

# Assessment timing: student preferences and its impact on performance

Richard McManus
Canterbury Christ Church University

Developments in Economics Education conference 2015
University of Birmingham

$10^{th}$ & $11^{th}$ September 2015

# General idea and research questions



## General idea

To give students the option of when to take their first assessment in a first year core Business School module (Economics) to identify preferences and their impact on performance.

## Research questions

1. Do students have preferences on when they take their assessments?
2. If so, what explains these preferences?
3. Is there an impact between the timing of and performance in assessments?
4. Are there pedagogic advantages to allowing flexibility in assessment timing?

# General idea and research questions



## General idea

To give students the option of when to take their first assessment in a first year core Business School module (Economics) to identify preferences and their impact on performance.

## Research questions

1. Do students have preferences on when they take their assessments?
2. If so, what explains these preferences?
3. Is there an impact between the timing of and performance in assessments?
4. Are there pedagogic advantages to allowing flexibility in assessment timing?

# Content

# Related literature



## Related literature

- Students have preferences on assessment question types (Zoller & Ben-Chaim 1989, Birenbaum 2007, Iannone & Simpson 2014) however, no research (to the best of our knowledge) is conducted on students' assessment *timing* preferences.
- Literature demonstrates that frequent and spaced tests, as opposed to more study, improve student outcomes (Carpenter & DeLosh 2005).
- Ariely & Wertenbroch (2002) show that given the choice of timing, some students will procrastinate and these students will have weaker attainment.

# Related literature



## Related literature

- Students have preferences on assessment question types (Zoller & Ben-Chaim 1989, Birenbaum 2007, Iannone & Simpson 2014) however, no research (to the best of our knowledge) is conducted on students' assessment *timing* preferences.
- Literature demonstrates that frequent and spaced tests, as opposed to more study, improve student outcomes (Carpenter & DeLosh 2005).
- Ariely & Wertenbroch (2002) show that given the choice of timing, some students will procrastinate and these students will have weaker attainment.

# Related literature



## Related literature

- Students have preferences on assessment question types (Zoller & Ben-Chaim 1989, Birenbaum 2007, Iannone & Simpson 2014) however, no research (to the best of our knowledge) is conducted on students' assessment *timing* preferences.
- Literature demonstrates that frequent and spaced tests, as opposed to more study, improve student outcomes (Carpenter & DeLosh 2005).
- Ariely & Wertenbroch (2002) show that given the choice of timing, some students will procrastinate and these students will have weaker attainment.

# Related literature: inspiration



Ariely and Wertenbroch (2001) - three assessments in a one 'semester' module; three student cohorts:

- One group imposed deadlines evenly spaced throughout the teaching ('Group A');
- One group given the final day as a deadline for all assessments ('Group B');
- A final group got to self impose deadlines, but these were binding with penalties ('Group C').

Results

- Performance: Group A > Group C > Group B;
- 'Non-rational' self imposed deadlines in Group B;
- Those who did self impose deadlines performed better.

# Related literature: inspiration



Ariely and Wertenbroch (2001) - three assessments in a one 'semester' module; three student cohorts:

- One group imposed deadlines evenly spaced throughout the teaching ('Group A');
- One group given the final day as a deadline for all assessments ('Group B');
- A final group got to self impose deadlines, but these were binding with penalties ('Group C').

Results

- Performance: Group A $>$ Group C $>$ Group B;
- 'Non-rational' self imposed deadlines in Group B;
- Those who did self impose deadlines performed better.



Section 2

# Methodology: experiment design

# Methodology



Students offered the choice of whether to take the first test (based on the material from the first term) either during the second or penultimate week of the second term (8 weeks separation):

- Worth 20% of final mark;
- Numerical questions (11% - 33%); graph drawing based question (22%); short answer question (22% - 44%); and MCQs (22%);
- Students informed of the style and structure before;
- All tests unique through randomising questions (no 'information advantage');
- Decisions were binding (reveal true preference);

# Methodology



Students offered the choice of whether to take the first test (based on the material from the first term) either during the second or penultimate week of the second term (8 weeks separation):

- Worth 20% of final mark;
- Numerical questions (11% - 33%); graph drawing based question (22%); short answer question (22% - 44%); and MCQs (22%);
- Students informed of the style and structure before;
- All tests unique through randomising questions (no 'information advantage');
- Decisions were binding (reveal true preference);

# Methodology



Students offered the choice of whether to take the first test (based on the material from the first term) either during the second or penultimate week of the second term (8 weeks separation):

- Students given the choice of no option:
  - Might not have a preference;
  - If so, randomly allocated with 50% probability of each date;
  - Allows for a controlled experiment.
- Choices communicated through Blackboard (or verbally or by email);
- Date of second test communicated prior to choice - to be taken after the Easter break, 8 weeks after the latter date of the first test.

