Classroom Games in Economics :
A Quantitative Assessment of the `Beer Game'

Michael McMahon

No 964

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

Classroom Games in Economics: A Quantitative Assessment of the 'Beer Game'*

Michael McMahon
University of Warwick
m.mcmahon@warwick.ac.uk

April 22, 2011

Abstract

Using an experiment, I compare the use of the 'Beer Distribution' classroom game with the more traditional 'chalk and talk' approach to teach students about inventories and the macroeconomy. My empirical results confirm and extend our understanding of the relative strengths and weaknesses of the use of classroom games: the game tends to improve interest and motivation on average, though some students dislike their use; the game is effective at driving home its key messages, but it may wrongly lead students to disregard other important factors; the game is inferior where facts mastery or definitional learning is required. Rather than an endorsement or a criticism of classroom games, the conclusion is cautionary advice on how to best make use of games within an overall course.

Keywords: Classroom experiments and games; motivation; student learning outcomes.

JEL Codes: A22; C90.

*I would like to thank my EC108 students at the University of Warwick for being both my motivation and my subjects. I thank Robin Naylor for early comments on this work. I am also very grateful to Kai Riemer for making available his excellent Beer Game software for use in my class which generated the data which was used in this study. Any errors are my own.

1 Motivation

In the mid-1990s, Becker and Watts (1995) and Becker and Watts (1996) observed an apparent reluctance of economics faculty to adopt alternatives to the traditional 'chalk-and-talk' approach to third level education. Since then however, both in economics and other subjects I suspect, departures from pure lecture based courses are becoming more and more widespread and the use of classroom games and experiements is no exception.[1] In fact, the expanding field of experimental economics actually has its roots in Chamberlin (1948) and his use of classroom games and experiments to enhance learning (Fels 1993).

These games are often played in order to immerse the students in the economic environments (Holt 1999) and because games are seen as increasing student motivation (Whitton 2007b). The main benefit of them, relative to the more traditional 'chalk and talk' approach, is that they promote an active learning environment. As noted by Holt (1999), students learning actively tend to grow in terms of student interest and learning; Ball, Eckel, and Rojas (2006) assess the use of wireless handheld devices to facilitate in-class experiments and quizzes and greater active learning and find positive effects from this approach. Trigwell K. and Waterhouse (1999) find evidence that suggests that pure information conveying approaches by the lecture, in which students are more passive, tend to lead to surface, rather than deep, learning as characterised by Marton and Säljö (1976).

But given there is a cost to setting up and running a new form of class/lecture, and given that third level students may not necessarily benefit as much as younger (such as secondary school) students from the use of experiments and games in class (Whitton 2007b), it is important to consider the balance of costs and benefits in the use of such games more carefully. In order to pursue a scientific approach to evaluating our own teaching (as suggested by Wieman (2009)), we need to carry out an experiment on the use of experiments.[2]

The aim of this paper is to assess the extent of, and quantify, the benefits of one such classroom game on student outcomes in both the short- and medium-run. I make use of a controlled experiment using a large class (about 200 students) of relatively homogenous undergraduates. This approach allows me to compare the use of a classroom

[1]See for example, Greg Delemeester and Jurgen Brauer maintain a website called 'Games Economists Play: Non-Computerized Classroom-Games for College Economics' (`http://www.marietta.edu/~delemeeg/games/`) which contains over 170 classroom games for economists to play in class, and the Economics Network has published a variety of case studies which outline the instructions for a selection of economics classroom games and experiments (`http://www.economicsnetwork.ac.uk/themes/games`), as well as a full chapter on the subject in their Lecturer's Handbook (Balkenborg and Kaplan 2009). A recent issue of the *International Review of Economics Education* discusses a number of aspects of using games to teach economics (see Watts and Guest (2010) for an overview).

[2]The ironic lack of an experimental approach to evaluate these experiment-based teaching methods was noted by Fels (1993).

game to the more traditional teaching methods. Given the cost involved in establishing and setting up a new approach to a class, a necessary (but not sufficient) condition for any new approach to be worth it is that it generates positive benefits over the existing approach (a sufficient approach is that the total benefits, howsoever weighted and summed, exceed the total costs). I shall not address the difficult question as to how to weight different types of benefits but rather I simply compare the two teaching methods along many dimensions; I assess how the classical approach of teaching differs from the use of a classroom game in both the extent to which the method effectively achieves its learning outcomes as well as the extent to which participants prefer, or not, the teaching method employed.

Other papers have examined the benefits of using classroom experiments. Both Emerson and Taylor (2004) and Dickie (2006) employ an experimental design similar to mine, although the focus of the experiments in these papers is microeconomics, whereas my experiment is macroeconomic in nature.[3] Emerson and Taylor (2004) find that substituting a series of lectures with a series of different games boosted scores on standardised tests. Dickie (2006) finds a similar result which is boosted by the use of incentives or rewards to the experiments.

My objective is to assess the ability of one such experiment to influence knowledge about that specific subject matter. My study is not focused on the overall outcomes of the course as these other papers. It would be unrealistic to expect a single game session to alter the students learning over the entire course. Instead, similar to Lowry (1999), my focus is on the effects and outcomes of a single game.

Gremmen and Potters (1997) is probably closest in spirit to my work. They compare a macroeconomic game in terms of its relative effectiveness over a more traditional approach. Unlike their groups, I have a larger sample, and I also have the benefit of taking the same group of students and splitting them randomly over the different approaches. I, like them, try to compare two teaching approaches with the same objectives. However, unlike their approach, I allow for some objectives that do not necessarily favour games such as, in their words, ‘fact mastery’. I believe this allows me to assess the appropriateness of games as a direct substitution for certain types of learning objective.

Motivated by Durham, Mckinnon, and Schulman’s (2007) finding that classroom games and experiments help students retain economics knowledge, I explore both the immediate effect of the class and also the extent to which the knowledge is retained for an exam which takes place the next term.

To be precise, my analysis compares the use of the ‘Beer Distribution Game’ (described in detail below) to teach about inventory management and volatility, with a

[3]There remain fewer macroeconomic oriented classroom games although the numbers are growing. For example, the developments by Denise Hazlett.

more direct seminar version using prepared slides and class discussion. I find evidence, in line with previous research, that games tend to improve interest and motivation on average, although not everyone finds them such an improvement. In terms of educational outcomes, the game is very effective at driving home one of its key messages. However, I also find that it may wrongly lead students to disregard other important factors (factors whose significance is down-played to focus on the role of broader factors). Similarly, where facts mastery or definitional aspects are required, it may be important to supplement the games with slides or other ways of ensuring that students get firm guidance on the facts or definitions.

The bottom line is that while potentially very useful to supplement to a more standard 'chalk and talk' approach, it is necessary to carefully design and select the games used, and to ensure that every care is taken to avoid these potential pitfalls (for example, stressing the important role of the down-played factors in addition to the role of those factors that are emphasised).

The remainder of the paper is structured as follows. Section 2 discusses the two types of class I analyse as well as the common teaching outcomes I pursue. Section 3 presents my experimental, difference-in-difference approach to the analysis. Section 4 presents the data and my analysis of it. Section 5 discusses the findings and Section 6 concludes.

2 The Different Teaching Approaches

In this section I present my intended learning outcomes for the class, and then, separately for each type of class, the exact teaching method used.

2.1 Class Objectives and Intended Learning Outcomes

The subject matter covered was the behaviour of inventories in the macroeconomy. Students, by the end of the class, should be able to answer the following questions:

1. What are inventories and what is their role in the behaviour of the macroeconomy?
2. What is the 'Bull Whip' effect which is considered by other academic disciplines and practitioners?
3. How have inventory management techniques have changed? What does this mean for the behaviour of the business cycle?

Of course, there are many ways to answer some of these questions; for example, there are some models which admit no role for inventories in the macroeconomy and therefore improvement in inventory management techniques have no effect on business cycles. However, my interest is in whether the different teaching techniques have a differential influence on leading the students to think along a particular line. Moreover, as I described above, I will be able to use my pre-test to control for how they thought prior to the class session and therefore more clearly capture the effect of my class.

As a summary of the line of argument I take in the class, the following would be my summary answers to the above questions for the students:

1. Inventories are very important for the behaviour of the macroeconomy and account for about 45% of the volatility of real GDP growth in the US over the last 40 years.

2. Macroeconomists tend to focus on the speculative motives for holding inventories whereas other disciplines think more about the effect of supply chains and logistic/distribution issues in how inventories arise. The "bull whip effect" which arises in these other disciplines predicts that the further up the supply chain a firm is, the more volatile its activity (such as orders). This occurs even in the face of fairly stable customer orders and results from imperfect information flows, lags along the supply chain and a lack of coordination between elements of the supply chain.

