USING GAMES TO IMPROVE STUDENTS' ENGAGEMENT AND UNDERSTANDING OF STATISTICS IN HIGHER EDUCATION

Joshua Fullard[1]

Abstract

In this article we present a series of light-touch, easy to implement, classroom activities designed to improve students' understanding and interpretation of regression coefficients. Using an experiment, we evaluate the effectiveness of the activities by randomly allocating first-year undergraduate students into a seminar with the activities (treatment group) or not (control group). We demonstrate that the classroom activities have a large, positive effect on several domains including engagement (measured by seminar attendance), enjoyment of learning (measured by an end-of-term student satisfaction survey), and their understanding of the material (measured by test scores). These results suggest that adopting more varied pedagogical approaches to teaching statistics/econometrics can benefit students.

Keywords: Teaching Statistics, Classroom Activity, Experiment

JEL Classification: A22

Introduction

Regression analysis is a vital statistical tool used to quantify the relationship between variables and is a staple across undergraduate social science programs across the world. However, both the approach to teaching and the content (a focus on models and math rather than causal questions and empirical examples) remains largely unchanged since the 1990s (Angrist & Pischke, 2017; Asarta et al., 2021; Becker & Watts, 1996; Watts & Schaur, 2011). Today, the majority of economics and business undergraduate statistics courses (83 percent) continue to be taught exclusively using the traditional "chalk-and-talk" approach - a formal method of teaching where the focal points are the blackboard (or a PowerPoint presentation) and the teacher's voice (Harter & Asarta, 2022).

Students in higher education generally hold strong, and often quite varied, preferences over how learning activities are structured and delivered (Alaoutinen, 2012; Ng et al., 2011). While there is an ongoing debate over how, or if, these preferences adapt to different learning contexts or environments, there is a consensus that there are benefits to designing learning activities that incorporate different pedagogical approaches in the classroom (Donche et al., 2013; Middleditch et al., 2022; Richardson, 2011). On the one hand, if students' preferences are adaptive there are a range of benefits to using different pedagogical approaches including the development of student ability to acquire (visual vs verbal), process (active vs reflective) and

[1] Assistant Professor of Behavioural Science, Warwick Business School
University of Warwick, Scarman Road, Coventry, CV4 7AL
Research Associate, Research Centre on Micro-Social Change, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ

understand (sequential vs global) information. Even if students' preferences are not adaptive, using different pedagogical approaches ensures that all students have opportunities to engage in activities that most suit their preferences.

While students do hold strong preferences to be taught in certain ways, and designing learning activities to use students preferred approach can be beneficial for educators in higher education (e.g., higher student evaluations), there is limited evidence that it improves student learning: all students benefit from receiving information in multiple ways (Kozhevnikov et al., 2002; Lapp et al., 1999; Morehead et al., 2016; Nancekivell et al., 2020; Tardif et al., 2015; Zrudlo, 2023).

While using more creative pedagogical approaches presents certain challenges (e.g., adapting teaching material), the advantages they offer are significant. Using more creative pedogeological approaches serves as an effective way to engage students and cultivate a dynamic learning environment, as well as improve students understanding of the material and can be used as an informal method of assessment - the answers that students give in the activities can help guide the direction of future learning (Leach & Sugarman, 2005; Schwartz & Martin, 2004; Sekwena, 2023).

In this paper, we present a series of easy to implement activities designed to improve students' ability to interpret regression coefficients, understanding of the potential mechanisms driving relationships, and interpreting effect sizes. The activities were designed to incorporate three approaches known to improve learning in higher education: repetition, active and reflective learning, and gamification. Repetition is shown to support a range of educational outcomes, including students' clarity of thought and engagement (Bruner, 2001). Including opportunities for active learning (such as group discussions) and reflective learning (such as individual reflection) supports the development of students critical thinking skills, interpersonal skills, and knowledge retention (Perusso et al., 2020; Prince, 2004), and gamification is shown to improve student engagement, motivation, and enjoyment of learning (Huang & Soman, 2013; Oliveira et al., 2023; Surendeleg et al., 2014).

Our paper belongs to a long tradition of work investigating how a more varied pedagogical approach, and games in particular, can be used in higher education to improve student learning. There is a rich history of using games in higher education to teach statistics and related topics. Early activities include the pit market games at Harvard University (Chamberlin, 1948) and the traveler's dilemma (Basu, 1994) to more recent activities such as The Invisible Hand in Action (Settlage & Wollscheid, 2019) as well as using popular TV shows such as Breaking Bad (Muchiri et al., 2020), the Big Band Theory (Geerling et al., 2018) and Parks and Recreation (Wooten & Staub, 2017). While games and activities, such as these, are shown to improve pupil outcomes they can present challenges for the teacher. Firstly, these activities are often time-consuming to plan and are not necessarily transferable across cohorts or programs – particularly programs with a higher proportion of international students who might not have the same cultural references. Second, these activities are often around 60 minutes in duration. Given the limited time that teachers have with their students it is often challenging to know what content to remove to make space for such a long activity. We add to this literature by providing exploratory evidence of how a series of easy to implement activities, used at the end of seminars, can improve student outcomes across several different domains. In a wider sense, this paper suggests there are benefits to using a more varied pedagogical approach to teaching statistics.