# Methodology



Students offered the choice of whether to take the first test (based on the material from the first term) either during the second or penultimate week of the second term (8 weeks separation):

- Students given the choice of no option:
  - Might not have a preference;
  - If so, randomly allocated with 50% probability of each date;
  - Allows for a controlled experiment.
- Choices communicated through Blackboard (or verbally or by email);
- Date of second test communicated prior to choice - to be taken after the Easter break, 8 weeks after the latter date of the first test.

# Methodology

Figure : Research methodology

| | | |
|---|---|---|
| Students | The option of choosing the timing of the first assessment in 'Economics' (a first year undergraduate module) was offered to students. ⇨ | Students take the test at their chosen time. |
| Results | Preferences over assessment timing can be revealed. | Observe the impact on attainment |
| Two timing options were offered: either the second or penultimate week of the second term of teaching (a time separation of 8 weeks), which will be referred to as 'Early' and 'Late' respectively. | | |



Section 3

# Revealed preferences

# Quick survey



## Quick survey

With limited knowledge of the students and module in question.... what do you expect the split to be between:

- Those who chose to take the test early;
- Those who chose to take it late;
- Those who have no preference?

# Student preferences



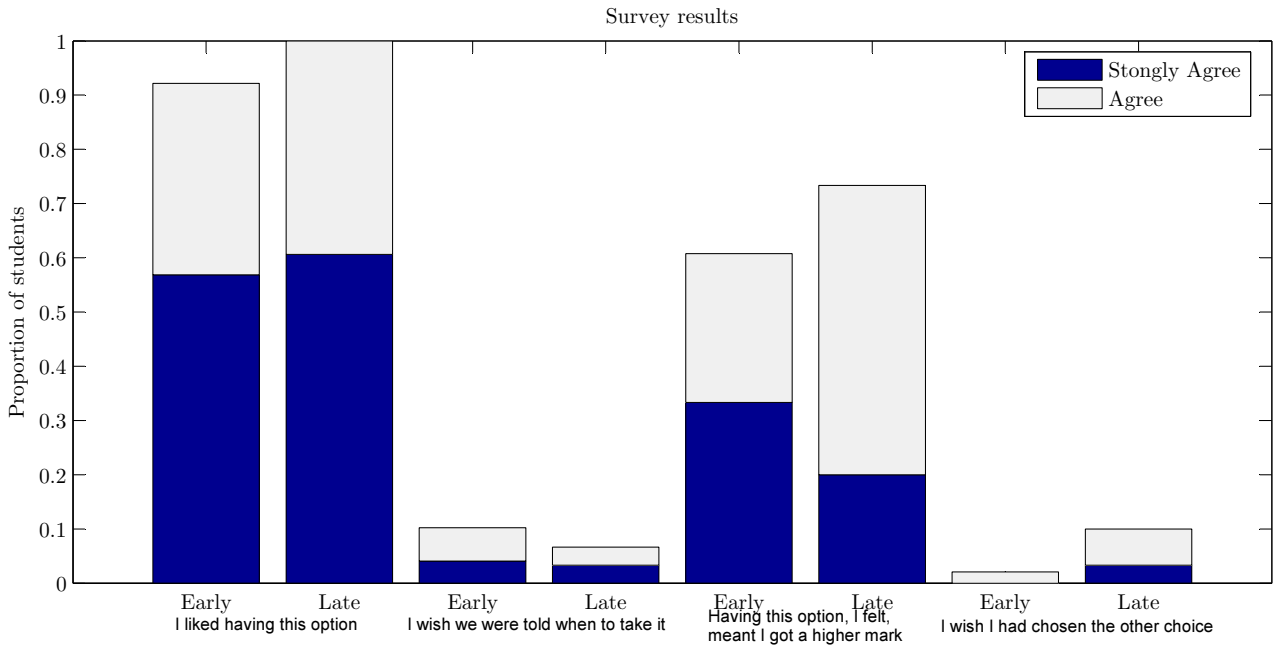
Table : Student choices

| | (1) | (2) |
|---|---|---|
| Early | 136 | 52% |
| Late | 110 | 42% |
| No preference | 4 | 2% |
| No decision | 13 | 5% |

