

Learning by doing: do economics students self-evaluation skills improve?

Jon Guest^a, and Robert Riegler^{b*}

^a*Warwick Business School*

^b*Coventry University*

Jon.Guest@WBS.ac.uk

Warwick Business School

Warwick University

Scarman Road

Coventry

CV4 7AL

United Kingdom

**Corresponding author*

Robert.Riegler@coventry.ac.uk

Faculty of Business and Law

School of Economics, Finance and Accounting

Priory Street

Coventry

CV1 5FB

United Kingdom

Abstract

This paper attempts to (1) measure the students' ability to accurately self-evaluate the quality of their own work, (2) see if this level of accuracy changes when students evaluate a second year essay, having evaluated a similar piece of work in the first year, (3) Investigate whether there is any significant variation in any of the observed changes and (4) identify any factors that might explain any of the observed variation. The data is generated from one cohort of students who were studying for an economics degree at a UK university. The self-evaluation exercise was introduced on two out-of-class essay assessments – one in the first year and one in the second year. Statistical analysis revealed that, on average, the students were significantly more accurate at self-evaluating the quality of their work in the second year than they had been in the first year. However there was considerable variation in this improvement. Those students who demonstrated the greatest improvement were firstly those who were awarded higher marks by the tutor for their second year essay and secondly, those who had been the least accurate at judging the quality of their first year essay. Other student characteristics such as different measures of student ability and gender had no significant impact on the changes in accuracy. However, there is no clear picture about what exactly is driving the improvement.

Keywords: Student Self-Evaluation; Feedback, Independent Learning Skills, Economics Teaching

JEL code: A220

1. Introduction

A key objective of any course in higher education is to develop the independent learning skills of the students enrolled on the programme. In order to meet this objective academic departments need to implement strategies that encourage and enable students to take greater control of their own learning rather than remaining dependent on others to direct and evaluate it for them. In order to become effective independent learners, students need to acquire good self-evaluation skills (Nicol and MacFarlane-Dick, 2006). They must develop the ability to monitor the quality of their own work as they are producing it. As Sadler (1989) argues, they need to understand the standard (i.e. know what a good piece of work looks like), identify any weaknesses in their own work (i.e. recognise gaps between their own and a good piece of work) and be able to take the actions required in order to rectify any perceived weaknesses and close the gap.

In many undergraduate courses it is assumed that these self-evaluation skills will improve as the students act upon the information they obtain from grades and written feedback on their

assessments. In this case one would expect the self-evaluation skills of second year students to be significantly better than those of first year students. First year students may initially find it difficult to adjust to the very different standards required for success in higher education. The speed of adjustment will be influenced by the exact nature of their previous study. However, after a year of producing assessments and receiving grades/feedback, tutors would hope that the majority of students had adapted to the new benchmarks. However, part of the motivation for this study comes from an increasingly common type of e-mail received by both authors from students who appear both surprised and bemused about the low marks they have been awarded for a particular assignment. Their incredulity seems so great that they are convinced that the only explanation for the low mark must be some type of transcription error by the tutor. This suggests that traditional methods of feedback are becoming less effective at improving the students' self-evaluation skills. This may be the result of rapidly rising numbers causing the quality of feedback to fall. Module/course evaluations and questionnaires often find deep dissatisfaction amongst the student body with the quality of the feedback they receive.

This paper attempts to measure the extent to which the self-evaluation skills of students improve as they progress from the first year to the second year of an undergraduate economics degree programme at a UK university. In theory the marginal impact of the grades and the feedback received from first year assessments on the students' perceptions of the standards required in higher education should be relatively large. One might expect the marginal impact from assessments in the second year to be somewhat smaller as students will hopefully have already adjusted to the requirements of the educational environment.

This study attempts to (1) measure the students' ability to accurately self-evaluate the quality of their own work, (2) see if this level of accuracy significantly changes when the same group of students progress from the first year to the second year, (3) investigate whether there is any significant variation in any of the observed changes, and (4) Identify any factors that might help to explain any of the observed variation.

In contrast to much of the existing literature, which uses the assignment mark as the only measure of ability (Cassidy, 2007) we argue that an observation from one piece of coursework might not be a reliable measure. Instead we used a range of different variables to try and capture ability. In addition to the assignment grade these include, UCAS points and the average mark across all the modules studied in the first year.

Our analysis revealed that, on average, the second year students were significantly more accurate at self-evaluating the quality of their work than they had been in the first year. There was also

considerable variation in the extent to which this accuracy improved. We found two key factors that could help explain this variation – the level of inaccuracy in the first year essay and the mark awarded by the tutor for the second year essay. The greater the size of both of these variables, the bigger the improvement in accuracy. Other student characteristics such as different measures of academic ability, gender and nationality had no significant impact on the changes in accuracy.

In section 2 we review the literature, in section 3 we discuss our methodology, and in section 4 we discuss the data and the characteristics of our sample. In section 5 we present our results which focus on changes in the level of self-evaluation accuracy between first and second year students. In section 6 we summarise our results and provide some conclusions.

2. Literature Review

Numerous studies have attempted to measure how accurately students self-evaluate the quality of their academic work. Much of this research has been based on a single piece of assessed work such as an essay, in-class test or examination. One of the earliest and most widely cited surveys of this literature was carried out by Boud and Falchikov (1989). They reviewed over 50 papers that covered a wide range of academic disciplines. Unfortunately none of them involved economics students. The survey reported a number of frequently observed findings in the literature. For instance, studies tend to find that academically weaker students, as judged by the marks awarded by the tutor, are less accurate than their academically stronger counterparts. They also tend to be overconfident about the quality of their work whereas the academically stronger students tended to slightly underrate theirs. This same result has been repeatedly found across a broad range of social and intellectual tasks in the psychological literature (Ehrlinger et al, 2008; Kruger and Dunning, 1999). The experience of the student in their field of study also appears to have a positive impact on accuracy. However the authors found no overall trend for students on average to either under or overestimate the quality of their work. They also concluded that many of the studies were “both conceptually and methodologically unsound”. Limitations with previous research include a lack of any incentives for students to take the activity seriously and potential bias in the data caused by the self-evaluation exercise being used by students as an impression management tool. These issues will be discussed in more detail later in this paper. Grimes (2002) carried out one of the first studies to focus specifically on the self-evaluation skills of economics students. He analysed the ability of 253 students to predict their performance on an in-class midterm examination in a Principles of Macroeconomics module. The findings were similar to those in other disciplines. The academically more able students, as measured by ACT (American College Testing) and GPA (Grade Point Average) scores, tended to be more accurate. Statistical analysis indicated that the students who over

predicted their exam scores were less accurate than those who under predicted their exam scores. Rather than predicted marks on one piece of assessed work, Nowell and Alston (2007) examined data on the difference between predicted and actual grades at a module level. Using data from 715 completed surveys the authors found that 58 per cent of students accurately predicted their grade while one-third were overconfident and just 9 per cent were under confident. In contrast to the results reported by Grimes (2002), Nowell and Alston (2007) found no statistically significant impact of GPA scores on over-confidence.