Our activities were designed for a first-year undergraduate module at a leading business school in the UK. The module is designed to teach students how to interpret statistical results and is compulsory for management students. The module runs for ten weeks where students attend one lecture (weeks 1 – 10) and one seminar (weeks 2 – 10) each week. The activities generally lasted for about 10 minutes and were used at the end of seminars in week 2, week 4, week 7, and week 9. The activities were used in three seminar groups - 27 percent of the students in this cohort (the treatment group). The remaining eight seminar groups did not do the activities (control group). Students were randomly allocated into seminars and there is very little switching between groups.

Using data on student seminar attendance, an end-of-term student satisfaction survey administered by the school, and performance on the final exam, we can investigate the impact the activity had on these three separate domains. Because students are randomly assigned to seminar groups, we argue that any differences observed could be interpreted as causal, but, due to some limitations, we are suggesting that these results should be considered exploratory. This is discussed further in the discussion section.

We find that our activity has a large positive effect on student engagement (measured by attendance), enjoyment of learning (measured by an end of term student satisfaction survey), and their understanding of the material (measured by test scores). Finally, we show that this activity can be easily modified to meet different learning objectives and levels of study. In the following sections, we describe the activity, discuss the impact it had on student outcomes, discuss our results, and conclude.

Activity

For ease of interpretation, we present the activities that we used in the seminars as a single activity with multiple rounds. In practice, the activity presented here was split into smaller sections and used over a 9-week period at the end of seminars (in week 2, week 4, week 7 and week 9). This activity was used in seminars but could easily be scaled up to be used in a lecture. For instance, students could be presented with questions using Vevox (or a similar software).

Set-up

Split students into small teams. We used this activity in seminars (groups of around 20 students) and separated students into teams of 4. To ensure each student can activity participate we would recommend having teams of no more than 6 students. Each team must have something to write with and on (i.e., a piece of paper and a pen or a whiteboard and marker).

Once students are split into teams, create a scoreboard (see Table 1). With each column representing a team and each row representing a round of game play. While you create the scoreboard on the whiteboard or flipchart, it might be nice to let the students come up with their own team names.

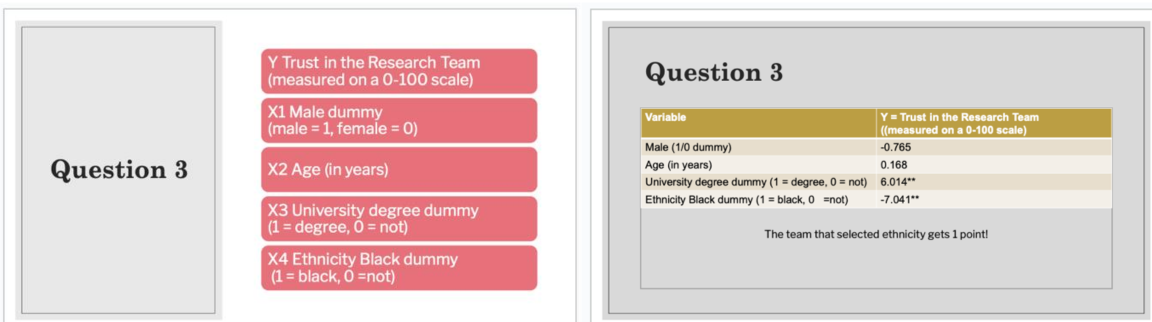
Table 1. Scoreboard for the activity. Each column represents a team, and each row represents a round of game play.

	Team 1	Team 2	Team 3	Team 4	Team 5
Round 1					
Round 2					
Round 3					

Round 1: Interpretation of coefficients

This first round is designed to help students interpret regression coefficients. Students are presented with an outcome variable and four (or more) explanatory variables. Each team must predict which of these variables has the strongest relationship (positive or negative) with the outcome variable. As illustrated in Figure 1, the outcome variable could be survey respondents' trust in the research team (measured on a 0 – 100 scale), and the explanatory variables could be sex (Male 1/0 dummy), age (in years), educational attainment (university degree 1/0 dummy), and ethnicity (black 1/0 dummy).

Figure 1. Shows an example question (LHS) and answer (RHS) used in an Undergraduate seminar. Source: Fullard (2022).



After a period for individual reflection (e.g., 30 seconds) and team discussion (e.g., 60 seconds), each team writes down an answer. Once complete, each team reveals their answers. Although it does not affect the outcomes (teams have already written down their answers), to help with the game play each team reveals their answers sequentially (i.e., Team 1, Team 2, Team 3…) with a different team going first for each question (e.g., for question 1, Team 1 goes first, for question 2, Team 2 goes first. and for question 6, Team 1 goes first again).