The first column represents absolute numbers, and the second column represents these are percentage of the total.

Peers' predictions

Figure : Survey results on the option of assessment timing




Results from an end-of-year survey conducted by 88 students answering questions on a 'Likert' scale, with respect to having the choice of when to sit the first assessment. Analysis separate between those students who took the 'Early' and 'Late' sitting of the test respectively, and the question asked can be found along the x-axis.

# Who chose when?



Students who took the earlier time of the test (intuitively) tended to be more engaged and comfortable with the module content:

- Prior post compulsory Economics education: 86% early (0.002);
- Prior post compulsory Mathematics education: 76% (0.007);
- Attendance 79% early versus 63% late (0.000);
- Blackboard clicks nearly 50% (0.000) higher;
- A-Levels: 61% early (0.056);
- BME: 35% early (0.000).

# Who chose when?



Students who took the earlier time of the test (intuitively) tended to be more engaged and comfortable with the module content:

- Prior post compulsory Economics education: 86% early (0.002);
- Prior post compulsory Mathematics education: 76% (0.007);
- Attendance 79% early versus 63% late (0.000);
- Blackboard clicks nearly 50% (0.000) higher;
- A-Levels: 61% early (0.056);
- BME: 35% early (0.000).

# Who chose when?

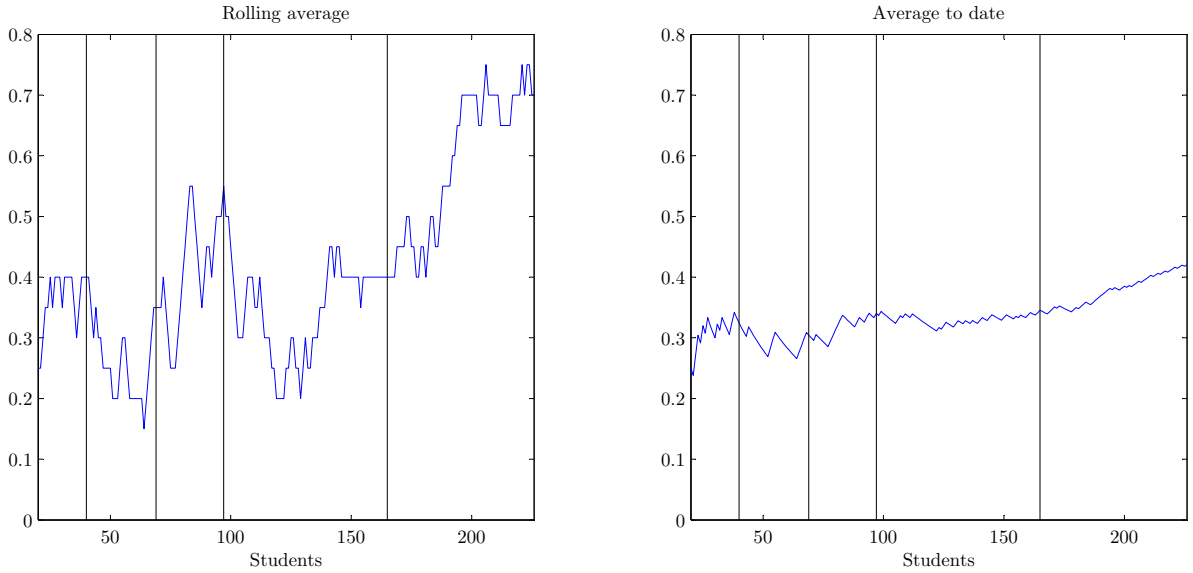


Students who took the earlier time of the test (intuitively) tended to be more engaged and comfortable with the module content:

- Prior post compulsory Economics education: 86% early (0.002);
- Prior post compulsory Mathematics education: 76% (0.007);
- Attendance 79% early versus 63% late (0.000);
- Blackboard clicks nearly 50% (0.000) higher;
- A-Levels: 61% early (0.056);
- BME: 35% early (0.000).