3. Technological improvements driven by improved I.T. have led to huge changes in inventory management techniques. These changes, such as the adoption of Collaborative Planning and Forecasting, use of barcodes, scanners and Radio Frequency Identification (RFID) tags, have reduced informational problems, increased collaboration and even facilated reductions in the the length and delays along the supply chain. As such, in line with the idea of the 'bull whip' effect, such improvements reduce economic volatilty.

2.2 About the Seminar/Lecture Version

The seminar/lecture delivery is very standard. I used a set of 17 slides which, sequentially, worked through the questions above providing data (such as the contribution of inventories to average growth and the variance of that growth) and definitions (such as the explicit definition of inventories as used by the Bureau of Economic Analysis). I also discussed, using graphs generated in previous runs of the 'beer game', the concept of the 'bull-whip' effect and described the reasons that this arises. I concluded with a discussion of the changes in inventory management and what these mean for macroeconomic developments and especially the so-called 'Great Moderation'.[4]

[4] The 'Great Moderation' refers to the marked decline in macroeconomic volatility after 1983 in the US and other developed economies.

While I tried to ensure that the class remained interactive, it was slightly closer to a lecture style delivery than a typical seminar, but I did ask questions of the students during the session and there was considerable opportunity for them to ask questions about the material.

2.3 About the Game Version

The 'beer game' is a role-play simulation game used, particularly in business schools. It was originally by MIT's Sloan School (see Forrester (1961) and Sterman (1989)). It is used to allow students "to experience first hand the typical coordination problems of (traditional) supply chains, in which information sharing and collaboration does not exist" (Beer Game Website 2011).

In the 'beer game', students each take one role in a four stage supply chain that consists of a Brewery, Distributor, Wholesaler and Retailer:

- Producer (makes the item and sells to distributors)
- Distributors (buy from producers and sells to wholesalers)
- Wholesalers (buy from distributors and sells to retailers)
- Retailers (buy from wholesalers and sells to consumers)

These four units, and the actions and movements of goods and orders within each stage, are shown in Figure 1. Each unit is charged with supplying beer to meet orders of its customers while minimising their costs; the brewery produces the beer and sells it to the distributor, the distributor sells to the wholesaler, the wholesaler sells to the retailer and the retailer sells to the customer who has exogenous demand. Each unit has the same basic decision to make each period shown in the golden box in Figure 1; they place orders with the next upstream party (the brewery 'orders' an amount of production to undertake). Each unit fulfill open orders of beer out of the beer inventory that they have available for sale; these inventories are replenished when the unit receives fulfilled orders from their suppliers (the brewery gets newly produced beer).

There are two main sources of difficulty in the game; the first is that there are numerous logistics and production lags at each stage of the game, and the second is that communication and/or collaboration is not allowed between stages of supply chain. in terms of the first difficulty, there are many sources of delay. Order and production requests made in period t take a period to be processed and so the order only makes it upstream in period $t+1$ (brewery production requests take a period before the production enters the raw material phase). There are further delays in the process; once processed, the order takes a period to be transmitted and received by the supplier who then will ship as much beer as possible to fulfill any open orders but the deliveries of these orders are subject to 2 periods of shipping delay before they

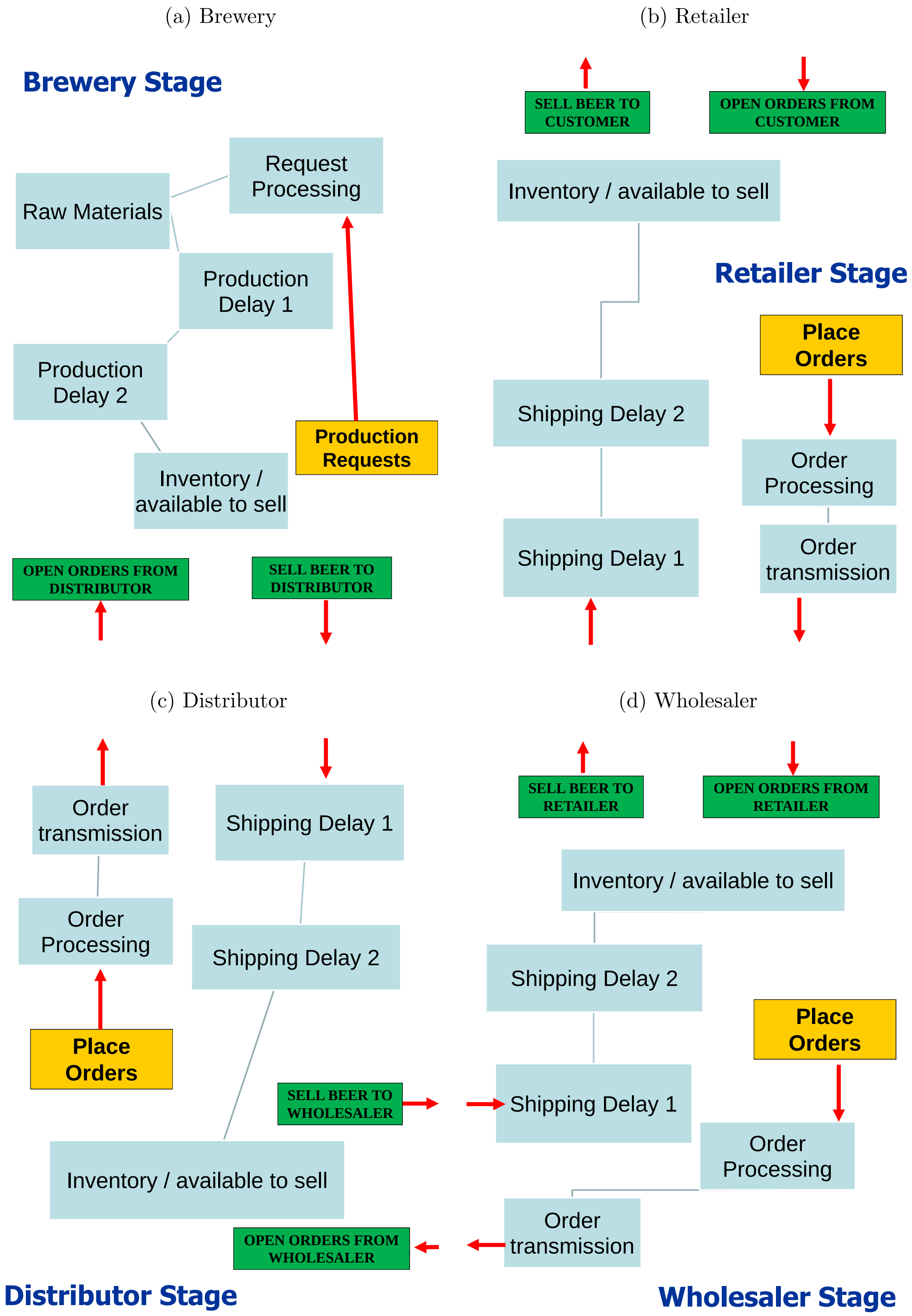

Figure 1: Beer Game stages and structure of distribution

are received (the brewery takes 2 periods to produce the beer once raw materials are prepared). Altogether, at each stage, an order placed at t will be received (subject to available supplier inventory) in period $t + 4$.

Once received, every order has to be fulfilled, either immediately (should the supplier have sufficient inventory) or later on; unfilled orders become an order backlog (which is effectively a negative inventory).[5] The only costs in this inventory management game relate to the holding of inventories (each case costs \$0.50) while case that the customer has ordered but that has not been dispatched, the backlog, costs \$1.00. Given that the objective is to minimise costs, the optimal strategy for the players is to run their business with as little stock as possible without being left unable to 'meet an order'.

On the second difficulty, the only information they are allowed to exchange is the order amount, and this is only transmitted upstream. There is no transparency as to what stock levels or other units demand is; only the retailer knows the demand of the final consumer. The demand of the final demand is exogenously controlled in the game. In fact for most of the uses of the game, it is predetermined and changes only once and permanently. I followed this typical pattern and orders were initially 5 cases of beer per period but then rising to 9 cases of beer per period after 5 periods where it stays for the remainder of the game (a total of about 40 periods are played though the players do not know this in advance as it would induce strategies to wind down stock before the end). In the beginning, the supply chain is pre-initialised with the right amount of inventory levels (15 units), orders (5 units) and beer units in the shipping delay fields (5 units) such that if every unit orders 5 units for the first 3 rounds, there is no accumulation or decumulation of inventory along the whole supply chain.