Once all the teams have stated their answers, the teacher then reveals the actual coefficients. The team(s) that selects the correct variable received 1 point. Teams that guessed correctly can also receive a bonus point if, after a short team deliberation, they correctly interpret the regression coefficient. To illustrate this using the example from Figure 1, the team(s) that selected ethnicity would get 1 point and would get a bonus point if they correctly stated that black participants' trust in the research team is 7 percentage points lower than white participants on average.

For practical purposes, if only one team gets the correct answer, they are asked to interpret the coefficient in class. If multiple teams get the correct answer, we ask a team representative to share their answer with the teacher individually. Once all answers are given, the points are awarded for the question, and the teacher moves onto the second question.

In Round 1, we vary the scale of the dependent and independent variables to test the students understanding. For example, the outcome variables we have used in this section include GDP per capita, Grade 4 Math Score (standardized), and number of hours university students study per week. These examples are all presented in Appendix A. It is important to note that the explanatory variables that we use are not always directly comparable by design (i.e., a male 1/0 dummy compared to an age in years variable). This gives us the opportunity to discuss what these coefficients mean. For instance, you could have a class discussion about what information you would need to evaluate the effect sizes more accurately. This is what we do in Round 3.

Round 2: Understanding mechanisms

In our experience teaching econometrics and statistics, students generally find the math straightforward but struggle with the interpretation – although we recognize this might not be the experience of all practitioners as we are at a highly selective institute. This second round is intended to improve this.

In this round, students are shown one outcome variable and one explanatory variable, and each team must guess if there is a positive relationship, a negative relationship, or no relationship between these variables. After a period of individual reflection (e.g., 30 seconds) and team discussion (e.g., 60 seconds), each team writes down their answer. Once that is complete, each team reveals their answers. To make it more challenging, each team must also give a plausible mechanism that might explain the relationship they predict (no repeats allowed).

For example, if the outcome variable is a 1/0 dummy that indicates a university graduates' decision to enroll into postgraduate education or not, and the explanatory variable is the unemployment rate at time of graduation, the following mechanisms might be discussed (source Fullard (2021a)):

Positive:

- Graduates are aware that it is more challenging to get a job during periods of high unemployment, so they are more likely to continue in education.
- Universities want to "protect" their graduate employment statistics so, during periods of high unemployment, they offer larger incentives to encourage graduates to continue into postgraduate study.
- Reduced opportunity cost of postgraduate study. Finding a job during a period of high unemployment is hard, and those who do generally earn less. Therefore, the forgone earnings from continuing in education are lower.

None:

- Students have inaccurate beliefs about labor market opportunities – they do not realize that it is harder to get a job.
- Universities have capacity constraints (there are limited places). While there might be an increase in demand (more applications), there might be no change in the quantity of graduates enrolled.

Negative:

- Concerns about employment opportunities might make students more focused on finding a job.
- Concerns about employability and finances might make students less willing to take out loans to fund their postgraduate studies.

After all the answers are revealed and potential mechanisms introduced, points are awarded for correct answers (1 point) and the "best" explanation(s) (1 point). The second part can be subjective, and we award points generously if the mechanisms are plausible and clearly articulated.

In Round 2 we use a range of outcomes and interesting relationships including the relationship between: physical attractiveness and performance evaluations (i.e., does beauty help or hinder), a fine and parents picking their children late from day-care, watching eyes on bins and littering, and incentivizing dishonesty. These examples are all presented in Appendix B.

Round 3: Effect size and statistical significance

Recall that Round 1 was designed to help students interpret regression coefficients. Round 3 is intended to build on this. Instead of showing students some coefficients with stars that indicate statistical significance, students in Round 3 are shown a coefficient with a p-value (so they can identify if the effect is statistically significant or not) and the dependent variable's mean and standard deviation (to help them identify the relative effect size). Using this information, students are asked to identify whether the effect is statistically significant (or not) and if the effect size is significant (or not).

Figure 2.
Shows an example question used in an Undergraduate seminar.
Source: Fullard (2023b)

Question 1

	Intentions to leave teaching In the next 12 months (0 – 100)	
	Coefficient	P-Value
Log(Teachers Pay)	-13.73	0.041
DV mean	14.18	
DV Standard Deviation	25.47	

As before, after a period for individual reflection (e.g., 30 seconds) and team discussion (e.g., 60 seconds), each team writes down their answers on a piece of paper, and the answers are revealed sequentially. Once all the teams have stated their answers, the teacher then reveals the correct answers. As illustrated in Figure 2, the coefficient of -13.73 would be identified as statistically significant at the 5% level. The team(s) that said it is statistically significant would receive 1 point. The effect size would also be identified as significant. The team(s) that said the effect size is significant would receive 1 point.