# Time of decision

Figure : Choices made over time




Results obtained in the left hand pane by taking a rolling average of the last 20 decisions where choosing to take the test early was assigned zero, and taking the test later assigned 1, and in the right hand pane through taking an accumulative average of these results. Each vertical line represents a week in time and demonstrates how many decisions were made each week.

# Logit regression results



Table : Logit regression predicting student choices

| | (1) | (2) |
|---|---|---|
| Attendance | 0.235*** | (0.000) |
| VLE | 2.232** | (0.022) |
| Economics | 1.233* | (0.072) |
| Mathematics | 0.945* | (0.053) |
| Ethnic minority | -1.000*** | (0.005) |
| | | |
| Psuedo-$R^2$ | 0.170 | n = 239 |

Results obtained from a Logit regression where the dependent variable takes the value 1 if the student chose the early test, and zero chose the late one. The values in the first column represent coefficients and in the second column (and in parenthesis) p-values of individual significance: a standard star convention is applied; *** signifies that the difference is statistically significant to 1% confidence, ** to 5% significance and * to 10%.

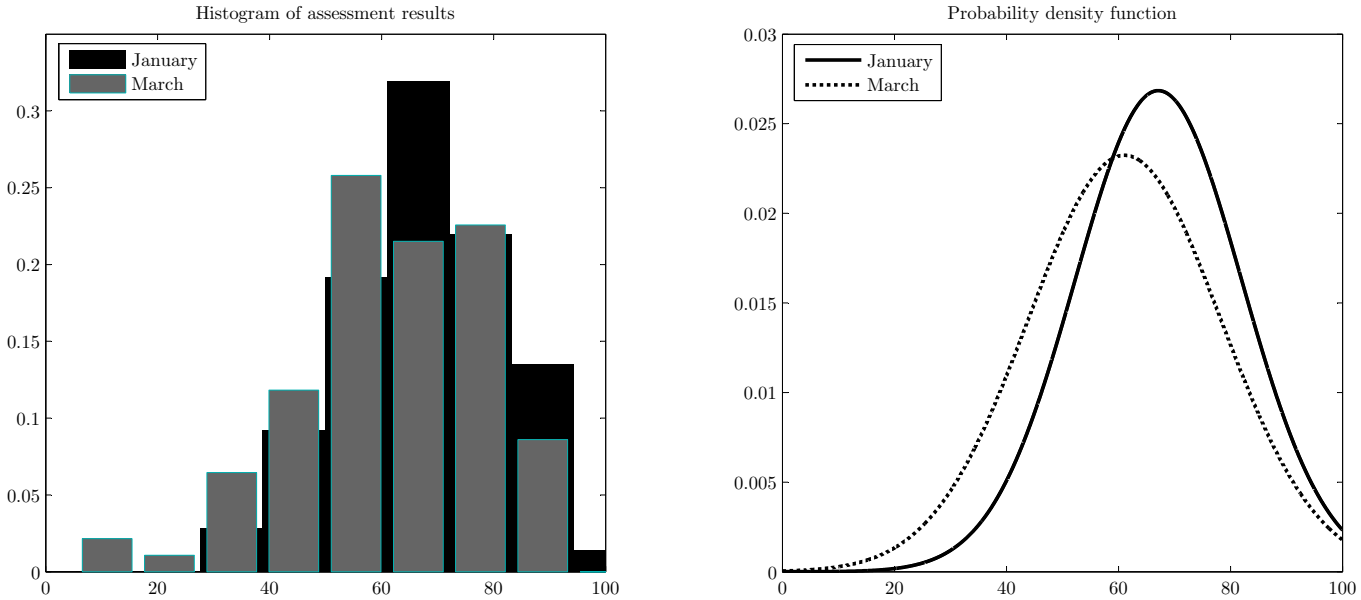


Section 4

# Assessment timing and performance

# Distribution of marks

Figure : Mark distribution from the two cohorts




The left hand pane represents a histogram of assessment results from the two cohorts: 'January' represents those taking it early and 'March' representing those taking it later. The right hand pane shows a probability density function of marks imposing a normal distribution on the two cohorts.