Despite this relatively stable consumer demand (a one off level shift), the supply chain invariably ends up generating huge volatility. Moreover, this volatility increases the further up the supply chain (from the consumer) is the unit; this is called the 'bull-whip' effect. Figure 2 shows typical results from the game and illustrates the 'bull-whip' effect. The black line shows the demand of the final consumer which increases once after period 5 and thereafter is constant. Though the retailer's order from the wholesaler (red line) show some volatililty, they are less volatile than the orders from the wholesaler to the distributor, and so on up the supply chain until the most volatile actions are the production requests taken by the brewery (blue line).

The game is played in rounds which each represent one period (typically considered a week). Each period, a student controlled unit will receive incoming orders, receive incoming deliveries, send out deliveries, and decide on the amount to be ordered. The main advantage of the computer-based game version is that the recording of outstanding inventory and backlog, as well as keeping track of what materials are at

[5]Customers are assumed to wait forever for their beer.

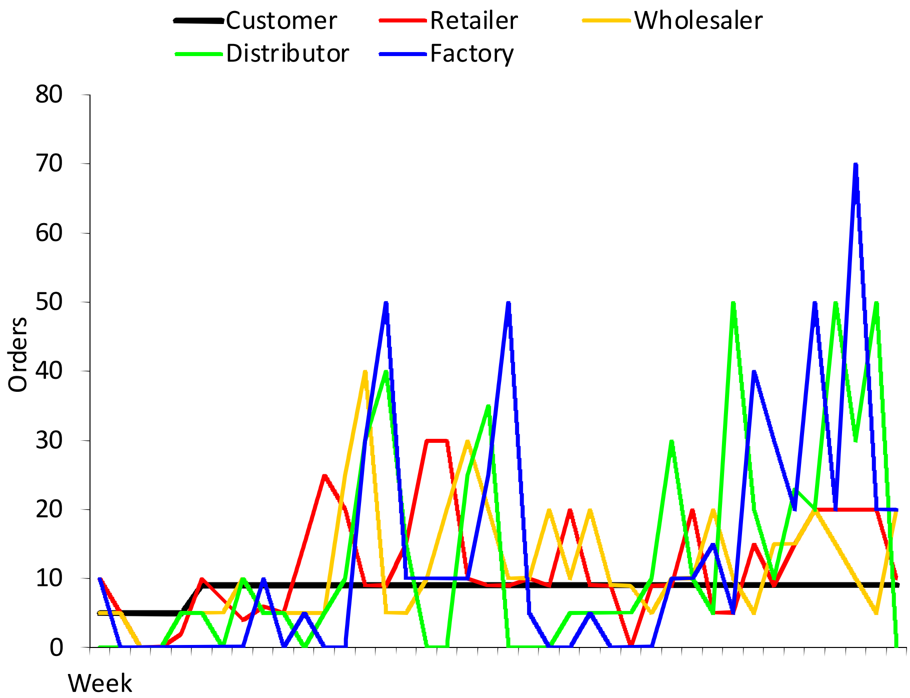

Figure 2: Typical Beer Game Data

what stages of the distribution chain, is all done by the computer which speeds up proceedings greatly.

When I ran the game for this experiment, the session began with an introduction simply of the rules, and then the players were allowed to play for about 30-40 rounds to experience the behaviour of the system. After this, I led a discussion of what happened (showing them data from their own play) and asking to discuss where the problems arose and how we might fix these problems. It was during this discussion that I attempted to introduce any concepts from the literature. I also asked then to think about inventory improvements such as email and GPS and barcodes and then consider how this would make the supply chain issues less great.

2.4 The costs associated with different types of teaching delivery

In each session, I covered the same material but just in a different fashion. As well as any differential benefits, the other key issue in choosing whether, or not, to adopt an alternative approach to delvering the material is the costs of different approaches. The costs include changeover and preparation costs (set-up costs), as well as delivery costs. In the case of my game, the set-up costs were only a few hours due to the availability of specialist software to implement the game. Nonetheless, a couple of hours might be enough of a fixed cost to disuade potential adopters if they do not feel that the benefits are sufficiently large. This is especially the case where exisiting resources for the more traditional 'chalk and talk' approach are available.

In terms of the delivery, however, there were significantly higher costs. Due to the requirement of having all students in front of their computer, I was limited by the size of the largest computer rooms available for teaching in the university. This would mean a maximum of 50 students at a time could complete the game; relative to the lecture version which can be presented to hundreds of people in a large lecture theatre at the same time, this would mean I would probably have to deliver 4 or 5 times as many hours. Of course, during the delivery, there was much less work for me to do and, apart from getting things going, I was silent for much of the session until the discussion at the end. In this regard, it might be possible for these sessions to be delivered by well-trained graduate students. Or, if they were replacements for seminar classes, then the group sizes would be already much more manageable.

3 Experimental Design

As I wish to analyse and quantify the extent of any difference between two different approaches to teaching, I pursue a simple difference-in-difference approach to the problem. In particular, I proceed using the following randomised experiment:

1. All of my EC108 class were instructed to sign up for one of 4 computer room sessions in order to cover some material for the EC108 course. No other detail was provided at this point.
2. This allowed me to create three groups of student with each broad group defined by the type of class that the student attended. Those who did not sign up and/or sign up for a class, and then those who signed up were allocated randomly to two groups (and each group was then split into two sessions for the purposes of the teaching). Therefore, I have the following three groups of student:

   Normal One of the 'sign-up' groups was taught the material on inventory behaviour via standard lecture / seminar format (lasting a single class session in total).

   Game The other 'sign-up' group played the 'beer game' and had a post-game discussion lead by me (again lasting a single class session).

   None Those who did not sign up were not provided access to the slides, nor the game.

3. On arrival at the computer room, students, regardless of the session they are in, take a short test to assess existing knowledge (success will be measured on a value-added basis). I invited those who did not sign up, and those who signed up

but didn't turn up, to complete the pre-class survey. A small group of students (10) did this.

4. The sessions then proceded according to the session specific format.
5. In the last few minutes of the session, the students retook the test, with a few additional questions, in order to assess learning in the short-term.
6. At the beginning of the next term (about 6 weeks later), all students took a mid-term test. This mid-term was part of the standard assessment used in the course and the only different was that on this occassion I included 5 questions from the pre- and post-class surveys in order to assess medium-term retained knowledge. Although they did not know it at the time, the students answers to these questions were recorded but did not count toward the final grade on the course.

At this point it is worth mentioning a few advantages of this approach. The classes, regardless of type, take the same amount of time - 50 minutes - and this is the standard allocation for such classes.[6] The students are all from the same course and were randomly assigned to the groups - this means that selection issues (see below) are of less concern and all students (about 200 of them in total) *should* have the same preparation in advance of the experiment. Finally, the topic is an important one for economics, but it is not covered in most undergraduate modules;[7] there are, therefore, no ethical considerations about preparing the treatment group more effectively for important exams or assessments.

There are two selection-bias concerns. The first concern the extent to which the allocation of session type is random. While the session a person signed up for was not random, I chose randomly what the teaching method would be in each session; I chose to teach the 4 sessions in the following order:

1. Seminar style
2. Game
3. Game
4. Seminar style

I chose this order because I worried that students may gravitate toward the later sessions, and away from the first sessions and so this ordering should ensure that I have a broadly similar total number of students (across the two sessions) in each type of session. As I made this decision before the students signed up, but the students were not aware either of the type of session or even the fact that there would be any differences across the sessions, I believe that the students should be randomly

[6]In fact, because of the pre- and post-sessional tests/surveys, the actual amount of time spent on the 'teaching' part of the sessions is less than 50 minutes.

[7]In fact Blinder (1981) notes that macroeconomic research pays too little attention to the behaviour of inventories.

assigned. Nonetheless, I shall include student fixed-effects to control for unobserved heterogeneity, and I show that the randomisation worked well.