In this round we found it useful to include questions with a range of different answers and, after each question, have a class discussion about how students evaluated the magnitude of the effect size. For instance, to revisit our example in Figure 2, the effect size, intuitively, sounds very large (a 1 percentage point increase in teachers' pay is associated to a 13-percentage point lower probability of leaving the profession), but it is important to put the coefficient into context by looking at the mean and standard deviation of the dependent variable. In this case, the effect size is 0.54 of a standard deviation (13.73/25.47) – which confirms our intuition. Other examples that we used in Round 3 include job satisfaction and occupational choice, and inequality preferences and mental health. Both these examples are available in Appendix C.

Results

In this section, we describe the effect that the activity has had on student engagement (measured by seminar attendance), enjoyment (measured by a student satisfaction survey), and attainment (measured by performance on the end-of-term exam).

Student Engagement

Attendance in each seminar is collected by the teacher at the start of the seminar and recorded online by the business school. Using the weekly attendance data, we can investigate if the seminars in the treatment group had higher attendance than those in the control group. We recognize that we have a relatively small sample (11 seminar groups) and that it is entirely possible that some seminar groups might have systematically higher (or lower) attendance than others. To adjust for this, we measure attendance in seminars 2 - 9 as a proportion of seminar 1 attendance. For example, if a seminar group has 15 students attend out of a possible 20 in the first seminar, then attendance rate in the subsequent seminars is calculated by the number of students who attend divided by 15.

We find that the seminars in the treatment group had, on average, 6 percentage points higher attendance each week than those in the control group (94.4% vs 88.5%). To put this into perspective, seminars that used the class activity had 1.7 more students in each seminar each week. This could be driven by two, potentially nonexclusive, mechanisms. First, students might find the seminars with the activity more enjoyable or more engaging. Alternatively, students might find seminars in the treatment group more useful which increases the expected returns to attendance – or the cost of non-attendance. While we are not able to distinguish between these two plausible mechanisms, we can investigate the effect on student enjoyment directly using the results of the student satisfaction survey.

Student Enjoyment

At the end of each term, students complete a school-administered student satisfaction survey, where they respond to questions about the module they have just taken. These questions include details about the seminar they attended measured on a 5-point scale (1 lowest, 5 highest). The questions we are interested in include; if they are satisfied with the teaching they received, if the material was interesting, and if they were encouraged to ask questions. Using this data, we can investigate if seminars in the treatment group performed better on these metrics than those in the control group. Note that we only have access to the aggregate level student satisfaction data (the mean scores by treatment and control group) so this limits the statistics we can report using this data.

Table 2 shows the student satisfaction scores from those in the treatment group (column 1), those in the control group (column 2), and the difference between the groups (column 3). Two main things stand out. First, students in the treatment group are more satisfied with the teaching they received (4.76 vs 3.91), they found the subject more interesting (4.80 vs 3.74), and the content was more clearly explained (4.64 vs 3.97). This suggests that the class activity made the seminars more interesting and engaging, and this improved the quality of the instruction they received. This is supported by some of the students' qualitative feedback in the student satisfaction survey where they highlight the role of the class activity directly: "[the teacher] even made his own exercises for seminars to help us maintain knowledge … made everyone engage in seminars to ensure that we all learnt," and "The quiz competition we had about regression was fun".

Second, we also observe positive effects on encouraging students to ask questions (4.80 vs 3.91). This suggests that the class activity also helped create a more open learning environment. This is supported by students' comments stating that the class activity made learning more approachable: "There's also been several occasions where [the teacher] used an activity to engage the class further and make learning as approachable as possible," and "I always felt included, which made me more motivated and engaged… [including] visual examples that helped me understand the topic better."

The results from the student satisfaction survey suggest that the activity improved the quality of the instruction the students received in seminars. Next, we investigate if the activity improves student attainment in the module exam.

Table 2. Shows the results from the end of term student satisfaction survey for the students who are in seminars with the class activity (column 1), those who are not (column 2) and the difference between them (column 3).

	(1)	(2)	(3)
Category	Treatment (Class Activity)	Control (All other Students)	Difference
I am satisfied with the teaching	4.76	3.91	0.85
Made the subject interesting	4.80	3.74	1.06
Encouraged me to ask questions	4.80	3.91	0.89
Good at explaining things	4.64	3.97	0.67
Max students	50	115	

Note: the students are asked these questions about the seminar they attended on a 5-point scale where 1 is worst and 5 is best. For the student satisfactions data we only have access to the aggregate level statistics presented above (means for each group and approximate sample size).

Student Attainment

The main method of assessment in this module is a 90-minute open book exam that takes place in January. The exam is designed to test students' ability to estimate and interpret different statistics – with a focus on interpreting regression coefficients. The exam was marked on a 0 – 100 scale by members of the module's teaching team. For those not familiar with the UK marking system, a mark of 70 or higher is a first (the equivalent of an A), a mark of 60 – 69 is an upper second (the equivalent to a B), a mark of 50 – 59 is a lower second (the equivalent of a C), a mark of 40 – 49 is a third (the equivalent of a D), and a mark below 40 is a fail (the equivalent of an F). It is important to note that we restrict our analysis to students who sat for the exam in January. We do not have data for the other students who sat for the exam later. Using the results from this exam we can investigate if the students in the treatment group had a stronger understanding of the material than those in the control group.

