	2.5	0.6									**7.3**
	A-	2.2	3.2	2.5	0.3								**8.3**
	B+		2.9	2.5	1.0	0.3							**6.7**
	B	0.3	0.3	1.9	2.2	2.9	1.0						**8.6**
	B-		0.6	1.6	0.6	2.9	1.0	1.0					**7.6**
	C+		0.3	0.6	1.3	1.3	2.5	2.2	0.3	0.3			**8.9**
	C				0.6	0.3	0.6	1.3	1.6	1.0		0.3	**5.7**
	C-					0.6	1.6	2.9	2.2	1.6	0.6	0.3	**9.8**
	D						0.6	1.0		2.2	0.6	1.0	**5.4**
	E					0.6	0.3	1.0	0.6	1.9	1.3	12.4	**18.1**
Tot.		**19.6**	**10.1**	**10.0**	**6.0**	**8.9**	**7.6**	**9.4**	**4.7**	**7.0**	**2.5**	**14.0**	**100**

Table (b) Macroeconomics (105) 2011-S2 Using 2011-S1 Weights (n=349)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	12.0	1.7	1.4									**15.2**
	A	3.2	3.7	2.0	0.9								**9.7**
	A-		1.4	3.4	5.2	1.4	0.6						**12.0**
	B+		0.6	0.3	3.7	2.3	2.0		0.3				**9.2**
	B			0.3	1.7	2.6	2.6	0.3	1.4				**8.9**
	B-				0.6	0.6	2.3	1.7	2.0	0.6			**7.7**
	C+					0.3	1.1	1.7	2.0	1.1	0.3		**6.6**
	C					0.3		0.3	0.9	1.1	0.9	0.6	**4.0**
	C-						0.6	0.3	1.1	1.1	1.4	1.1	**5.7**
	D								0.6	1.1	0.9	0.3	**2.9**
	E							0.3	0.6	0.9	1.4	14.9	**18.1**
Tot.		**15.2**	**7.4**	**7.4**	**12.1**	**7.5**	**9.2**	**4.6**	**8.9**	**5.9**	**4.9**	**16.9**	**100**

Using Historical Grade Distribution

Table (c) Microeconomics (104) 2011-S2 Using 2011-S1 Weights (n=315)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	2.5	3.2	3.8	2.2	1.6	0.3						**13.7**
	A			0.6	2.5	1.9	1.6	0.6					**7.3**
	A-				1.0	3.2	1.9	2.2					**8.3**
	B+					0.6	2.9	2.9	0.3				**6.7**
	B					0.3	1.3	2.2	3.8	1.0			**8.6**
	B-						1.0	1.9	2.9	1.3	0.6		**7.6**
	C+						0.6	1.0	1.6	3.8	1.0	0.6	**8.9**
	C								1.0	1.6	1.0	2.2	**5.7**
	C-								1.0	2.2	3.2	3.5	**9.8**
	D									1.3	0.3	3.8	**5.4**
	E								0.6	0.3	1.0	16.2	**18.1**
Tot.		**2.5**	**3.2**	**4.4**	**5.7**	**7.9**	**9.6**	**10.8**	**11.2**	**11.5**	**7.1**	**26.3**	**100**

Table (d) Macroeconomics (105) 2011-S2 Using 2011-S1 Weights (n=349)

		Alternative Grade											
		A+	A	A-	B+	B	B-	C+	C	C-	D	E	**Tot.**
True Grade	A+	4.9	4.3	4.6	1.4								**15.2**
	A	0.6	2.0	3.4	2.9	0.9							**9.7**
	A-			0.9	4.0	5.2	2.0						**12.0**
	B+			0.3	0.3	3.7	3.4	1.1	0.3				**9.2**
	B				0.3	1.7	2.9	2.6	1.4				**8.9**
	B-					0.6	1.1	3.4	2.3	0.3			**7.7**
	C+						0.6	2.6	2.6	0.9			**6.6**
	C						0.3	0.3	1.4	1.4	0.6		**4.0**
	C-							0.9	1.1	2.0	1.4	0.3	**5.7**
	D								0.9	1.7		0.3	**2.9**
	E							0.3	0.9	1.4	2.6	12.9	**18.1**
Tot.		**5.5**	**6.3**	**9.2**	**8.9**	**12.1**	**10.3**	**11.2**	**10.9**	**7.7**	**4.6**	**13.5**	**100**