The second concern is that the group of students who did not show up are the weakest / laziest students but these are also the students who would not fill in the pre-survey to allow them to enter my control group. Since I have no information on these students except the answers that they provide in the exam, I am limited in the analysis I can do with them.[8] This would particularly be of concern if we believed that the game format would be of most benefit to these students. On one hand, we might dismiss this concern because we are interested in the effective outcomes and so need to measure the effect on those students who would in practice attend regularly; in this way the study is *more* likely to reflect the real outcomes. On the other hand, it might be that games work through the attendance; knowing it was a game, attendance may increase boosting the total amount of students taught regardless of whether there is no change in the per student outcome. Studying this latter channel is beyond the scope of this paper but merits further work.[9]

This experimental design provides me with the information to compare the traditional learning group with those in the classroom game groups on both a short- and medium-term basis. This analysis is based on a difference-in-difference (diff-in-diff) approach to identify the different effects. The idea in this analysis is that we can write a person's knowledge (y_{it}) depends on factors such as natural ability and which type of class they attended, as well as their study effort before the exam. Table 1 expresses the knowledge (y_{it}) of different individuals over time using the following factors affecting measured knowledge:

- natural ability (λ_i) which is assumed constant across time;
- the average short-term effect of attending a class (γ^s);
- the average medium-term effect of attending a class (γ^m);
- the average short-term *marginal* effect of attending the game class (β^s);
- the average medium-term *marginal* effect of attending the game class (β^m);
- the average effect of exam study effort (e).
- individual- and time- specific deviations from these averages (ϵ_{it})

To illusatrate the power, of the diff-in-diff approach, consider a researcher trying to estimate the marginal effect of the 'Game' (β^s); looking at the 'Game' group on its own we cannot identify β^s uniquely as it would be mixed up with the effect of both the

[8]In particular, with only a single observation of these students, no person fixed effect can be included. However, I can examine these data cross-sectionally and I show that there is no average difference in their performance on questions unrelated to the subject matter being explored.

[9]Emerson and Taylor (2010) shows that, in an American University, the take-up of economics classes and choosing economics as a majors is overall unaffected by students being taught by classroom games rather than more traditional approaches. But this says nothing of the effect on the extent to which students attend during the courses that they choose.

Table 1: Table of Effects

Person	Group	Before Class ($t = 0$)	After Class ($t = 1$)	Exam ($t = 2$)
h	None	$\lambda_h + \epsilon_{h0}$		$\lambda_h + e + \epsilon_{h2}$
i	Normal	$\lambda_i + \epsilon_{i0}$	$\lambda_i + \gamma^s + \epsilon_{j1}$	$\lambda_i + e + \gamma^m + \epsilon_{j2}$
j	Game	$\lambda_j + \epsilon_{j0}$	$\lambda_j + \gamma^s + \beta^s + \epsilon_{i1}$	$\lambda_j + e + \gamma^m + \beta^m + \epsilon_{i2}$

average natural ability of group members and the effect of having any class. Diff-in-diff analysis is able to get around this problem since if the errors are independent and mean zero (so the average of errors is zero as long as I have sufficiently large number of students in each group), we can take the differences (Post-teaching - Pre-teaching) within both types of class (these are the first *'diff'* calculations and in the example here would be, for Normal and Game students respectively, γ^s and $\gamma^s + \beta^s$), and then take the difference of these two differences which allows us to identify the desired effect β^s.[10]

While this is the basic idea, it is often easier to use a regression approach rather than taking the group means and differencing appropriately. In particular, in a world of multiple treatment dummies, and multiple different time periods in which the treatments affect different individuals, the regression framework allows the estimation of a single regression, it becomes easier to get the standard errors, and it is possible to add extra correlates as control variables. As such, I proceed with such an approach in this paper and I estimate the following basic regression:[11]

$$\begin{aligned} y_{it} = &\alpha_i + \tau_1 \mathrm{T(Post)} + \tau_2 \mathrm{T(Exam)} + \eta_1 \mathrm{T(Post)} \times \mathrm{D(Game)} \\ &+ \eta_2 \mathrm{T(Exam)} \times \mathrm{D(Class)} + \eta_3 \mathrm{T(Exam)} \times \mathrm{D(Game)} + \varepsilon_{it} \end{aligned} \tag{1}$$

[10]The key identification assumption in the this analysis is that, absent the different type of class, the groups would have behaved similarly, on average. The randomisation and the use of invidual fixed-effects should address most concerns in this regard.

[11]Importantly, by including individual fixed effects, I do not need to include D(Class) and D(Game) on their own in the regression; the fixed effect captures these effects. Also, as I do not have Post-Teaching data on those that did not take part in a session, I do not need the interaction term T(Post) × D(Class).

where ε_{it} is an individual i, time t error term, and:

$$
\begin{aligned}
\text{T(Post)} &= \begin{cases} 1 & \text{if } t = 1 \\ 0 & \text{otherwise} \end{cases} \\
\text{T(Exam)} &= \begin{cases} 1 & \text{if } t = 2 \\ 0 & \text{otherwise} \end{cases} \\
\text{D(Class)} &= \begin{cases} 1 & \text{if type } = \text{`Normal' or `Game'} \\ 0 & \text{if type } = \text{`None'} \end{cases} \\
\text{D(Game)} &= \begin{cases} 1 & \text{if type } = \text{`Game'} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Using this regression, we can fully map each group mean into coefficients of equation (1) as shown in Table 2 using the assumption that $\mathbb{E}[\,\varepsilon_{it} = 0\,]$. This then allows us to map the regression coefficients to drivers of the educational outcomes which were our original interest; these are presented in Table 3. The exercise is now simply to examine the regression results for any outcome variable of interest to identify differences between the groups.

Table 2: Table of Effects in Terms of Regression Coefficients

Person	Group	Before Class ($t = 0$)	After Class ($t = 1$)	Exam ($t = 2$)
h	None	α_h		$\alpha_h + \tau_2$
i	Normal	α_i	$\alpha_i + \tau_1$	$\alpha_i + \tau_2 + \eta_2$
j	Game	α_j	$\alpha_j + \tau_1 + \eta_1$	$\alpha_j + \tau_2 + \eta_2 + \eta_3$

4 Data and Results Analysis

In order to measure outcomes in this experiment, participants were required to answer a series of questions. As outlined above, the participants were required to answer the questions both in advance of, and immediately after, the class session in which they partake, as well as in the midterm exam. Those who did not attend a class were given the chance to complete the pre-class form (although very few took this opportunity). In total, there are five questions to measure student's knowledge and beliefs about the behaviour of inventories, and there is a question to measure the students' view of the class session that they attended. In this section I shall beging by introducing the basic aspects of the data, and then, by question, I provide details of the question, the way in which I assess the extent to which the class achieves its objectives, and I then, while the specifics of the question are still fresh in the readers mind, complete the econometric analysis using the methodology laid out above.

Table 3: Mapping between regression coefficients and underlying factors of interest

Description	Symbol	Coefficient	Variable
Natural ability within each group	λ_k	α_k	FE
Average short-term effect of attending either class	γ^s	τ_1	T(Post)
Average medium-term effect of attending either class	γ^m	η_2	T(Exam) $\times$ D(Class)
Average short-term *marginal* effect of the game class	β^s	η_1	T(Post) $\times$ D(Game)
Average medium-term *marginal* effect of the game class	β^m	η_3	T(Exam) $\times$ D(Game)
Average effect of exam study effort	e	τ_2	T(Exam)

4.1 Summary Statistics

Table 4 shows how the sample is split between the different groups, as well some of the summary statistics of the final exam marks (which do not include the responses of the main inventories questions). In total there are 198 students for whom I have usable data. Not all of the students came to the extra sessions; 78 in fact missed it, and then only 10 filled in the pre-class form which I sent around. 62 students took the "Class" version of the experiment, and 58 took the "Game". In terms of exam marks, the statistics are very similar across groups; if anything, the mean is lower for those who participants in the "Game" session. But as I will include individual fixed-effects in most of the regressions, I do not need to worry about the effects of person-specific heterogeneity.

Table 4: Summary Statistics

Group	**Count** Total Students	**Mean** Exam Mark, %	**Median** Exam Mark, %	**Sd** Exam Mark, %
Just Exam	68	67	69	13
Form	10	67	69	13
Normal	62	65	69	14
Game	58	64	66	16
Total	198	65	69	14

Of course, the different groups have different amounts of data. Those for whom I have only exam data are represented only once. Those who did not attend a class but have filled in the Pre-class assessment form, have two entries per student. Finally, those who attended a class, have three entries per student. These data are represented in Table 5; in total, I have 448 responses to each of the main questions.

Table 5: Timing of the Observations

	Group				
Time of Response	**Just Exam**	**Form**	**Normal**	**Game**	**Total**
	No.	No.	No.	No.	No.
Before Class	0	10	62	58	130
After Class	0	0	62	58	120
Exam	68	10	62	58	198
Total	68	20	186	174	448

4.2 Question 1

Question 1 Which, if any, of the following make up the inventories as measured by the National Accounts?