Table 3 shows the students exam scores (row 1) and degree classifications (row 2 – 6) for students in the treatment group (column 1), control group (column 2) and the difference between them (column 3). Looking at the marks achieved, we observe that students in the seminars with the class activity achieved higher marks than those in the control group (65.4 vs 58.5), and the difference is large (7 percentage points). To put the magnitude into perspective, the median student in the treatment group received a mark that is almost a full degree classification higher than the median student in the control group (69 vs 60).

Turing our attention to the distribution of marks, we observe that a lower proportion of students failed (7.5 percent vs 18 percent), and a higher proportion achieved a first (46.3 percent vs 31.2 percent) in the treatment group compared to the control group. This suggests that students across the ability distribution uniformly benefited from the more varied pedagogical approach.

Table 3. Shows the students mean marks (standard error) (row 1) and proportion of students within each grade boundary (row 2 – 5) in the module exam for students who are in seminars with the class activity (column 1), those who are not (column 2) and the difference between them (column 3).

	(1)	(2)	(3)
Exam Outcome	Treatment (Class Activity)	Control (All other Students)	Difference
Mark (0 – 100)	65.44 (16.57)	58.52 (17.50)	6.92***
Degree Classification			
Fail (Mark Less than 40)	7.46 (26.47)	17.99 (38.51)	10.52**
Third Class (Mark 40 to 49)	11.94 (32.67)	11.64 (32.15)	0.30
Lower Second Class (Mark 50 to 59)	17.91 (38.63)	19.57 (39.78)	-1.66
Upper Second Class (Mark 60 to 69)	16.42 (37.32)	19.58 (39.78)	-3.16
First Class (Mark 70 or higher)	46.26 (50.23)	31.21 (46.46)	15.05**
Number of Students	67	189	

Note: This excludes students who did not sit the exam in January and completed their exam at a later period (e.g., due to mitigating circumstances). Stars indicate statistical significance to the usual levels: $^{*}P < 0.10$, $^{**}P < 0.05$, $^{***}p < 0.01$

Discussion

In this paper, we present an easy to implement series of classroom activities that are designed to improve students understanding and interpretation of regression coefficients in higher education. We also provide evidence that these activities were highly successful at improving student engagement (measured by seminar attendance), enjoyment of learning (measured by a student satisfaction survey), and attainment (measured by performance on the end-of-term exam). Moreover, the effect sizes are large.

As students are randomly assigned to seminar groups, and there is very little switching between seminar groups, the difference in outcomes could be interpreted as the causal effect of the seminar activities. However, we recognize that we have a small sample, and it is possible that the randomization was not successful - students in the treatment group might be systematically different than those in the control group. Ideally, we would be able to assess the comparability of the students in the treatment and control groups directly by investigating if they are balanced on observable characteristics such as age, ethnicity, and prior attainment, but we do not have access to this information. The consistent size and direction of the effects, across different dimensions, combined with the qualitative feedback received in the student satisfaction survey, suggest that using a more varied pedagogical approaches in the classroom can have a positive effect on student learning in our setting. We also recognize that these results are also consistent with students in the treatment group being systematically more engaging that those in the control group. For these reasons our results are intended to be interpreted as exploratory evidence that using light touch activities in the classroom can have a positive effect on student learning.

While the activities presented are intended to improve students' understanding of regressions and regression coefficients, it is unlikely that the benefits are driven by participating in this activity alone. It is more possible that the activity also impacted students' behavior outside of the classroom. One plausible mechanism is that the activities improved students' initial understanding of the material so that their independent study/revision time was more productive. Alternatively, the interactive elements of the activities might have encouraged more peer learning and revision outside of the seminars. This is particularly important for first year undergraduates in the first few weeks of term. Investigating the role that activities have on a wider range of student outcomes, such as studying behavior and the perceived returns to different university-related behaviors seems like a promising area of future research.

Concluding remarks

Despite the evidence that using games and activities in the classroom can improve pupil outcomes, the way statistics is taught in higher education remains largely unchanged since the 1990s. While teachers might be reluctant to include long activities into an already cramped teaching schedule, we provide suggestive evidence that even short, easy to implement, activities can positively impact pupil outcomes.

References

Alaoutinen, Satu. 2012. "Evaluating the effect of learning style and student background on self assessment accuracy". *Computer Science Education*, *22*(2), 175-198. https://doi.org/10.1080/08993408.2012.692924

Angrist, Joshua D, & Pischke, Jörn-Steffen. 2017. "Undergraduate econometrics instruction: through our classes, darkly". *Journal of Economic Perspectives*, *31*(2), 125-144. https://www.aeaweb.org/articles?id=10.1257/jep.31.2.125

Asarta, Carlos J, Chambers, Rebecca G, & Harter, Cynthia. 2021. "Teaching methods in undergraduate introductory economics courses: Results from a sixth national quinquennial survey". *The American Economist*, *66*(1), 18-28.