Although there have been many studies that have collected and analysed data from one piece of work or module there is a far smaller literature that investigates how self-evaluation skills of a given group of students change as they progress through a course. For example, does the accuracy of self-evaluation improve as students complete more assessments and receive feedback? Can completing a number of self-evaluation exercises have a positive effect? Perhaps the largest study undertaken in the impact of repeated self-evaluation was carried out by Lew et al. (2010). The authors analysed data generated from 3,588 first year students at a polytechnic in Singapore. The study is unusual in the type of self-evaluation skills that it seeks to measure. Most studies have focussed on quantifying the ability of students to judge the outcome or product of their learning i.e. the quality of a piece of coursework or an exam answer they have written. However in this study students were asked to make judgements about the quality of the learning activities they had undertaken on a particular day. Rather than the outcome of learning the students' were asked to make judgements about the process of learning.

One advantage of measuring the process rather than the outcome is that it generates far more observations. The process of learning can take place on a daily basis whereas the outcome of learning (i.e. producing an assignment) tends to happen far less regularly - typically only 1-3 times per module per semester. In this study first year students made approximately 80 self-assessment judgements over a 16 week semester. Using correlation coefficient analysis the authors found only a weak relationship between self- evaluation judgements of the learning process and tutor evaluation of the learning process. In order to test the accuracy of the self-evaluation scores over time, the authors averaged and grouped the data into four, four week periods. The correlation coefficients were then compared over these 4 aggregated time periods and the differences were not found to be statistically significant. The authors concluded that the accuracy of the student's self-evaluation of their learning activities did not improve over time.

Whereas the study by Lew et al. (2010) focussed on the learning process, Boud et al. (2013) analysed the impact of repeatedly self-evaluating assignments. They analysed data from 182 students on a

four year undergraduate Design programme. Unfortunately participation in the self-evaluation activity was purely voluntary so there are significant sample selection issues. The study focussed on a number of issues. These included the following: What happens to self-evaluation accuracy as students complete more assessed assignments within a module? What happens to the accuracy of self-evaluation scores as students progress onto further modules in different semesters? What impact does academic ability have on any changes in self- evaluation accuracy? The authors concluded that the accuracy of the students' self-evaluation scores improved significantly within a module once they had completed the first assessment. In particular using a paired t-test they found a statistically significant difference between the self-evaluation estimates and tutor marks on the first assessed task in a module with students tending to overestimate the quality of their work. This difference disappeared in the 2nd, 3rd and 4th assignment where there was no statistically significant difference between the self-evaluation estimate of the students and the tutor mark.

Although the evidence indicates that the accuracy of the students' self-evaluation seems to improve with repeated experience within a module the authors find that it initially deteriorates once they start a new module in the following semester. A statistically significant difference between the tutor mark and self-evaluation estimate in the first assignment in a module in each of the first three semesters was found. Accuracy then improves in each module in each semester as the second and third assessments are completed.

The authors only used one indicator in order to measure the students' academic ability – the marks they obtained in the assignments. Based on this measure the data was split into three groups – high achievement, mid-achievement and low-achievement. The study compared the accuracy of the students' self-evaluation in the first assignment in the semester with the last assignment in the semester in each achievement group. The key finding was that it was only the mid-achievement group that demonstrated improvements in self-evaluation accuracy.

The only study the authors could find that used data from economics students to analyse the change in self-evaluation accuracy over time was that by Ferraro (2010). He analysed the self- evaluation accuracy of 105 students who took three non-cumulative multiple choice tests in an introductory microeconomics module. At the end of the test the students were asked to (1) estimate how many of the questions they had answered correctly and (2) their percentile ranking. They were also provided with some descriptive statistics on the grade distribution after they had been marked. Small monetary incentives ranging from \$1 to \$25 were offered to those students who provided accurate estimates. Results from regression analysis used to control for subject differences found no evidence that the students became more accurate over time.

In a summary numerous studies have found evidence that the least able students tend to be more over-confident and less accurate with their self-evaluation estimates. However issues remain over which measures should be used to capture for academic ability. There are far fewer studies that attempt to measure how the self-evaluation accuracy of a given group of students changes as they progress through a course or module. The evidence that does exist is mixed with some papers finding improvement over time both within and across modules while others, including the only study that focussed on economics students, found no improvement within a module.

3. Methodology

Our data was collected from two pieces of assessed coursework. One was completed by students on a mandatory first year module and the other was completed by the same group students when they had progressed onto a mandatory second year module on an economics undergraduate degree programme.¹ Attempts were made to make sure the two assessments were as similar as possible to increase the potential likelihood that any feedback received on the first piece of coursework would be relevant for the second piece of coursework. They were both 1,500 word essays set on a topic from the microeconomics syllabus. Written feedback was provided in the form of handwritten comments under the following two headings – ‘Key strengths –what you did best’ and ‘key areas to work on to improve your performance in future assessments’. An open response style of assessment was used instead of fixed response multiple choice questions to try and capture a deeper and more multifaceted measure of learning. However, this does introduce a greater level of subjectivity into the marking process. Many studies have found that essay inter-marker reliability is relatively low (e.g., Meadows and Billington, 2005; Newstead, 2004; Bloxham, 2009). In this study only two tutors completed the marking and they both used the same very detailed assessment criteria – see appendix B. Great care was also taken to moderate the work and make sure it was marked in a consistent manner. Coursework was also chosen rather than tests/examinations as it was thought that students would be better able to judge the quality of their work in a less time constrained environment. Added time pressures and stress would probably make it more difficult for students to accurately assess their own performance in examination conditions. Also to the best of our knowledge all previous self-evaluation studies with economics students have only ever used fixed response style assessments in examination conditions.

The self-evaluation data was collected by asking the students the following question which was printed on a separate assessment form:

¹ The modules were Principles of Economics and Intermediate Economics.

“What do you honestly consider would be a fair and appropriate mark for the essay you have handed in?”

In response to this question the students provided a percentage mark and also graded themselves against each of the descriptors in the assessment criteria (e.g., Sauer and Mertens, 2012; Wakefield et al., 2014).

There are a number of key issues that have to be considered when collecting data for the purpose of estimating self-evaluation accuracy (Ward et al., 2002). For example, is submission of the self-evaluation estimates voluntary or is it a compulsory part of the coursework? Are marks deductions applied if the exercise is not completed? If it is a purely voluntary activity then there may be sample selection issues with the data as only the most motivated students might complete the task while the least motivated fail to submit completed forms. In this study the self-evaluation exercise was a ‘gatekeeper’ activity for the coursework. If the students failed to submit a completed form then they received a mark of zero for the essay. Another issue is the extent to which the marks or grades submitted by the student accurately reflect their true beliefs about the quality of their work. Are the students given any incentives to take the self-evaluation activity seriously? Research in the pedagogy literature suggests that students find it a cognitively difficult to understand and apply assessment criteria (Nicol and Macfarlane-Dick, 2006). If the students only perceive there to be limited marginal benefits from carrying out the activity then they may not exert the effort required to carry out the task effectively. Some anecdotal evidence suggests that self-evaluation forms are often completed just before the work is submitted with very little thought being given to the estimates provided. In most previous studies there is little evidence of any incentives being provided for the accurate completion of the task apart from the small monetary payments made in the study by Ferraro (2010). In this research a marks incentive was provided. If the students’ self-evaluation estimate was within three percentage points of the mark awarded by the tutor then three marks were added to this score. For example if a student estimated that an appropriate mark for their essay was 56 per cent and the tutor awards it 55 per cent then they would receive a final mark of 58% i.e. 3 percentage points are added to the mark provided by the tutor. It is also often unclear in previous work whether the potential learning benefits from engaging in the activity were explained to the students. The authors in this study used 30 minutes of lecture time to explain the potential advantages of fully participating with the self-evaluation activity.