- Firms work-in-progress
- Household purchases of baked beans for consumption later
- Firm holdings of oil for use in production
- A private households car
- Purcahses of cans of coke by Tesco for sale
- None of the above
- I dont know

This is a definitional question and according to the national accouts conventions of the Bureau of Economic Analysis, "the stock of private inventories consists of materials and supplies, work in process, finished goods, and goods held for resale. The change in private inventories is included in the NIPA measure gross private domestic investment." As it is part of investment, it does not include household purchases of finished goods which are not yet consumed. As such, the correct answers are 'Firms work-in-progress', 'Firm holdings of oil for use in production', and 'Purcahses of cans of coke by Tesco for sale'. The others, 'Household purchases of baked beans for consumption later' and 'A private households car', are incorrect.

To measure this answer, I create a score which also ensures that the 'none of the above answer' and the 'I don't know' are the least knowledgeable. This score, labelled $Q1s$, is given by the following equation:

$$
\begin{aligned}
Q1s = &\sum \text{correct} - \sum \text{incorrect} - 3 \times \text{`none of the above'...} \\
&\text{...} - 3 \times \text{`I don't know'}
\end{aligned}
\tag{2}
$$

I analyse the answers to this question in Table 6. In the first column I use student Fixed-Effects (FE) while in the second I use Random-Effects (RE). There are some small differences between the two estimated coefficients but a Hausman test suggests that there is no statistical difference between the coefficients; the null hypothesis of the test is that the difference in coefficients not systematic and we cannot reject it at reasonable levels of statistical significance ($Probability > \chi^2 = 0.41$). This suggests that using either FE or RE is suitable in analysising these data and confirms that the randomisation worked well.

The results from the analysis of the $Q1s$ variable suggest that, if anything, the classroom game is less good at conveying the definition of inventories; the η_1 coefficient on the 'T(Post) × D(Game)' interaction term is negative and marginally significant (at 11% level when using fixed effects). Despite me explaining the definition in the course

Table 6: Question 1 Analysis

	(1) Q1s	(2) Q1s	(3) D(Q1p)	(4) D(Q1p)
T(Post) - τ_1	0.89***	1.06***	1.45***	0.29***
	[0.00]	[0.00]	[0.00]	[0.00]
T(Exam) - τ_2	1.20**	1.35***	0.72	0.10
	[0.03]	[0.00]	[0.24]	[0.52]
T(Post) x D(Game) - η_1	-0.51	-0.89***	-1.20***	-0.24***
	[0.11]	[0.00]	[0.00]	[0.01]
T(Exam) x D(Class) - η_2	-0.30	-0.27	0.57	0.14
	[0.62]	[0.59]	[0.36]	[0.40]
T(Exam) x D(Game) - η_3	0.27	-0.11	-0.31	-0.052
	[0.40]	[0.68]	[0.30]	[0.56]
Constant	0.55***	0.55***	-1.84***	0.076**
	[0.00]	[0.00]	[0.00]	[0.01]
Observations	380	380	380	380
R-squared	0.178			0.123
Number of students	130	130	130	130
Estimation Method	Panel	Panel	Panel Probit	Panel
Student Effects	FE	RE	RE	FE

P-values reported in brackets
*** p<0.01, ** p<0.05, * p<0.1

of the discussion, it is likely that students' attention is wrongly focused on the type of inventory involved in the game. Notably, this lesser knowledge of the definition is offset students preparation for the exam (η_3 is not significant except at 40% statistical significance).

To explore further this 'facts mastery' aspect of teaching, I also test the probability that the participant gets the perfect answer by defining:

$$D(Q1p) = \begin{cases} 1 & \text{if } Q1^{score} = 3 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

To analyse this dichotomous dependent variable I use a Panel Probit model. As there is no suffficient ststistic for the conditional fixed-effect probit estimator, I proceed under the assumption of random effects in Column (3), although I also include a linear probability model with student fixed effects in Column (4). The result from the Q1s variable is confirmed using either econometric model. Looking at the implied average probabilities in Table 7, exam preparation adds, on average, about 10pp to probability of being correct, while the normal class raises the probability of being of being correct by 32pp. The game class has very little effect on the probability of being correct

Table 7: Probability that $D(Q1p) = 1$

	Time of Response			
Group	**Before Class**	**After Class**	**Exam**	**Total**
	Prob	Prob	Prob	Prob
Form	0.03		0.13	0.08
Normal	0.03	0.35	0.29	0.22
Game	0.03	0.06	0.19	0.09

(+3pp). The medium term effect of the class is, however, smaller.

4.3 Question 2

Question 2 How important do you think inventory changes are in affecting the business cycle?

- "Very Important"
- "Quite Important"
- "Somewhat Important"
- "Unrelated"
- "I don't know"

As part of the objective of the class is to emphasise the importance of inventories in the business cycle, the ideal outcome of the teaching is to lead students to answer either "Quite Important" and more precisely "Very Important". I create a score, ($Q2s$), which ranges from "I don't know" being worth 0, "Unrelated" counts as 1 and then linearly in unit increments to "Very Important" being worth 4. A higher score is an indication that the student believes that inventories are more important for the business cycle.

Additionally, I can create a dichotomous variable according to:

$$D(Q2c) = \begin{cases} 1 & \text{if "Very Important" or "Quite Important"} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The results, shown in Table 8, indicate that while attending either class type increases the student belief that inventories are important, there is no difference between the game and normal class. This effect does not, however, last to the exam by which the people who attended no class give the same average answer. In terms of probability of high importance, the effect of attending a class is only significant at the 20% significance level.

Table 8: Question 2 Analysis

	(1) Q2s	(2) D(Q2c)
T(Post) - τ_1	0.47***	0.34
	[0.00]	[0.17]
T(Exam) - τ_2	0.10	0.032
	[0.77]	[0.95]
T(Post) x D(Game) - η_1	0.12	0.043
	[0.56]	[0.89]
T(Exam) x D(Class) - η_2	0.32	0.39
	[0.40]	[0.48]
T(Exam) x D(Game) - η_3	0.20	0.16
	[0.32]	[0.63]
Constant	2.92***	0.85***
	[0.00]	[0.00]
Observations	380	380
R-squared	0.131	
Number of students	130	130
Estimation Method	Panel	Panel Probit
Student Effects	FE	RE

P-values reported in brackets
*** p<0.01, ** p<0.05, * p<0.1

4.4 Question 3

Question 3 Please rate the following in terms of importance for generating changes in the levels of inventory held by firms. Rate each one separately:

	Very Important	Quite Important	Somewhat Important	Unrelated	I dont know
a lack of information	4	3	2	1	0
variable consumer demand	4	3	2	1	0
a lack of collaboration	4	3	2	1	0
production delays	4	3	2	1	0
structure of the supply chain	4	3	2	1	0

First, for each answer, I create a variable that plots the score according to the table above. I, therefore, have five variables ($q3^{\text{info}}$, $q3^{\text{consumer}}$, $q3^{\text{nocollab}}$, $q3^{\text{delays}}$ and $q3^{\text{supplychain}}$) that measure scores and can be used to assess the effect of different teaching methods on students percewptions of the importance of different sources of volatilty. This regression results using these five variables are reported in Columns (1) to (5) of Table 9.

Table 9: Question 3 Analysis

	(1) $q3^{\text{info}}$	(2) $q3^{\text{consumer}}$	(3) $q3^{\text{nocollab}}$	(4) $q3^{\text{delays}}$	(5) $q3^{\text{supplychain}}$	(6) D($q3^{all}$)	(7) D($q3^{\text{cons unrelated}}$)
T(Post) - τ_1	0.56***	-0.34*	0.68***	0.23	0.065	0.67***	0.38
	[0.00]	[0.05]	[0.00]	[0.14]	[0.70]	[0.00]	[0.30]
T(Exam) - τ_2	-0.30	-0.10	-0.40	0.20	0.20	-0.46	0.70
	[0.47]	[0.82]	[0.32]	[0.60]	[0.64]	[0.31]	[0.30]
T(Post) x D(Game) - η_1	0.61**	-0.35	0.79***	0.33	0.38	0.33	0.33
	[0.01]	[0.17]	[0.00]	[0.14]	[0.12]	[0.26]	[0.41]
T(Exam) x D(Class) - η_2	0.30	-0.13	0.87**	-0.41	-0.52	0.61	-1.04
	[0.50]	[0.79]	[0.04]	[0.32]	[0.25]	[0.20]	[0.19]
T(Exam) x D(Game) - η_3	0.43*	-0.50**	0.84***	0.30	0.22	-0.13	1.34**
	[0.07]	[0.05]	[0.00]	[0.18]	[0.37]	[0.61]	[0.02]
Constant	2.82***	3.42***	2.02***	3.09***	2.90***	0.13	-2.16***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.29]	[0.00]
Observations	380	380	380	380	380	380	380
R-squared	0.207	0.090	0.318	0.084	0.066		
Number of students	130	130	130	130	130	130	130
Estimation Method	Panel	Panel	Panel	Panel	Panel	Panel Probit	Panel Probit
Student Effects	FE	FE	FE	FE	FE	RE	RE

P-values reported in brackets
*** p<0.01, ** p<0.05, * p<0.1

Although the Bull-whip effect that was discussed in the class actually emphasises that all of these things are important, the main effects of the game are to stress the informational sources of volatility; Column (1) and (3) both show significantly more emphasis placed by game students on the role of lack of information and the fact that collaboration is not allowed (η_1 is positive and significant in both cases). This emphasis continues to the exam, even though students who attended the normal class also stress the lack of collaboration in the exam relative to those who attended no class.