Basu, Kaushik. 1994. "The traveler's dilemma: Paradoxes of rationality in game theory". *The American Economic Review*, *84*(2), 391-395. https://www.jstor.org/stable/2117865

Bateson, Melissa, Robinson, Rebecca, Abayomi-Cole, Tim, Greenlees, Josh, O'Connor, Abby, & Nettle, Daniel. 2015. "Watching eyes on potential litter can reduce littering: evidence from two field experiments". *PeerJ*, *3*, e1443. https://peerj.com/articles/1443/

Becker, William E, & Watts, Michael. 1996. "Chalk and talk: A national survey on teaching undergraduate economics". *The American Economic Review*, *86*(2), 448-453. https://www.jstor.org/stable/2118168

Boulhol, Hervé, De Serres, Alain, & Molnar, Margit. 2008. "The contribution of economic geography to GDP per capita". https://doi.org/10.1787/18151973

Bruner, Robert F. 2001. "Repetition is the first principle of all learning". *Available at SSRN 224340*. https://www.researchgate.net/publication/228318502_Repetition_is_the_First_Principle_of_All_Learning#:~:text=...-,Repetition%20reinforces%20learning.,%2C%202018).%20...

Chamberlin, Edward H. 1948. "An experimental imperfect market". *Journal of Political Economy*, *56*(2), 95-108. https://www.jstor.org/stable/1826387

Cohn, Alain, Maréchal, Michel André, Tannenbaum, David, & Zünd, Christian Lukas. 2019. "Civic honesty around the globe". *Science*, *365*(6448), 70-73. https://www.science.org/doi/10.1126/science.aau8712

Delavande, Adeline, Del Bono, Emilia, & Holford, Angus. 2022. "Academic and non-academic investments at university: The role of expectations, preferences and constraints". *Journal of Econometrics*, *231*(1), 74-97. https://doi.org/10.1016/j.jeconom.2020.03.019

Donche, Vincent, De Maeyer, Sven, Coertjens, Liesje, Van Daal, Tine, & Van Petegem, Peter. 2013. "Differential use of learning strategies in first-year higher education: The impact of personality, academic motivation, and teaching strategies". *British Journal of Educational Psychology*, *83*(2), 238-251. https://doi.org/10.1111/bjep.12016

Fullard, Joshua. 2021a. "Bad economy, good teachers? The countercyclicality of enrolment Into Initial Teacher Training Programmes in the UK". *ISER Working Paper Series*. https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2021-06

Fullard, Joshua. 2021b. "Relative wages and pupil performance, evidence from TIMSS". *ISER Working Paper Series*. https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2021-07

Fullard, Joshua. 2022. "Tell Me Who You Are and I Will Give You My Consent: A Light-touch Intervention on Consent to Data Linkage". https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2022-10

Fullard, Joshua. 2023a. "Does the Beauty Beat the Beast?". *Working Paper*.

Fullard, Joshua. 2023b. "Labour market expectations and occupational choice: evidence from teaching". *Available at SSRN 4384928*. https://doi.org/https://dx.doi.org/10.2139/ssrn.4384928

Fullard, Joshua. 2023c. "Unhappy, uninformed and inequality averse. The role of wellbeing, labour market beliefs and inequality preferences on occupational choice". *Working Paper*.

Geerling, Wayne, Mateer, G Dirk, Smith, Ben O, Tierney, James E, & Wooten, Jadrian J. 2018. "Lesson plans for teaching economics with The Big Bang Theory". *Lesson Plans for Teaching Economics with The Big Bang Theory*, 163. https://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1060&context=econrealestatefacpub

Gneezy, Uri, & Rustichini, Aldo. 2000. “A fine is a price”. *The journal of legal studies*, *29*(1), 1-17. https://www.jstor.org/stable/10.1086/468061

Harter, Cynthia, & Asarta, Carlos J. 2022. “Teaching methods in undergraduate intermediate theory, statistics and econometrics, and other upper-division economics courses: Results from a sixth national quinquennial survey”. *The American Economist*, *67*(1), 132-146. https://journals.sagepub.com/doi/full/10.1177/05694345211037904

Huang, Wendy Hsin-Yuan, & Soman, Dilip. 2013. “Gamification of education”. *Report Series: Behavioural Economics in Action*, *29*(4), 37. https://mybrainware.com/wp-content/uploads/2017/11/Gamification-in-Education-Huang.pdf

Kozhevnikov, Maria, Hegarty, Mary, & Mayer, Richard E. 2002. “Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers”. *Cognition and instruction*, *20*(1), 47-77. https://doi.org/https://doi.org/10.1207/S1532690XCI2001_3

Lapp, Diane, Flood, James, & Fisher, Douglas. 1999. “Visual literacy: Intermediality: How the use of multiple media enhances learning”. *The Reading Teacher*, *52*(7), 776-780. https://doi.org/https://www.jstor.org/stable/20204683