Another reason why the grades submitted by the student might not accurately reflect their true beliefs about the quality of their work are concerns that it might influence the person marking their work. For example students who believe they have submitted a low quality piece of coursework may

not wish to communicate these beliefs to the marker. They may fear that honest self-evaluations of low quality work will negatively influence the assessment judgements of the tutor. In these circumstances some of the participants may deliberately overrate their work in the hope of positively influencing the marker. Students who believe that they have produced high quality work may also want to avoid communicating these beliefs to the tutor. In this case there is an aversion to 'showing off' and concerns that the tutor will perceive them as 'big headed'. The participants may then deliberately under report their true beliefs about the quality of their work. A focus group study by Lew and Schmidt (2007) found some evidence to support this argument. A number of the participants in their study explained how they believed that that any self-evaluation data they provided would influence the tutor when they were deciding on a final grade. Systematic biases may therefore exist in a poorly designed study where the self-evaluation process ends up being used by students as an impression management tool.

This problem is most likely to occur in situations when the self-evaluation form is submitted with the piece of coursework. In an attempt to overcome this problem in our study, the students submitted completed self-evaluation forms completely separately from their essay at the School reception. It was also repeatedly emphasised in lectures and seminars that the tutors would not actually see the forms. Instead the data was input by a course administrator who added the 3 point bonus when appropriate. Note that in our analysis we used the coursework mark before any bonuses were added. Written explanations of the bonus scheme and submission arrangements were also provided with the coursework. Those can be found in appendix B. Appendix C contains the self-evaluation form given to students.

4. Data

There were originally 101 first year students in the cohort that were tracked over a 2 year period. Unfortunately we only had complete data on 78 students for a number of reasons - some failed to proceed to the second year, others did not submit a coursework in one of the two years, or were repeating the first year. A decision was made to drop the repeaters from the sample, as it could be argued that they had experienced more opportunities to engage in self-evaluation than their peers. If a student failed the assessment but passed the course they were still included in the data. The

sample of students had the following characteristics - 81% were male; 82% were UK students; 63% had 'A' level² Economics and 45% had 'A' level Maths.

Table 1: Descriptive statistics of the sample (78 students)

Variables	Students	Mean	S.D.	Min.	Max.
Male students	78	0.81	0.4	0	1
A Level Maths	78	0.45	0.5	0	1
A Level Economics	78	0.63	0.49	0	1
British Students	78	0.82	0.39	0	1
Overall 1 st Year Course Mark	78	58.8	8.64	40.1	75.9
1 st Yr Essay Tutor Mark	78	51.3	8.4	34	71
1 st Yr Essay Student Self-Eval.	78	60.6	7.2	41	88
1 st Yr Inaccuracy	78	10.8	8.4	0	32
2 nd Yr Essay Tutor Mark	78	50.7	9.0	19	66
2 nd Yr Essay Student Self-Eval.	78	58.8	5.4	37	70
2 nd Yr Inaccuracy	78	9.2	7.0	0	28

Table 1 compares the average first year and second year tutor mark and students' self-evaluation mark. The average tutor mark is significantly lower than the student self-evaluation estimate for both the first and second year essays, indicating overconfidence.

Our measure of inaccuracy is the absolute value of the difference between the student's self-evaluation estimate and the mark (or grade – see below) awarded by the tutor. By taking the absolute value we do not differentiate between those students who are over-confident and those that are under-confident even though there is some evidence that the two groups differ (Grimes, 2002). Unfortunately, because the sample of under-confident students in our study is very small a thorough analysis is not possible. The average level of inaccuracy fell from 10.8 to 9.2, which is significant at the 10% significance level. This provides some evidence that the students' ability to accurately self-evaluate the quality of their written work does improve when they progress onto the second year of the degree programme.

For the remainder of the analysis the overall mark was converted into a classification³ as assessment judgements by both tutors and students tend to reflect beliefs about grades. Once an expected or

² 'A' level is the term commonly used in England for the General Certificate of Education Advanced Level. They are usually two year courses that are normally studied by 16-18 year olds. Grades in these courses are often used as a basis to determine entry onto courses in higher education.

³ The grading classification used in the UK system of higher education works as follows: Marks of 70% and over are classified as first-class (1st), 60-69% are classified as a second-class upper division (2:1), 50-59% are classified as second class lower division (2:2) and 40-49% are classified as third-class (3rd). A mark below 40% is a fail.

actual grade is chosen a mark tends to be randomly selected from a range of possibilities that are consistent with this grade. For example imagine a situation where a tutor decides that a piece of student work is of a high 2:2 standard. He/she then makes the rather arbitrary decision of whether to award a mark of 57%, 58% or 59%. The student may also believe their work is of a high 2:2 quality but may provide a self- evaluation estimate that is a couple of marks different from that of the tutor. This can be interpreted as just noise in the data rather than real differences in the perception of the quality of the work. This situation could be contrasted to one where the tutor awards 58% while the student provides a self-evaluation estimate of 60%. Once again there is just a difference of a couple of percentage points but in this case it does reflect a difference in beliefs about the grade.

To control for this potential problem we converted the marks into the following 8 grade categories: First-class (70 and above), High 2:1 (65 to 69), Low 2:1 (60 to 64), High 2:2 (55 to 59), Low 2:2 (50 to 54), High Third (45 to 49), Low Third (40 to 44) and Fail (less than 40). These eight grades were then ranked from 1 for a 'fail' through to 8 for a 'first'. One potential problem with this measure is that the category size is not consistent, i.e. the 'first' and the 'fail' category cover a much larger range of marks than the other categories. However, in our data both students who were awarded a first class grade received a mark of 71 per cent while out of the 12 'fail' grades, only four were below 35 per cent. Therefore the impact on the descriptive statistics is relatively minor.

In the first part of the analysis we provide a descriptive analysis of our sample to observe differences in the first and second year self-evaluation accuracy of the students. In the second part we explain in more detail characteristics that might explain variations in the changes of self-evaluation accuracy. Boud et al. (2013) and Ferraro (2010) find some evidence that academic ability is a factor that influences changes in accuracy. Most often the mark awarded by the tutor on a particular assessment is used to capture ability. However, an observation from just one piece of work may be a noisy and unreliable measure. Instead we use a range of different measures including the students' average overall grade in their first year⁴ (see table 1) and the UCAS⁵ points achieved before for entering university.

⁴ This included the marks from the following five modules – Principles of Economics 1, Principles of Economics 2, Mathematics for Economists, Statistics for Economists, and Current Issues in Applied Economics.