In the beer game, in order to stress the role of the other factors, consumer demand is very stable. In fact, when students finish the game they often believe that consumer demand (which has up until then only been know by the retailer) is very volatile; the revalation of the actual volatility is usually very surprising. The downside of this fact is evident in the results in Column (2); after the class, both the groups put less importance on the consumer demand channel but by the exam, the game students are the ones who continue to believe that the channel is less important; η_3 in column (2) is negative and statistically significant.

Although important, the role of the supply chain and the delays that arise within it are less actively discussed at the end of the game session. It appears that students do not seem to think differentially on the role that these factors play (although the mean answer is that they are 'quite important' (Columns (4) and (5)).

I can complete two cross-checks on results above with Question 3. First, I can check the effect of the different teaching methods on the probability that all are at least somewhat important:

$$D(q3^{all}) = \begin{cases} 1 & \text{if all at least `Somewhat Important'} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Second, because one concern would be that the beer distribution game, by emphasising the other channels, leads students to wrongly believe that consumer demand is unimportant, I also examine the likelihood that a student thinks that consumer demand does not generate any volatility using:

$$D(q3^{\text{cons unrelated}}) = \begin{cases} 1 & \text{if } q3^{\text{consumer}} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The results are reported in Columns (6) and (7) in Table 9, while the implied probabilities from the regressions are reported in Tables 10 and 11. While the class is successful at boosting student beliefs that all are important, the effects are very short-term and do not persist to the exam. More worryingly, the game causes a significant increase in the medium-term probability that a student wrongly considers consumer demand unimportant. The probability rises from 2% before the class, to 12% in the exam if the student has taken the game class; if the student had taken the normal class, the probabilit woould be around the same at the exam as before the class.

Table 10: Probability that $D(Q3^{all}) = 1$

	Time of Response			
Group	**Before Class**	**After Class**	**Exam**	**Total**
	Prob	Prob	Prob	Prob
Form	0.55		0.37	0.46
Normal	0.55	0.79	0.61	0.65
Game	0.55	0.87	0.56	0.66

Table 11: Probability that $D(q3^{\text{cons unrelated}}) = 1$

	Time of Response			
Group	**Before Class**	**After Class**	**Exam**	**Total**
	Prob	Prob	Prob	Prob
Form	0.02		0.07	0.04
Normal	0.02	0.04	0.01	0.02
Game	0.02	0.07	0.12	0.07

4.5 Question 4

Question 4 Imagine a supply chain for a drink comprising a producer, distributor, wholesaler and retailer. Please rate the following in terms of how volatile activity (production or orders for higher up in the chain) is in that part of the supply chain. Rate each one separately:

	Most Volatile	Average Volatility	Less Volatile	No Volatility	I dont know
Retailers	4	3	2	1	0
Wholesalers	4	3	2	1	0
Distributors	4	3	2	1	0
Producer	4	3	2	1	0

According to the Bull-Whip effect, the volatility increases as we move up the supply-chain with producers being more volatile than retailers. Therefore, I define a simple measures of this effect, , Q4s, which is a score which measures the volatility gap between producers and retailers (conditional on an answer to both parts of the question, this variable can range from 3 to -3):

$$D(q4^{score}) = \text{Producer Score} - \text{Retailer Score} \quad (7)$$

The results of this analysis are shown in Table 12. In this case the results are more favourable for the game approach. While both approaches boost the student belief in the key prediction of the bull-whip effect in the short-term (though not statistically significantly so at usually acceptable levels of significance), the effect of the game class is long-lasting such that it is revealed in the answers on the exam.

Table 12: Question 4 Analysis

	(1) Q4s
T(Post) - τ_1	1.45***
	[0.00]
T(Exam) - τ_2	1
	[0.23]
T(Post) x D(Game) - η_1	0.28
	[0.57]
T(Exam) x D(Class) - η_2	-1.19
	[0.19]
T(Exam) x D(Game) - η_3	1.36***
	[0.01]
Constant	-0.48***
	[0.00]
Observations	373
Number of students	130
R-squared	0.184
Estimation Method	Panel
Student Effects	FE

P-values reported in brackets
*** p<0.01, ** p<0.05, * p<0.1

4.6 Question 5

Question 5 Do you think improvements in inventory management techniques have made a difference to macroeconomic outcomes? Please select one of the the following answers:

- Yes

- No
- I don't know

This question measures the effect that the different teaching methods have had on student perceptions of the cause of macroeconomic volatilty. In both classes I discussed the conflicting views about whether inventories have caused macroeconomic volatility (as the bull-whip effect suggests and as it is argued in some of the literature), or whether macroeconomic volatility cause the volatility of inventories (as is prevalent in more recent macroeconomic models of inventories). This question lends itself to a dichotomous variable analysis on the likelihood that the students think that inventory improvements had an effect on macroeconomic volatility:[12]

$$D(Q5^{Yes}) = \begin{cases} 1 & \text{if "Yes"} \\ 0 & \text{if "No"} \end{cases} \tag{8}$$

I do not provide a regression analysis of this variable because as is evident from Table 13, there is very little variation across groups in their answer to this question. It seems that most students believe, regardless of attending a class, or not, that inventory improvements would affect macroeconomic volatility.

Table 13: Q5

	Time of Response			
Group	**Before Class**	**After Class**	**Exam**	**Total**
	Prob	Prob	Prob	Prob
Form	0.90		0.70	0.80
Normal	1.00	1.00	0.94	0.98
Game	0.98	0.98	0.97	0.97

4.7 Questions to Measure Student Views on the Teaching Method

In the survey completed at the end of the class sessions, I included a four-part question in which students could rate the class they had just finished relative to the usual lecture sessions that I gave for the course. One of the 120 students who took part in the classes failed to fill in the extra questions and therefore I lose this observation for this part of the analysis. The question that they were asked was:

Session Appraisal Please rate this session compared to our ordinary EC108 sessions. Rate each one separately:

[12] For this, I need to drop those few who wrote 'Unknown'.

	Much Better	Slightly Better	No Different	Slightly Worse	Much Worse
Overall session (S1)	5	4	3	2	1
Delivery and engagement (S2)	5	4	3	2	1
Content (S3)	5	4	3	2	1
Your interest in the material (S4)	5	4	3	2	1

The answers to these questions allow me to create variables which measure, for each of the four categories surveyed, the score given by the student; $S1$ is the score for the overall session, $S2$ for delivery and engagement, $S3$ for content and $S4$ for interest in the material. The first analysis is simply to look at the means of the variables. Table 14 shows that the mean answer to all four questions is higher in the Game group.

Table 14: Mean Responses to Session Evaluation Questions

Group	**Mean**			
	S1	S2	S3	S4
Normal	3.8	3.8	3.5	3.7
Game	4.1	4.2	3.7	4.2
Total	3.9	4.0	3.6	3.9

The problem with these variables is that they are only measured once (at the end of the class), and therefore, as there is no time-series data, I can only look at cross-sectional comparisons. This means that I cannot control for unobserved heterogeneity using fixed effects but since the earlier results suggested that the randomisation worked well, this is only of secondary concern. I estimate equation 9 on the cross-section of data using each of the scores twice; the first regression includes no controls, while the second includes as controls the students exam mark (as %) on the non-experiment related questions of the test, as well as session fixed effects to capture any group specific averages.[13]

$$Sj_i = \alpha + \beta_1 \mathrm{D(Game)} + \beta_2 \mathrm{Controls} + \epsilon_i \quad j \in (1, 2, 3, 4) \tag{9}$$

The results in Table 15 generally confirm the earlier results from Table 14. Notably, however, the effects of the game are only statistically significant in terms of the better delivery and engagement (S2) as well as in generating greater interest in the material (S4). The overall view (S1) was not statistically different and there was no significant difference in the views about the underlying macroeconomic content (S4).

[13] As there are only 4 sessions, the constant and the D(Game) variable capture the effects of Session 1 and Session 2. Therefore, I need only include dummy variables for Session 3 and Session 4.