Leach, Guy J, & Sugarman, Tammy S. 2005. “Play to win! Using games in library instruction to enhance student learning”. *Research strategies*, *20*(3), 191-203. https://doi.org/10.1016/j.resstr.2006.05.002

Middleditch, Paul, Moindrot, Will, & Rudkin, Simon. 2022. “Teaching with Twitter: An extension to the traditional learning environment”. *International Review of Economics Education*, *41*, 100252. https://doi.org/10.1016/j.iree.2022.100252

Morehead, Kayla, Rhodes, Matthew G, & DeLozier, Sarah. 2016. “Instructor and student knowledge of study strategies”. *Memory*, *24*(2), 257-271. https://doi.org/https://doi.org/10.1080/09658211.2014.1001992

Muchiri, Steve, Paraschiv, Mihai, & Wooten, Jadrian. 2020. “Teaching Economics with Breaking Bad”. *Available at SSRN 3525829*. https://www.researchgate.net/profile/Steve-Muchiri-2/publication/341118128_The_Economics_of_Breaking_Bad_A_Concept_Guide/links/60464f66299bf1e07865f01d/The-Economics-of-Breaking-Bad-A-Concept-Guide.pdf

Nancekivell, Shaylene E, Shah, Priti, & Gelman, Susan A. 2020. “Maybe they’re born with it, or maybe it’s experience: Toward a deeper understanding of the learning style myth”. *Journal of Educational Psychology*, *112*(2), 221. https://doi.org/http://dx.doi.org/10.1037/edu0000366

Ng, Pin, Pinto, James, & Williams, Susan K. 2011. “The effects of learning styles on course performance: A quantile regression analysis”. *Academy of Educational Leadership Journal*, *15*(1), 15-37. http://openknowledge.nau.edu/id/eprint/1533/7/Ng_P_Etal_2008_FrankeWPS_08-02(1).pdf

Oliveira, Wilk, Hamari, Juho, Shi, Lei, Toda, Armando M, Rodrigues, Luiz, Palomino, Paula T, & Isotani, Seiji. 2023. “Tailored gamification in education: A literature review and future agenda”. *Education and Information Technologies*, *28*(1), 373-406. https://link.springer.com/article/10.1007/s10639-022-11122-4

Perusso, André, Blankesteijn, Marlous, & Leal, Rafael. 2020. “The contribution of reflective learning to experiential learning in business education”. *Assessment & Evaluation in Higher Education*, *45*(7), 1001-1015. https://doi.org/10.1080/02602938.2019.1705963

Prince, Michael. 2004. “Does active learning work? A review of the research”. *Journal of engineering education*, *93*(3), 223-231. https://onlinelibrary.wiley.com/doi/10.1002/j.2168-9830.2004.tb00809.x

Richardson, John TE. 2011. “Approaches to studying, conceptions of learning and learning styles in higher education”. *Learning and Individual differences*, *21*(3), 288-293. https://doi.org/10.1016/j.lindif.2010.11.015

Schwartz, Daniel L, & Martin, Taylor. 2004. “Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction”. *Cognition and instruction*, *22*(2), 129-184. https://doi.org/10.1207/s1532690xci2202_1

Sekwena, Gailele L. 2023. “Active learning pedagogy for enriching economics students’ higher order thinking skills”. *International Journal of Learning, Teaching and Educational Research*, *22*(3), 241-255. https://scholar.ufs.ac.za/bitstream/handle/11660/12075/Sekwena_Active_2023.pdf?sequence=1&isAllowed=y

Settlage, Daniel M, & Wollscheid, Jim R. 2019. “The Invisible Hand in Action--An Interactive Classroom Experience”. *Journal of Economic Teaching*, *4*(1), 41-59. http://downloads.journalofeconomicsteaching.org/4/1/1-4.pdf

Surendeleg, Garamkhand, Murwa, Violet, Yun, Han-Kyung, & Kim, Yoon Sang. 2014. “The role of gamification in education–a literature review”. *Contemporary Engineering Sciences*, *7*(29), 1609-1616. https://www.researchgate.net/publication/266515512_Gamification_and_Education_a_Literature_Review

Tardif, Eric, Doudin, Pierre-André, & Meylan, Nicolas. 2015. “Neuromyths among teachers and student teachers”. *Mind, brain, and Education*, *9*(1), 50-59. https://doi.org/https://onlinelibrary.wiley.com/doi/pdf/10.1111/mbe.12070?casa_token=D7wc6n0M_W0AAAAA:AbVHeFeD-EhaajvN2u8xMtZ14zv4_YHYilMOkLqQcU7twFzzizyb52Cxelm5X5Hd7ISOd2mOCjAs6uZY

Watts, Michael, & Schaur, Georg. 2011. “Teaching and assessment methods in undergraduate economics: A fourth national quinquennial survey”. *The Journal of Economic Education*, *42*(3), 294-309. https://www.jstor.org/stable/23049285

Wooten, Jadrian, & Staub, Kalina. 2017. “Teaching Economics Using NBC's Parks and Recreation”. *Wooten, JJ, & Staub, K.(2019). Teaching economics using NBC’s Parks and Recreation. The Journal of Economic Education*, *50*(1), 87-88. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3002541

Zrudlo, Ilya. 2023. “Why the learning styles myth appeals and how to persuade believers otherwise”. *Teaching and teacher education*, *132*, 104266. https://doi.org/https://doi.org/10.1016/j.tate.2023.104266

Appendix A

Here we present some other regression tables that we used in Round 1.