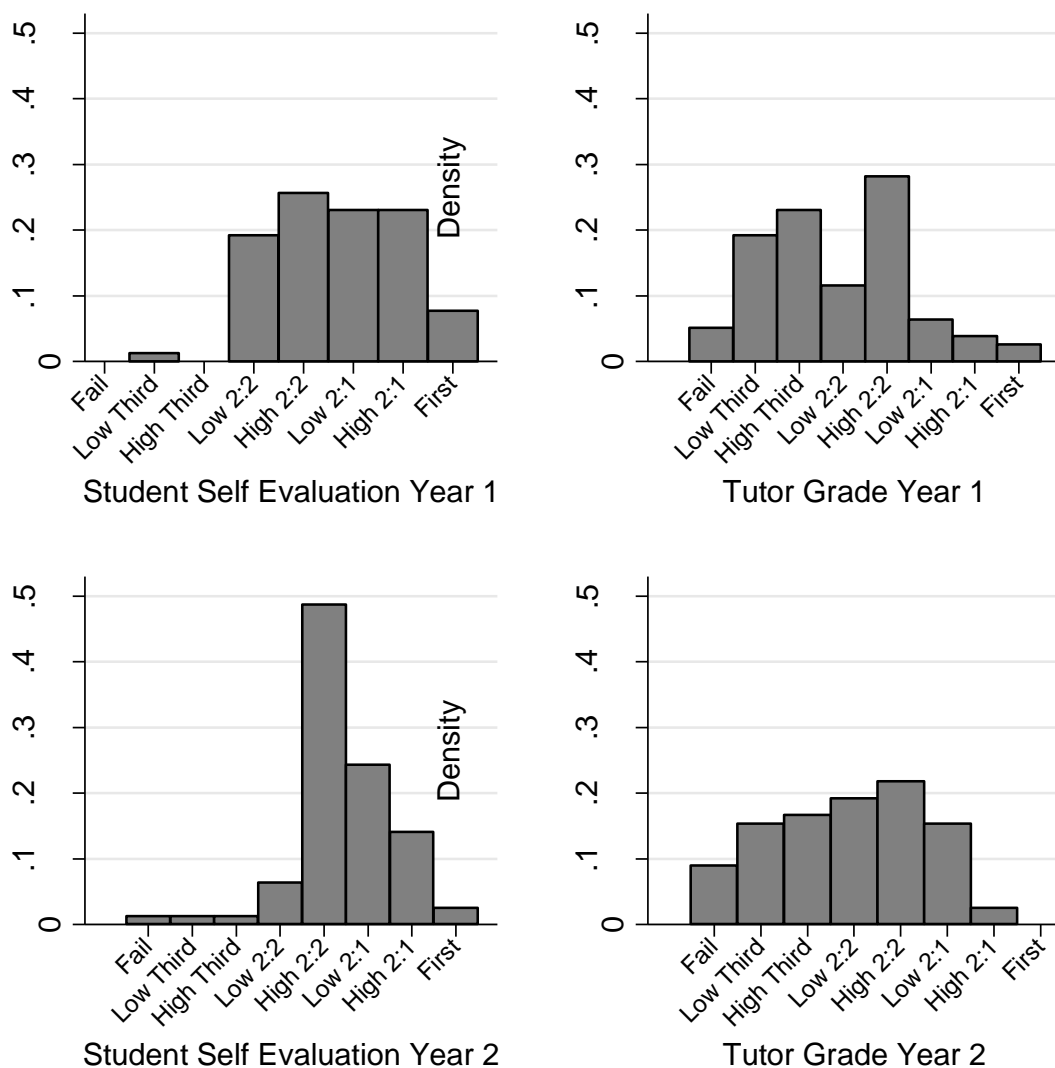
⁵ UCAS is the organisation that manages the application process for universities in the UK. A points system is used that converts grades into points using a tariff table. For example an 'A' grade is worth 120 points while a 'B' grade is worth 100 points.

5. Analysis and Results

5.1 Do the self-evaluation skills of students improve as they progress from the first to the second year?

The analysis in this section focuses on changes in self-evaluation accuracy as the students progress from the first to the second year of their degree programme. Figure 1 shows the distribution of the student self-evaluation grade (left panels) and the grades awarded by the tutor (right panels). The distribution of the second year's self-evaluation grades peaked at a high 2:2 with nearly 50% of the students expecting to receive this grade. However, just 20% were actually awarded this grade by the tutor.

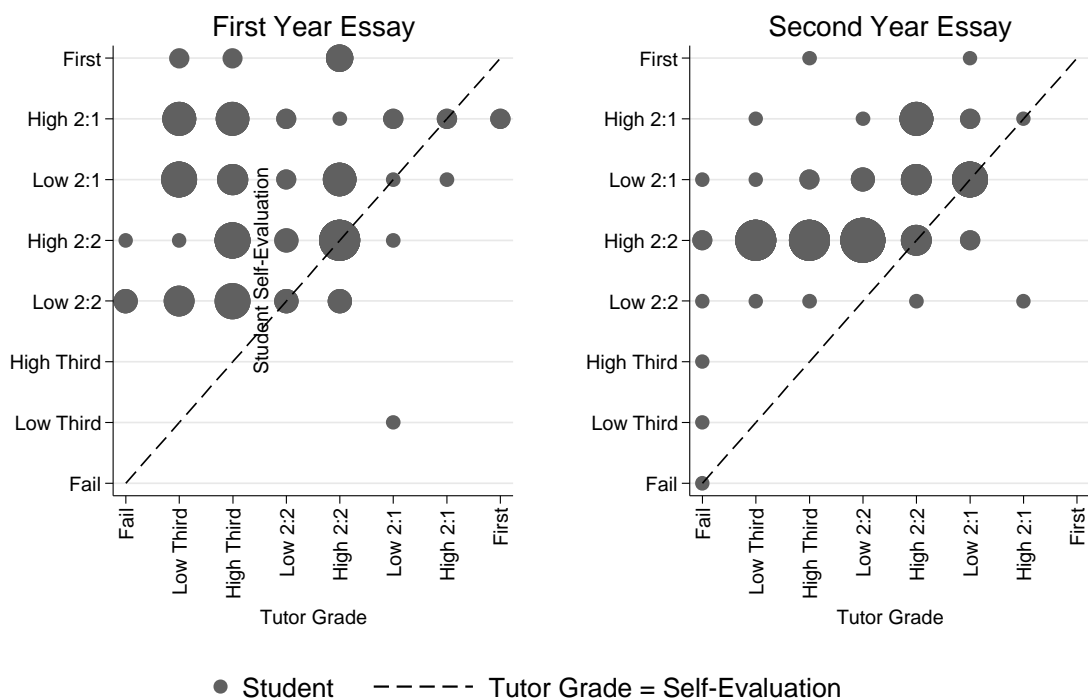
Figure 1: Distribution of student self-evaluation and tutor grade



For both first and second year assessments the distribution of the students' self-evaluation grades are concentrated in higher grade categories indicating over-confidence. There is also less variation in the student's second year evaluation estimates than in the first year

To obtain a clearer picture of the degree of both under and over-confidence the self-assessment estimates were plotted against the tutor grades and are shown in figure 2. The 'x' axis captures the tutor grade while the 'y' axis measures the students' self-evaluation. Any point on the 45 degree line illustrates a situation where the students' self-evaluation estimate is equal to the grade awarded by the tutor. Any point above the 45 degree line illustrates an over-confident student while any point below the 45 degree line illustrates an under-confident student. The larger the circle, in figure 2, the higher is the number of similar self-evaluation – tutor grade combinations. The figure clearly shows that the vast majority of students were over-confident about their performance.