Table 15: Session Question Analysis

VARIABLES	(1) S1	(2) S1	(3) S2	(4) S2	(5) S3	(6) S3	(7) S4	(8) S4
D(Game)	0.30	0.37	0.37**	0.55**	0.25	0.31	0.53***	0.48**
	[0.11]	[0.13]	[0.02]	[0.02]	[0.15]	[0.21]	[0.00]	[0.05]
D(Session 3)		-0.59**		-0.44*		-0.41		-0.14
		[0.03]		[0.06]		[0.12]		[0.56]
D(Session 4)		-0.44*		-0.086		-0.28		-0.24
		[0.06]		[0.69]		[0.18]		[0.33]
Exam Mark, %		-0.0093*		-0.010**		-0.011**		-0.00015
		[0.08]		[0.04]		[0.05]		[0.98]
Constant	3.80***	4.65***	3.82***	4.53***	3.49***	4.33***	3.66***	3.80***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	119	119	119	119	119	119	119	119
R-squared	0.022	0.101	0.043	0.100	0.018	0.078	0.075	0.085
Estimation Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Student Effects	None	None	None	None	None	None	None	None

P-values reported in brackets

*** p<0.01, ** p<0.05, * p<0.1

Perhaps surprisingly, these positive effects were limited to the first Game group (Session 2). Once we control for the session specific dummy variables (the even-numbered columns in Table 15), there is no difference between the second Game group (Session 3) and the normal class groups. Table **??** repeats Table 14 but by session rather than group and shows that Session 2 is the session in which people are most positive about the Game. Figures 3 and 4 shows the scores across sessions for the two dimensions along which the Game appears to perform better - S2 (Delivery and engagement) and S3 (Content). The striking difference is the percentage of people who report that session 2 was "much better".

Table 16: Mean Responses to Session Evaluation Questions, by specific group (1-4)

Session	**Mean**			
	S1	S2	S3	S4
1st Class (1)	4.0	3.9	3.6	3.8
1st Game (2)	4.4	4.4	4.0	4.3
2nd Game (3)	3.8	4.0	3.6	4.1
2nd Class (4)	3.6	3.8	3.4	3.5
Total	3.9	4.0	3.6	3.9

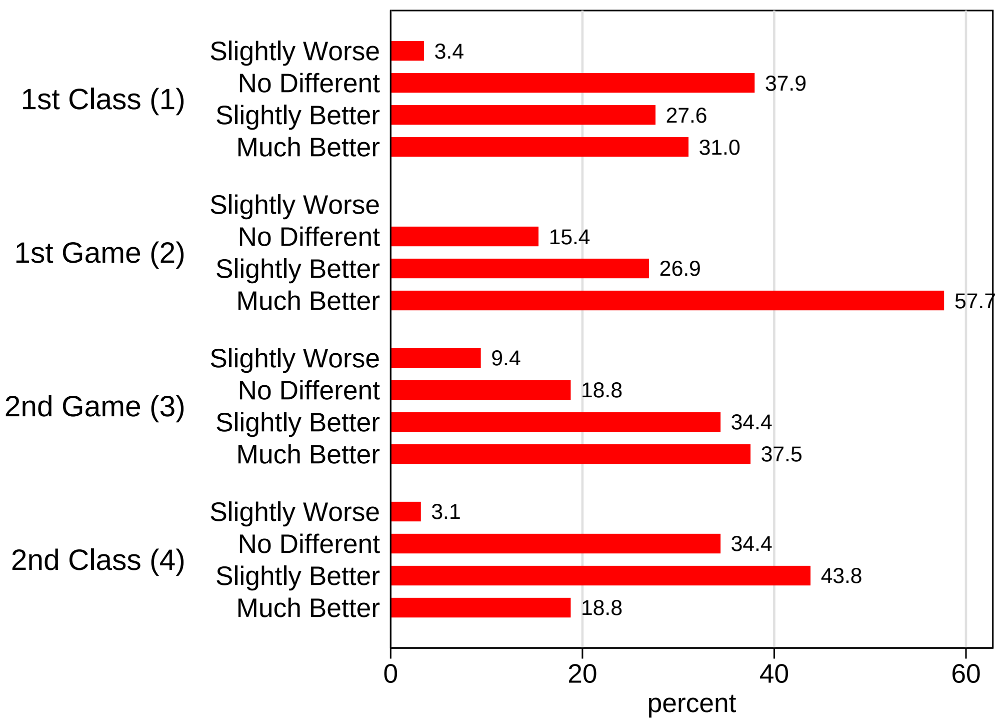

Figure 3: Delivery and engagement (S2) scores across session

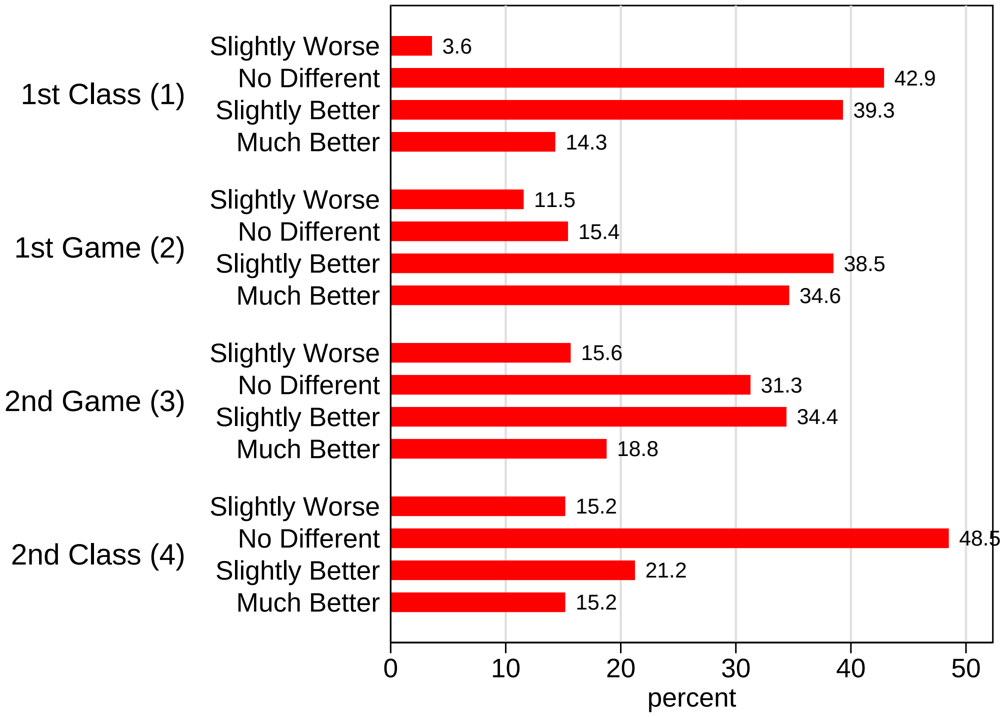

Figure 4: Content (S3) scores across session

This is particularly surprising given that, after the 4 sessions, I felt that the second game session (session 3) had run more smoothly than the first (session 2). In the first game session there had been some set-up technical difficulties with getting access for many of the students; these were resolved but it took about 10 minutes to do it meaning the group had less time to play the actual game. By the second game session, I knew the work around to the access problems and I could therefore get up and running immediately. These issues were reflected in the open comments section of the post-class form. For example, one student noted of the first game session that it "would be awesome if not for the techical issues" and another said "needs to be tested first"; both these students still reported the session as "Much Better" than our usual sessions. A different student regarded the second game session as "Slightly Worse" and openly stated that "Lectures are better".

5 Discussion

These empirical results confirm some previous findings while also extending our understanding of the relative strengths and weaknesses of the use of classroom games for teaching specific things. Motivation seems to increase on average; this is in line with previous beliefs and findings such as Whitton's (2007a) in-depth interview study in which she finds that older students, though not necessarily getting intrinsic motivation from the game, are reasonably positive about their use. But it appears that not everyone finds them such an improvement and that group dynamics may play a role; it was not obvious to me that the groups were experiencing any greater problems in the second game session, but the potential role for any group-specific dynamics warrant further investigation.

The fact that the game is more effective at driving home the key idea of the 'bull-whip' effect (Question 4 analysis) is confirmation that such games can be very useful additions to make particular points. However, as shown in the answers to Question 3, the game may have wrongly left students with the impression that some aspects are unimportant. Both of these points suggest the careful design and selection of games is necessary in order to ensure students get the right message and don't over-emphasise one aspect, at the expense of other aspects, of the teaching.