Example 1
Who studies more? Evidence from a recent survey of students in the UK.

Variable	Y = Hours studied per week
Female (1/0 dummy)	3.171***
White British low-SES (1/0 dummy)	-0.640
EU (1/0 dummy)	0.224
Overseas (1/0 dummy)	0.472

Source: Delavande et al. (2022)
Stars indicate statistical significance to the usual levels: *p<0.10, ** p<0.05, ***p<0.01

Example 2
Who enjoys volunteering the most? Evidence from a recent survey of students in the UK.

Variable	Y = enjoyment of volunteering on a 0-100 scale
Female (1/0 dummy)	5.785***
White British low-SES (1/0 dummy)	4.832
EU (1/0 dummy)	3.939
Overseas (1/0 dummy)	11.020***

Source: Delavande et al. (2022)
Stars indicate statistical significance to the usual levels: *p<0.10, ** p<0.05, ***p<0.01

Example 3
Who does the best at Math?

Variable	Y = Grade 4 (9-19 year old) Math scores (standardized to have a mean 0 and standard deviation of 1)
Male (1/0 dummy)	0.0876***
Relative Student age (difference in months from the median age)	0.0350***
200 + books at home (1/0 dummy)	1.059***
Class size above median (difference in class size from the median)	0.123***

Source: Fullard (2021b)
Stars indicate statistical significance to the usual levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Example 4
What drives prosperity?

Variable	Y = GDP per capita (relative to the OECD average)
Physical Capital (factories and machinery)	0.184***
Human Capital (education)	0.334***
Population growth	-0.006

Source: Boulhol et al. (2008)
Stars indicate statistical significance to the usual levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Appendix B

Here we present some other relationships that we used in Round 2.

Example 1
How does physical appearance influence performance evaluation in the context of Mixed Martial Arts?

Context:
Y indicates if an MMA fighter won a close fight (1 = win, 0 = loss)
X is the fighter's attractiveness measured on a 0 – 100 scale

Question:
Is the coefficient for X positive, negative or no relationship?
Does physical attractiveness help or hinder an MMA fighter?

Answer:
Negative. Less attractive fighters are perceived as performing better.

Source: Fullard (2023a)

Example 2
The more it costs, the less you do it?

Context:
Y is the number of pupils picked up late from a day-care facility
X is if the day-care has a fine for picking up children late (1 = fine, 0 = no fine)

Question:
Is the coefficient for X positive, negative or no relationship?
Do fines reduce the likelihood that parents will pick up their children late?

Answer:
Negative. The fine increased the number of parents picking up their child late. The introduction of a fine changed the perception of picking up a child late – for a fixed cost a teacher will now take care of my student after 4pm.

Source: Gneezy and Rustichini (2000)

Example 3
Watching eyes?

Context:
Y is the amount of littering in a park
X is if the bins have watching eyes on them or not (1 = watching eyes, 0 = no eyes)

Question:
Is the coefficient for X positive, negative or no relationship?
Do watching eyes on bins reduce littering in parks?

Answer:
Negative. Watching eyes on bins reduces littering.

Source: Bateson et al. (2015)

Example 4
Can we incentivize dishonesty?

Context:
Y indicates if someone attempted to return a "lost wallet" or not (1 = attempt to return, 0 = not)
X indicates if there was money in the wallet ($13 USD) or not

Question:
Is the coefficient for X positive, negative or no relationship?
Does money in the wallet impact the likelihood that someone will try to return the wallet? If so, what is the direction?

Answer:
Positive. Having money in the wallet increases the likelihood that someone will try to return it. The wallets with no money in were reported 40% of the time. Wallets with money were reported 51% of the time.

Source: Cohn et al. (2019)

Appendix C

Here we present some other examples that we used in Round 3.

Example 1

Job satisfaction and occupational choice

	Intentions to leave teaching in the next 12 months (0 – 100)	
	Coefficient	P-Value
Teacher Job Satisfaction (0 – 10)	-7.97	0.00034
Dependent Variable Mean	30.60	
Dependent Variable Standard Deviation	36.96	

Source: Fullard (2023c)

Example 2

Inequality preferences and mental health

	Mental Health Index (0 – 100)	
	Coefficient	P-Value
Inequality Index (0 – 1) (Leaky Bucket Experiment)	-6.639	0.112
Dependent Variable Mean	51.67	
Dependent Variable Standard Deviation	21.39	

Source: Fullard (2023c)