Figure 2: Tutor grade vs. student self-evaluation grade

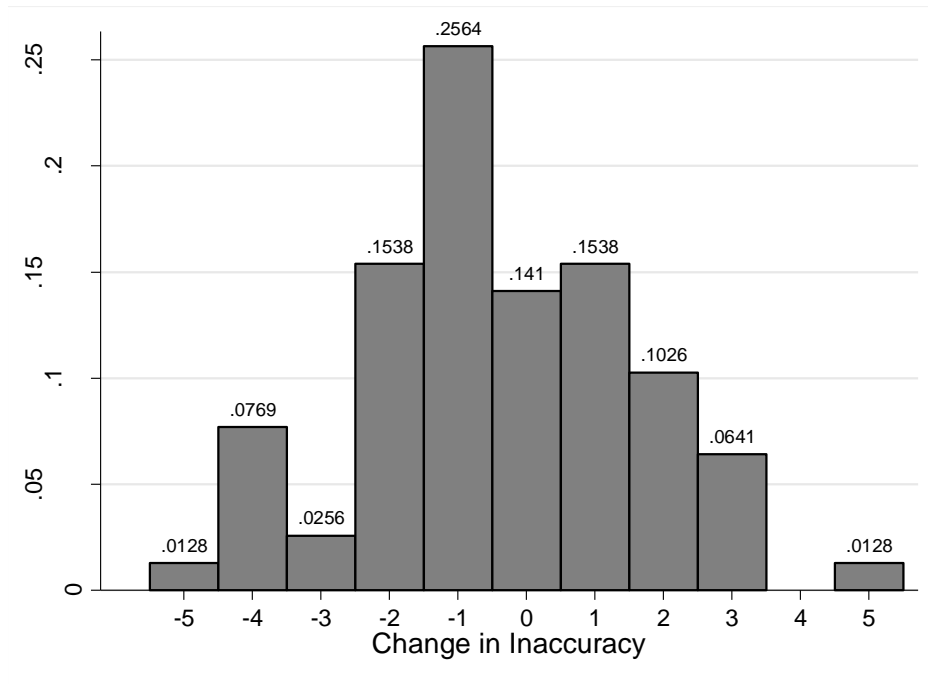


The average level of inaccuracy fell from 2.12 grades to 1.72. The improvement is statistically significant at the 10% significance level⁶ and provides some evidence that the students' ability to accurately self-evaluate the quality of their written work does improve when they proceed onto the second year of the degree programme. This could be evidence that the feedback provided by the tutors in the first year was effective. However, the distinctive bunching of the second year self-evaluation estimates around the high 2:2/low 2:1 grades lead the authors to consider an alternative explanation. Perhaps a significant number of students had no idea why their first year essay was awarded a particular grade. They may have failed to engage with or understood the feedback provided by the tutor. In this case, faced with the marks incentive, they may have simply chosen the most common grade awarded by the tutor in the first year when making their self-evaluation estimates in the second year i.e. a high 2:2. There frequent reporting of the 2:1 grade might be an example of the better than average (BTA) effect reported in the psychology literature. Unfortunately this alternative explanation was only considered after the data had been collected and analysed. We have no information on the extent to which students knew either the average and/or distribution of the grades awarded by the tutors for the first year essay. The tutors neither presented this information in class nor made it available on the virtual learning environment. The one grade students would definitely have known from the first year essay was the one awarded for their own work. Another possibility is that this grade influenced the self-evaluation grade chosen in the second year. This will be discussed in more detail in section 5.2.

Finally, we illustrate the extent to which students' self-evaluation improved between the first and second year by providing a graphical illustration of the change in accuracy in figure 3. A positive value on the x-axis indicates that the gap between the grade awarded by the tutor and the self-evaluation score recorded by the student increased over time. This would illustrate a situation where the ability of a student to accurately self-evaluate the quality of their own work actually deteriorated. A negative value indicates that the student has become more accurate. A value of zero indicates there has been no change in the accuracy of their estimates. Figure 3 illustrates that the self-evaluation accuracy of 41 (52.5%) of the students did get better with 20 (26%) of these improving by one grade. 26 (33%) of the students were actually more inaccurate in the second year than they had been in the first year. The accuracy of 11 (14%) students did not change over the two year period. On average, students improved their accuracy by 0.4 grades.

⁶ Due to the limited numbers of categories, we conducted a t-test using the percentage marks presented in table 1.

Figure 3: Changes in the inaccuracy of self-evaluation estimates



We find some evidence that students improved their self-evaluation accuracy over time. Factors which possibly influence the degree of improvement will be discussed in section 5.2.

5.2 What factors can explain any changes in the accuracy of students' self-evaluation estimates?

Two models are estimated in this section. Firstly, we analyse the factors that determine the level of self-evaluation inaccuracy on the second year essay. Secondly, we investigate what other factors can explain any of the changes in the levels of inaccuracy of students' self-evaluation estimates. Equation 1 and equation 2 will be estimated:

$$Inaccuracy_{it} = f(TA_{it}, Inaccuracy_{it-1}, Ability_i, X_i) \quad (1)$$

$$\Delta Inaccuracy_{it} = f(TA_{it}, Inaccuracy_{it-1}, Ability_i, X_i) \quad (2)$$

The level and the change of student i 's self-evaluation inaccuracy are measured as followed:

$$Inaccuracy_{it} = |SE_{it} - TA_{it}|, \text{ and}$$

$$\Delta Inaccuracy_{it} = (|SE_{it} - TA_{it}|) - (|SE_{it-1} - TA_{it-1}|)$$

where self-evaluation (SE) and tutor assessment (TA) are measured in grades.

Due to the ordinal dependent variable equation 1 and 2 are estimated by ordered Probit. Also because of the low numbers of observations the tail end categories are aggregated. *Inaccuracy* consists of 4 categories (accurate, 1 grade inaccurate, 2 grades inaccurate, 3 or more grades inaccurate). The change in *inaccuracy* has 5 categories (inaccuracy decreases by 2 or more grades, by 1 grade, does not change, increases by 1 grade, and by 2 or more grades).

The selected covariates are the following:

*Inaccuracy*_{*t*-1}: With regards to the level of inaccuracy (i.e. equation 1), we expect that the students who were accurate on their first year essay to also be accurate in the second year. However, with regards to changes (i.e. equation 2), it is expected that the higher the level of inaccuracy in the first year, the more likely it is that there would be a significant improvement in self-evaluation accuracy in the second year i.e. they have the greatest gap to close.

Ability: Evidence from the existing literature suggests that some measures of academic ability have a positive impact on the accuracy of self-evaluation estimates. More able students might be better at understanding and acting upon any of the feedback provided by the tutor. This causes them to become less over –confident and more accurate at judging the quality of their future work. Thus increasing academic ability will result in a faster decline in inaccuracy. However a counter argument is that the more able students are likely to be more accurate in the first year, and so have a smaller gap to close.⁷ Ability is a difficult variable to measure. We use a number of different proxies including – the average mark in the first year modules and Universities and College Admissions Service (UCAS) points. As UCAS points are only available for British students we also use the average overall mark for the first year modules.