# Regression analysis

back: scatter



Table : Regression analysis for mark achieved in the first assessment

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Early | 5.808*** | 2.984 | 0.848 | -19.994*** |
| | (0.005) | (0.167) | (0.702) | (0.019) |
| QOE | | 14.438* | 9.476 | -14.432 |
| | | (0.072) | (0.238) | (0.240) |
| Attendance | | | 0.497** | 0.516** |
| | | | (0.034) | (0.026) |
| VLE | | | 11.126* | 11.526* |
| | | | (0.078) | (0.064) |
| Early×QEO | | | | 0.090** |
| | | | | (0.011) |
| Other controls | No | Yes | Yes | Yes |
| $R^2$ | 0.035 | 0.175 | 0.253 | 0.280 |
| n | 222 | 183 | 181 | 181 |

# The timing of assessment and the impact on attendance



Figure : Attendance before and after option offered

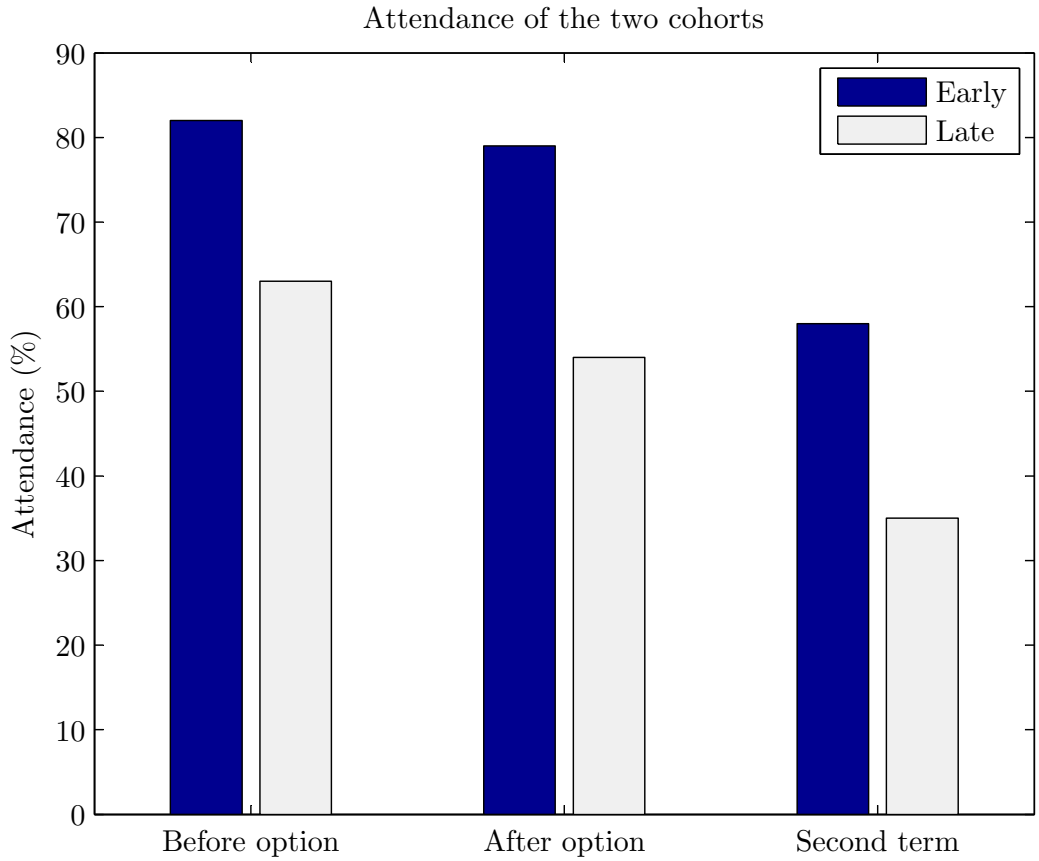
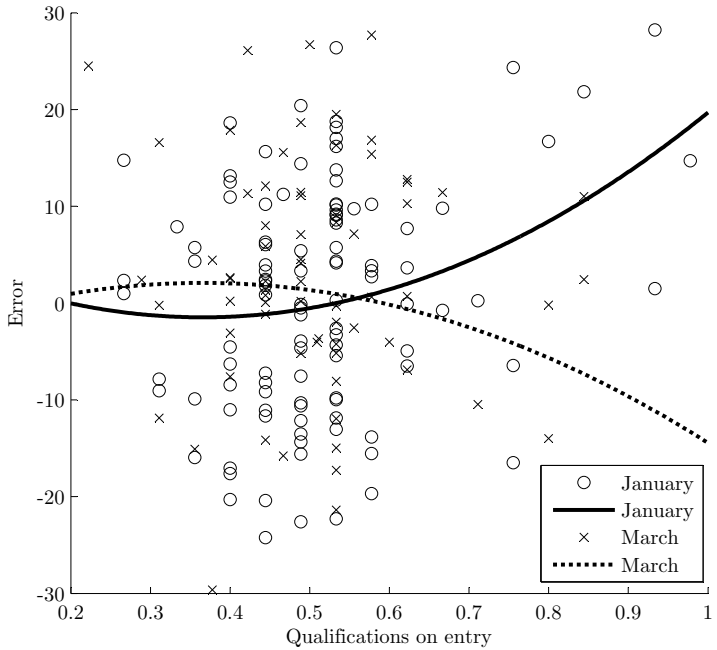