*TA*_{*t*}: The grade awarded by the tutor for the second year essay may have a positive impact on self-evaluation accuracy for number of reasons. Firstly, it could be a proxy for academic ability. However, it is questionable whether performance on one piece of work is a reliable measure. Secondly, the positive impact might be a statistical artefact of our data as the vast majority of students within our sample are over-confident. The higher the grade awarded by the tutor the less space there is for an over-confident student to be inaccurate. For example, the maximum level of inaccuracy for an overconfident student who is awarded a low 2:2 by the tutor is four grades. The maximum level of inaccuracy for an overconfident student who is awarded a high 2:1 by the tutor is just one grade. Finally, the tutor mark might capture how much effort was exerted on the essay. This might increase

⁷ We try to control for this by adding an interaction term between ability measures and the level of inaccuracy for the first year essay into equation 2. It was not significant and therefore dropped from the ordered probit regression.

because of a particular interest in the topic. The increased effort level might enable the student to more accurately judge the quality of the work.

X: Several student characteristics are also included in the model. To control for cultural factors we include a dummy for being British. Some previous research has found evidence that female students tend to be less over confident and more accurate than male students so a dummy variable was also included for gender. For the sample of British only students we also added a dummy variable for 'A' Level economics. This variable captures prior study of the subject before entering higher education. Those students with greater previous experience might be expected to improve more. Also a dummy indicating if a student was under-confident was added. This dummy should capture if over- and under-confident students behave differently.

Table 2: Marginal Effects of ordered Probit on self-evaluation inaccuracy in the second year

Dep. Var.: Grade Inacc_t				
	0	+1	+2	≥ +3
<i>Model 1a</i>				
Tutor Grade	0.101***	0.118***	-0.071**	-0.148***
Grade Inacc _{t-1}	0.007	0.008	-0.005	-0.010
Av. First Yr. Mark	0.001	0.001	-0.001	-0.002
Under-confident	0.009	0.011	-0.006	-0.013
Male	-0.008	-0.009	0.006	0.012
British	0.010	0.012	-0.007	-0.015
Observations	78	78	78	78
<i>Model 1b</i>				
Tutor Grade	0.081***	0.094***	-0.056**	-0.119***
Grade Inacc _{t-1}	0.002	0.003	-0.002	-0.003
UCAS points	0.001	0.001	0.000	-0.001
Male	-0.011	-0.013	0.008	0.016
Under-confident	0.015	0.017	-0.010	-0.021
Econ. A-levels	-0.009	-0.010	0.006	0.013
Observations	62	62	62	62

* p<0.05, ** p< 0.01, *** p<0.001

The results of equation 1, are presented in table 2. While Model 1a uses the full sample and the average first year mark as a measure of ability, Model 1b only considers British students and uses UCAS points as a proxy for ability. All values are marginal effects at the mean for students who are accurate (column 0), one grade inaccurate (column +1), two grades inaccurate (column +2) and 3 or more grades inaccurate (column ≥ 3). The results on the level of inaccuracy of Model 1a and Model 1b are similar, independent of the ability measure used. The only variable having a statistically significant impact on inaccuracy is the tutor grade on the second year essay. A better grade on the second year essay leads to a higher probability of being accurate or 1 grade inaccurate, and reduces the probability of being 2 or more grades inaccurate. Even though not statistically significant, the

signs suggest that higher ability, under-confident, female and British students are more likely to be accurate, however, students with A-level economics and who were accurate in the first year are less likely to be accurate in the second year.

Table 3: Marginal Effects of ordered Probit on the change in self-evaluation inaccuracy

Dep. Var.: Δ Grade Inacc.					
<i>Model 2a</i>	≤ -2	-1	0	$+1$	≥ 2
Tutor Grade	0.0738***	0.1250***	-0.0695**	-0.1045***	-0.0248*
Grade Inacc _{t-1}	0.1769***	0.2999***	-0.1667**	-0.2507***	-0.0594*
Av. First Yr. Mark	0.0015	0.0025	-0.0014	-0.0021	-0.0005
Under-confident	0.0157	0.0266	-0.0148	-0.0222	-0.0053
Male	-0.0268	-0.0454	0.0253	0.038	0.009
British	0.0446	0.0755	-0.042	-0.0631	-0.015
Observations	78	78	78	78	78
<i>Model 2b</i>	≤ -2	-1	0	$+1$	≥ 2
Tutor Grade	0.0660**	0.1050**	-0.0572**	-0.0862***	-0.0276*
Grade Inacc _{t-1}	0.1816***	0.2890***	-0.1573**	-0.2372***	-0.0761*
UCAS points	0.0002	0.0003	-0.0001	-0.0002	-0.0001
Male	-0.029	-0.0461	0.0251	0.0379	0.0121
Under-confident	0.0266	0.0423	-0.023	-0.0347	-0.0111
Econ. A-levels	-0.0423	-0.0673	0.0366	0.0552	0.0177
Observations	62	62	62	62	62

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 illustrates the marginal effects at the mean for equation 2. Model 2a uses the full sample with the overall average first year mark used as an ability measure while model 2b includes only British students and uses UCAS points as an ability measure. The first two columns present the marginal effects for students whose inaccuracy fell by 1 and 2 or more grades. The third column captures students whose level of inaccuracy remained unchanged while the last two columns capture students who became more inaccurate.

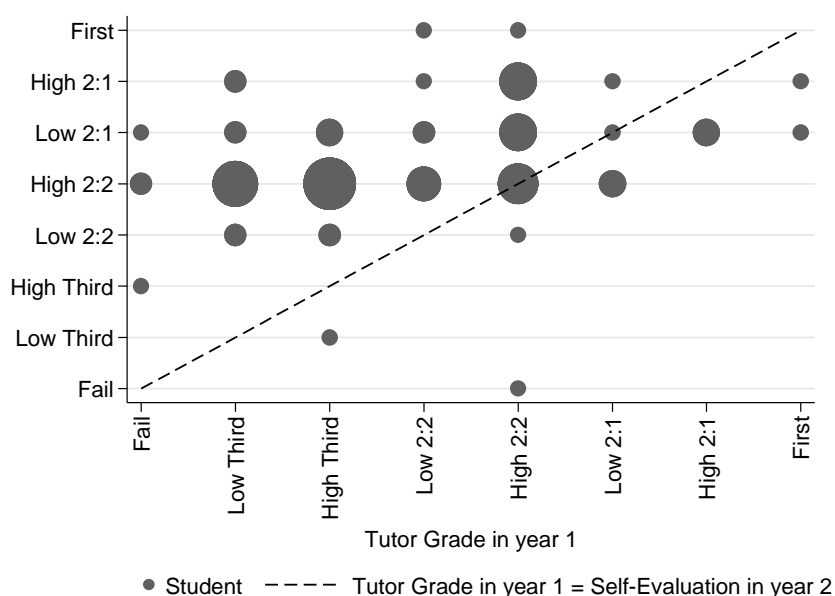
The only significant marginal effects are on the coefficients for the second year tutor grade and self-evaluation inaccuracy in the first year. Students who do well in their second year essay are more likely to improve their accuracy. For example, a one unit increase in the tutor grade leads to a 7% increase in the probability of a 2 or more grade improvement and a 13% increase in the probability of a 1 grade improvement. As expected, there is a strong and large significant relationship between inaccuracy of students' self-evaluation for the first year essay and the changes in accuracy. The higher the self-evaluation inaccuracy for the first year essay the higher is the probability of improving.

Summarising the main results, the assessment judgement of the tutor for the second year essay is the main factor that explains both the level and changes of accuracy in the second year. However

our more comprehensive measures of academic ability, such as the overall first year course mark, are not significant. Furthermore, the inaccuracy of the self-evaluation estimates for the first year essay is another factor that explains the changes in accuracy. Student characteristics such as gender, fee status and prior study of economics do not seem to have a significant impact.

In the next stage of the research we wanted to analyse if the self-evaluation estimates of the students were influenced by other factors. For example, do students base their self-evaluation estimate in the second year on the mark they were awarded by the tutor in the first year? This might occur for a couple of reasons: Firstly, the student believes that their second year essay is of a similar quality to their first year essay and the tutor mark from the first year provides an accurate indication of the appropriate standard. Alternatively the student may have very little idea of what constitutes a good piece of work and simply uses the grade they received from the tutor in the first year because they lack any other information or relevant knowledge. Figure 1 revealed that the mode tutor grade in year one was a high 2:2 (17 students), but more than 60% of students received a grade below this level. The mode of students' self-evaluation grade in the second year was also a high 2:2 (38 students), but this time only 10% of students expected a grade of low 2:2 or lower. Figure 4 links the students' self-evaluation grade for the second year essay on the y-axis and the mark awarded by the tutor for the first year essay on the x-axis. If the self-evaluation grade in year 2 is equal to the tutor grade in year 1 then the highest density of scatter points should be on the 45 degree line.

Figure 4: The Impact of the tutor grade on the 1st year essay on the self-evaluation on the 2nd year essay



With the exception of eight students, we find that the vast majority of students had a self-evaluation grade that was different to the previous year's tutor grade. Of those students that had a different self-evaluation grade, most of them remained overconfident and expected to receive a higher grade for their second year essay than they achieved in the first year essay.⁸ While the median tutor grade was a low 2:2, the median student self-evaluation was still a high 2:2. If students are influenced by the grade received in the previous year they appear to be anchoring but with a certain level of optimism. As previously discussed we have no data to test whether the students' self-evaluation estimate is influenced by the modal awarded by the tutor in the first year.

5.3 Robustness checks

Transferring the percentage marks into grade categories could lead to loss of information. Therefore Equation 1 and 2 were also estimated with an OLS estimator using percentage marks as a measure for inaccuracy. The results are presented in Table A1 in appendix A and support the main results. The higher the tutor mark, the lower the level of inaccuracy in the second year and the larger is the decrease in inaccuracy from the first year. An increase in the level to inaccuracy in the first year leads to a significantly higher decreases in inaccuracy, but does not have a significant impact on the level of inaccuracy in the second year. Besides those two variables, all other variables are insignificant. The results are robust using different ability measures.

Out of 101 students 78 students progressed onto the second year. The other 23 students, who dropped out of the course, had (i) significantly lower tutor marks (ii) significantly lower self-evaluation estimates, but were not significantly more inaccurate. To control for any potential selection bias, a Heckman model was employed. Even though there was evidence of sample selection, the results were very similar to those obtained using OLS without sample selection correction.

6. Conclusion

This paper aimed to analyse changes in the ability of students to accurately self-evaluate the quality of a first and second year essay in microeconomics. The sample consisted of 78 students studying for an economics degree at a UK university. The study was carefully designed to address some of the limitations with previous research in this area i.e. selection effects, time constraints, lack of incentives, impression management bias and marker reliability. We find some variation in the data but on average students self-evaluation estimates become more accurate in the second year. The

⁸ The tutor grade of the first year is also statistically different to the self-evaluation grade in the second year.

most important factor that determines this improvement is the mark awarded by the tutor for the second year essay. This may be a noisy measure of student ability but other proxies, such as the average overall mark in the first year modules, do not appear to have a significant effect. Another important factor that explains the improvement is the level of inaccuracy of the self-evaluation estimate in the first year. Those students who have the greatest gap to close, between their own and the tutor's perceptions of their academic work appear to do so. Our results are not consistent with those reported in the only other study that analysed changes in the self-evaluation accuracy of economics students. That paper differed from ours as it analysed data collected from three multiple choice tests within a single module whereas this study is based on an out-of-class essay submitted in two different academic years.

We cannot provide a conclusive answer as to why the accuracy of the self-evaluation estimates does improve. It is possible that the students' self-evaluation skills remain constant. The improvement in observed accuracy could simply be caused by the students anchoring to the modal mark awarded by the tutor in the first year. The increased bunching of the second year self-evaluation estimates around high 2:2/low 2:1 grades indicate that this might be a possibility. We currently have no way of testing this hypothesis as no data was collected on the students' awareness or knowledge of the grade distributions from the previous year's work. This could be one avenue for future research as it is not necessarily the case that increasing self-evaluation accuracy is solely determined by improvements in the self-evaluation skills of the students. What we could test for was the possibility that the students simply used their first year tutor mark as their second year self-evaluation estimate. There was no evidence to support this hypothesis. We also have no way of knowing how effective the students found the feedback provided by the tutor. The impact of different types of feedback on self-evaluation accuracy might also provide an interesting topic for future research.

The design of the study would have been improved if both written assessments had occurred within the same module but unfortunately the assessment strategy adopted by the university prevented us from doing this. It is possible that students experienced problems in adjusting to the higher standard required in written coursework in the second level of the course. Evidence from Boud et al. (2013) suggests that self-evaluation accuracy initially falls before improving when students start a new module.

Another direction for future research is to investigate the possibility that students have a better understanding of the standards required in Higher Education than their self-evaluation estimates might suggest. They may find it particularly difficult to objectively evaluate their own work. If this

were true we might expect them to be better able to accurately judge the quality of work produced by their peers.

Acknowledgements

We would like to express our gratitude to the Editorial Coordinator and the two referees of the IREE. Their careful reading and useful comments have helped to improve the quality of the paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Bloxham, S., 2009. Marking and moderation in the UK: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education*, 34(2), pp. 209-220.

Boud, D. and Falchikov, N., 1989. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), pp. 529-549.

Boud, D., Lawson, R. and Thompson, D. G., 2013. Does student engagement in self-assessment calibrate their judgement over time?. *Assessment & Evaluation in Higher Education*, 38(8), pp.941-956.

Cassidy, S., 2007. Assessing 'inexperienced' students' ability to self-assess: exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education*, 32(3), pp. 313–330.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. and Kruger, J., 2008. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behaviour and Human Decision Processes*, 105(1), pp. 98-121.

Ferraro, P.J., 2010. Know thyself: competence and self-awareness. *Atlantic Economic Journal*, 38(2) pp.183-196.

Grimes, P.W., 2002. The overconfident principles of the economics student: An examination of metacognitive skill. *Journal of Economic Education*, 33(1), pp. 15-30.

Kruger, J. and Dunning, D., 1999. Unskilled and unaware of it: How difficulties in recognising one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), pp. 1121-1134.

Lew, M.D., Alwis, W.A.M. and Schmidt, H.G., 2010. Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, 35(2), pp. 135-156.

Lew, M.D. and Schmidt, H.G., 2007. *Measuring students' beliefs about self-assessment*. Rotterdam; Erasmus University.

Meadows, M. and Billington, L., 2005. *A review of the literature on marking reliability*, Commissioned report to the National Assessment Agency, London: AQA.

Newstead, S., 2004. Time to make our mark. *The Psychologist*, 17(1), pp. 20-23.

Nicol, D.J. and Macfarlane-Dick, D., 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), pp. 199–218.

Nowell, C. and Alston, R., 2007. I thought I got an A! Overconfidence across the Economics Curriculum. *Journal of Economic Education*, 38(2), pp. 131-142.