Figure : Qualifications on entry and time of assessment




The y-axis ('error') is obtained from running a regression as in specification (3) in Table 3 without the 'Time' and 'QOE' variable'. Lines of best fit obtained using a second order polynomial.





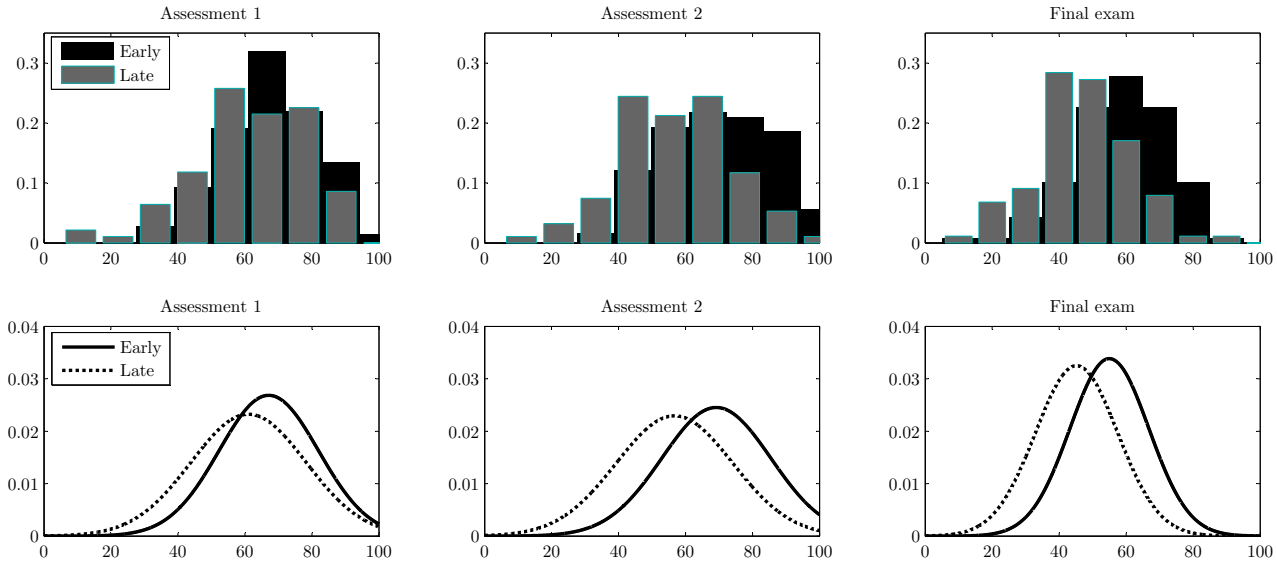
Subsection 1

# Impact in future assessments

# The effect on future assessments



Figure : Mark distribution from the two cohorts in all assessments




The top column represents a histogram of assessment results from the two cohorts: 'Early' represents those taking the first assessment in the first sitting, and 'Late' represents those taking it in the later sitting. The bottom column shows a probability density function of marks after imposing a normal distribution on the two cohorts, presented in chronological order of when the assessment taken. Note that 'Assessment 2' and the 'Final exam' were taken at the same time by all students. .

# The effect on future assessments

survey results

Table : The effect of assessment timing on future assessments

| | Assessment 1 | | Assessment 2 | | Final exam | |
|---|---|---|---|---|---|---|
| Early | 5.808*** | 0.848 | 12.482*** | 4.810** | 10.646*** | 4.076** |
| | (0.005) | (0.702) | (0.000) | (0.042) | (0.000) | (0.044) |
| Controls | No | Yes | No | Yes | No | Yes |
| $R^2$ | 0.035 | 0.253 | 0.121 | 0.353 | 0.133 | 0.309 |
| n | 222 | 181 | 216 | 197 | 207 | 191 |

'Early' represents a variable taking the value of 1 if the first assessment was taken early and 0 taken late; 'Controls' represents whether or not the regression specification includes other control variables, selected for each specific assessment using a general-to-specific approach. Figures in parenthesis represent p-values, and the star convention is the same as in Table 2.

# Potential explanations



## Potential explanations

Potential explanations for future performance:

- There is a **legacy effect** such that delaying engagement in module content 'catches up with' the students eventually;
- That performance in the first assessment (where timing was optional) was actually better as a result of providing students which flexibility. This is supported by the end-of-year survey (Figure 2) where 69% of students claimed that the option of timing meant they performed better.

# Potential explanations



## Potential explanations

Potential explanations for future performance:

- There is a **legacy effect** such that delaying engagement in module content 'catches up with' the students eventually;
- That performance in the first assessment (where timing was optional) was actually better as a result of providing students which flexibility. This is supported by the end-of-year survey (Figure 2) where 69% of students claimed that the option of timing meant they performed better.

# Potential explanations



## Potential explanations

Potential explanations for future performance:

- There is a **legacy effect** such that delaying engagement in module content 'catches up with' the students eventually;
- That performance in the first assessment (where timing was optional) was actually better as a result of providing students which flexibility. This is supported by the end-of-year survey (Figure 2) where 69% of students claimed that the option of timing meant they performed better.



Section 5

# Conclusions and discussion

# Conclusions



## Conclusions

- **Clear preferences** of having this option were shown (only 2% of students stated to be indifferent) with those more comfortable and engaged in the module electing to take an earlier sitting of the assessment.
- Those who took the **early test performed better** on average compared to those who took it later, however, after controlling for attendance, there is no statistical link.
- There was, however, evidence that later assessment *caused* lower attendance and moreover, evidence of a **legacy effect** of this timing where the out-performance of the early cohort grew over later tests, which all students took at the same time.

# Future work



## Future work

- Continue with different cohorts and spacing to see if results are robust;
- Try to identify methods to ascertain:
  - Which of the potential explanations on future performance is more likely;
  - Whether overall there is an improvement in marks.
- You comments and questions.

# References




Ariely, D. & Wertenbroch, K. (2002), 'Procrastination, deadlines, and performance: Self-control by precommitment', *Psychological Science* **13**(3), 219–224.

Birenbaum, M. (2007), 'Assessment and instruction preferences and their relationship with test anxiety and learning strategies', *Higher Education* **53**(6), 749–768.

Carpenter, S. K. & DeLosh, E. L. (2005), 'Application of the testing and spacing effects to name learning', *Applied Cognitive Psychology* **19**(5), 619–636.

Iannone, P. & Simpson, A. (2014), 'Students' preferences in undergraduate mathematics assessment', *Studies in Higher Education* (ahead-of-print), 1–22.

Zoller, U. & Ben-Chaim, D. (1989), 'Interaction between examination type, anxiety state, and academic achievement in college science; an action-oriented research', *Journal of Research in Science Teaching* **26**(1), 65–77.

# Expectations of peers back