Sadler, R.D., 1989. Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), pp. 119-144.

Sauer, K. M. and Mertens, W.G., 2012. The Test Assessment Questionnaire: a tool for student self-assessment after the midterm exam. *Journal of Economics and Economic Research*, 14(2), pp. 93-100.

Wakefield, C., Adie, J., Pitt, E. and Owens, T., 2014. Feeding forward from summative assessment: the Essay Feedback Checklist as a learning tool. *Assessment & Evaluation in Higher Education*, 39(2), pp. 253-262.

Ward, M., Gruppen, L. and Regehr, G., 2002. Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1), pp. 63-80.

Appendix A: More regression results

Table A.1: OLS regression results using percentage marks

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var:	Mark Inacc.	Mark Inacc.	Mark Inacc.	Δ Mark Inacc.	Δ Mark Inacc.	Δ Mark Inacc.
Tutor Mark _t	-0.578***	-0.508***	-0.501***	-0.576***	-0.506***	-0.489***
Mark Inacc _{t-1}	-0.059	-0.048	-0.042	-1.246*	-2.118***	-1.136***
Av. First Yr. Mark	-0.030			-0.065		
Mark Inacc _{t-1} × Av. First Yr. Mark				0.003		
UCAS Points		-0.009			-0.037	
Mark Inacc _{t-1} × UCAS Points					0.004	
Maths A Levels			-1.407			-3.219
Mark Inacc _{t-1} × Maths A Levels						0.179
Underconfident	0.930	0.310	0.334	0.962	0.741	-0.038
Male	-0.780	-0.940	-1.148	-0.693	-0.643	-1.042
British	-1.6481			-1.6732		
Econ. A Levels		1.093	0.441		0.731	0.586
Constant	42.760***	37.462***	36.072***	44.685***	45.666***	36.297***
Observations	78	62	62	78	62	62
R ²	0.58	0.51	0.52	0.82	0.80	0.79

Legend: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Appendix B: Assessment Criteria Explanations

The following information was given to students before they assessed their own coursework.

Your essay will be marked against each of the following five criteria (underlined and in bold below). Students often complain that they do not fully understand the exact meaning of statements included in the assessment criteria. To help overcome this problem a number of questions have been included below each statement. You should consider these questions when completing your essay and when carrying out the peer review process. They will help you judge the extent to which your own work and the other student's work you are reviewing meet each of these criteria. Hopefully this will enable you to achieve a higher mark. You will need to complete a self-evaluation exercise as part of the assessment for this module. **Your work will not be marked unless you have completed and submitted the self-evaluation exercise to reception.**

As an incentive to take the self-evaluation activity seriously, bonus marks will be available. The scheme will operate as follows. If your overall self-evaluation mark is within three percentage points either above or below the final mark awarded by your tutor then you will be awarded a 3 percentage point bonus. The three percentage points are added to the mark awarded by the tutor – not the mark you have awarded yourself. Some students appear to have been confused about this last year. Some examples below illustrate how the scheme operates.

- If you judge your own work as worth **59%** and the tutor awards a mark of **56%** you will receive a final mark of **59%** i.e. the three mark bonus is added to the mark awarded by the tutor not the mark you awarded yourself.

- If you judge your own work as worth **58%** and the tutor awards a mark of **56%** you will receive a final mark of **59%** i.e. the three mark bonus is added to the mark awarded by the tutor not the mark you awarded yourself.
- If you judge your own work as worth **49%** and the tutor awards a mark of **45%** you will receive a final mark of **45%** i.e. you do not receive a bonus as your estimated mark is more than 3 percentage points different from the mark awarded by the tutor.

A number of students last year thought there was some type of optimal strategy in order to game the system and maximise their marks. If you fully understand how the scheme works you will appreciate that the only incentive you have is to be accurate.

Concerns were also expressed last year that the self-evaluation marks provided by the students might influence the assessment judgments of tutor. Please note that the tutors marking your microeconomics essay will not see the self-evaluations sheets you have completed. When they mark your essay they will have no idea what you thought it was worth.

Quality of the introduction and conclusion.

- To what extent
 - o Does the opening paragraph clearly state the key issues that are relevant to the question?
 - o Does the opening paragraph outline how the essay will answer the question?
 - o Does the final paragraph draw together the various points/arguments made in the essay and demonstrate how the question has been answered?
 - o Does the essay come to an abrupt end?

Logical development of ideas and overall quality of the writing.

- To what extent:
 - o Are arguments and explanations in your essay easy to follow? Are the sentences in the essay well-constructed?
 - o Do arguments/points made in the essay follow on from one another in a logical manner?
 - o Do some arguments in the essay seem unrelated to previous points that have been made? ☹
 - o Have all arguments/points been explained in sufficient detail in each paragraph?
 - o Are arguments incomplete and not developed in enough detail? ☹
 - o Is there evidence that the work has been proof read and revised from an initial draft?
 - o Is it written in a way that gives the reader the impression that the material has been fully understood by the author?
 - o Is it written in a manner that suggests that the student does not fully understand the material and has simply re-written pieces of text from various sources? ☹

Relevance and application of economic theory and other material included in the essay.

- To what extent:
 - o Has relevant microeconomic theory been chosen and related to the question?

- Where appropriate, have clearly labelled diagrams and or maths been used to help illustrate and apply relevant economic theory to the question?
 - Have diagrams been included but not fully explained and applied to the question? ☹
 - Does the labelling on diagrams relate specifically to the question?
- **Depth of analysis**
- To what extent:
- Does the essay demonstrate a good understanding of the relevant microeconomic theory?
 - Does the essay include some discussion of any of the possible limitations of the microeconomic analysis included?
 - Has it evaluated the extent to which microeconomic theory can explain the phenomena referred to in the question?
 - Is there evidence of wider reading and research on the topic area?
 - Have the lecturer's notes been simply reproduced/ regurgitated? ☹
 - Is there evidence that the academic sources provided by the lecturer have been read and used to answer the question?
 - Has material from different academic sources been effectively integrated?
- **Presentation and layout**
- Do any sentences in the essay last for an entire paragraph! ☹
 - Is there correct punctuation, grammar and spelling throughout the whole essay?
 - Is the bibliography and in text referencing in the appropriate Harvard format?

Appendix C: Self Evaluation Sheet

The same evaluation sheet was given out to students for their first and second year microeconomics essay.

Assessed Microeconomics Essay - Self Evaluation Form

Name:

How do you rate your answer on each of the following criteria – please tick the appropriate box for each of the five criteria below.

	<u>Fail</u> <40%	<u>Third</u> 40 - 49%	<u>Low 2:2</u> 50 - 54%	<u>High 2:2</u> 55 - 59%	<u>Low 2:1</u> 60 – 64%	<u>High 2:1</u> 65 - 69%	<u>First</u> > 70%
Quality of the introduction and conclusion							
Logical development of ideas and overall quality of the writing							
Relevance and application of microeconomic theory and other material included in the answer							
Depth of analysis							
Presentation and lay-out							

What do you honestly consider would be a fair and appropriate mark for the essay you have written?

What percentile do you think you'll be in for the essay you have written?

PTO

What do you think was the most important thing you learned while researching and writing this essay?

What do you think you did best?

What do you think you did least well?

What did you find the hardest part of writing and researching this essay?