Similarly, where facts mastery or definitional aspects are required, it may be important to supplement the games with slides or other ways of ensuring that students get firm guidance on the facts or definitions. Again, careful selection of games that are appropriate for purpose is required; games and other teaching approaches appear to have different strengths in terms of Bloom's (1956) cognitive levels.

One concern may be that the effects are different between stronger and weaker students; Emerson and Taylor (2004) find that experiments benefit weaker students more. I have also analysed the effects of the experiments on those who had lower grades on the non-inventory material in the exam by splitting the marks on the final exam into those who scored above the median (68.75%) and those weaker students who scored below it; there are 68 strong students and 62 weak students.[14] I compare the regressions of both the upper and lower halves of the group. A selction of the results are reported below in Tables 17 for the effect on outcomes and Table 18 for the effect on student satisfaction.[15] The odd-numbered columns report results for the upper half of the class, while the even-numbered columns report the results for the weaker half of the class.

I find that, if anything, the results are strongest for the better students; perhaps these students don't feel pushed by the usual lectures and seminars and therefore enjoyed the challenge of the game more. There is less evidence that the students who subsequently performed weaker on the test found the game sessions better and the other results are similar. Of course, I am limited in comparing stronger and weaker students within the reasonably homogenous group of first year economics undergraduates; all students will have received very similar, and high, grades at school (the typical offer for the L100 economics degree at Warwick for this intake was at least AAA at A-level (including an A in Maths) or 38 points at IB (including a 6 in higher Maths)).

[14]The total number of students is 130 made up of 120 from the classes and 10 from the not turn up group.

[15]The full results are not reported here in the interests of space but they are available on request.

Table 17: Investigating a Weaker Student Effect on Outcomes: Comparision of effect with full responses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Q1s	Q1s	Q2s	Q2s	D($Q3^all$)	D($Q3^all$)	Q4s	Q4s
T(Post) - τ_1	0.85***	0.93**	0.55***	0.38*	0.67**	0.70**	1.32***	1.60***
	[0.00]	[0.01]	[0.00]	[0.08]	[0.05]	[0.02]	[0.01]	[0.00]
T(Exam) - τ_2	0.83	1.75*	0	0.25	-0.78	-0.13	1	1
	[0.20]	[0.07]	[1.00]	[0.67]	[0.25]	[0.84]	[0.35]	[0.46]
T(Post) x D(Game) - η_1	-0.61	-0.41	-0.063	0.31	0.72	0.0016	0.99	-0.45
	[0.14]	[0.41]	[0.82]	[0.31]	[0.13]	[1.00]	[0.15]	[0.53]
T(Exam) x D(Class) - η_2	0.29	-1.09	0.55	0.026	1.21*	0.051	-1.21	-1.17
	[0.69]	[0.29]	[0.25]	[0.97]	[0.10]	[0.94]	[0.30]	[0.41]
T(Exam) x D(Game) - η_3	0.12	0.45	0.041	0.38	-0.19	-0.086	1.90***	0.83
	[0.77]	[0.38]	[0.88]	[0.22]	[0.63]	[0.80]	[0.01]	[0.24]
Constant	0.84***	0.23	2.96***	2.88***	0.14	0.12	-0.48**	-0.48*
	[0.00]	[0.19]	[0.00]	[0.00]	[0.47]	[0.45]	[0.04]	[0.05]
Observations	198	182	198	182	198	182	194	179
R-squared	0.238	0.148	0.148	0.126			0.234	0.153
Number of students	68	62	68	62	68	62	68	62
Estimation Method	Panel	Panel	Panel	Panel	Panel Probit	Panel Probit	Panel	Panel
Student Effects	FE	FE	FE	FE	RE	RE	FE	FE
Student Percentile	>50th	<50th	>50th	<50th	>50th	<50th	>50th	<50th

P-values reported in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Investigating a Weaker Student Effect on Enjoyment: Comparision of effect with full responses

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	S1	S1	S2	S2	S3	S3	S4	S4
D(Game)	0.77**	-0.048	0.75**	0.36	0.53	0.082	0.35	0.58*
	[0.03]	[0.88]	[0.03]	[0.27]	[0.12]	[0.82]	[0.35]	[0.08]
D(Session 3)	-0.82**	-0.35	-0.68**	-0.16	-0.53	-0.26	0.034	-0.30
	[0.04]	[0.35]	[0.05]	[0.62]	[0.14]	[0.51]	[0.93]	[0.32]
D(Session 4)	-0.34	-0.58*	-0.044	-0.13	-0.23	-0.35	-0.26	-0.22
	[0.33]	[0.07]	[0.88]	[0.69]	[0.45]	[0.28]	[0.43]	[0.56]
Constant	3.81***	4.33***	3.75***	4***	3.47***	3.85***	3.73***	3.85***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	62	57	62	57	61	58	61	58
R-squared	0.126	0.067	0.105	0.049	0.082	0.032	0.072	0.108
Estimation Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Student Effects	None	None	None	None	None	None	None	None
Student Percentile	>50th	<50th	>50th	<50th	>50th	<50th	>50th	<50th

P-values reported in brackets

*** p<0.01, ** p<0.05, * p<0.1

6 Conclusion

The main conclusion of this paper is that while positive in some respects, experiments and games for teaching use need to be carefully adopted. Their use must be integrated into courses with due consideration for the teaching objectives and what the games are effective at delivering. Moreover, the appreciation of the change in teaching approach by the average game student may be subject to diminishing returns; a few games might be welcome breaks in teaching style and therefore help motivate students and reinvigorate their interest in the subject, but too many and the effects may be diminished.[16]

One aspect that I could not explore in this study is the role of lecturer delivery in determining the benefits of the games. As I mentioned above, I had less active delivery in the game class compared to the normal class. This means that any passion and enthusiam that I can transmit in a lecture would be harder to get across as the students worked through the game without my direct involvement. In some cases, however, less lecturer involvement may be beneficial! If it is more difficult to make mistakes or to be boring in the game version, there may be large gains over lectures for weaker lecturers; similar arguments are put forward in Fels (1993). It may also be a way to equalise the delivery of classes by graduate teaching assistants.

Finally, while I have tried to keep this study focused on particular objectives, it cannot be considered a complete evaluation of all possible uses for classroom game and experiments. For example, I cannot perfectly measure the costs and if repeated, the second and future times the costs would surely decrease. Moreover, as suggested by Fels (1993), a complete and comphrensive evaluation would ideally require a multi-year, multi-course, multi-lecturer design and proper measurement of the costs relative to the lecturer costs including follow-up time spent on answering student questions, etc. Additionally, to measure the long-term benefits accurately, the study would extend into future years of study and performance in later courses (Watts and Guest 2010).

Notwithstanding these issues, my empirical results confirm some previous findings while also extending our understanding of the relative strengths and weaknesses of the use of classroom games for teaching specific teaching outcomes. This paper represents niether an unconditional endorsement of classroom games, nor a criticism or rejection of their use. Rather, the message of this paper is cautionary - *caveat utilitor* (let the user beware).

[16]It may be that the use of first year students, in their first term, did not appreciate this break in teaching style (as they are still adjusting to the new lecture approach) and so my results underplay the motivational effect.

References

Balkenborg, D., and T. Kaplan (2009): "Economic Classroom Experiments," in *The Economics Network Handbook for Economics Lecturers*, ed. by P. Davies.

Ball, S. B., C. Eckel, and C. Rojas (2006): "Technology Improves Learning in Large Principles of Economics Classes: Using Our WITS," *American Economic Review*, 96(2), 442–446.

Becker, W. E., and M. Watts (1995): "Teaching tools: teaching methods in undergraduate economics," *Economic Inquiry*, 33(4), 692–700.

——— (1996): "Chalk and Talk: A National Survey on Teaching Undergraduate Economics," *The American Economic Review*, 86(2).

Beer Game Website (2011): "http://www.beergame.org/," .

Blinder, A. S. (1981): "Retail Inventory Behavior and Business Fluctuations," *Brookings Papers on Economic Activity*, 12(1981-2), 443–505.

Bloom, B. S. (1956): *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain.* Addison Wesley Publishing Company, Boston, MA, USA.

Chamberlin, E. H. (1948): "An Experimental Imperfect Market," *The Journal of Political Economy*, 56(2), 95–108.

Dickie, M. (2006): "Do Classroom Experiments Increase Learning in Introductory Microeconomics?ry Microeconomics," *Journal of Economic Education*, 37(3), 267–288.

Durham, Y., T. Mckinnon, and C. Schulman (2007): "Classroom Experiments: Not Just Fun And Games," *Economic Inquiry*, 45(1), 162–178.

Emerson, T. L. N., and B. A. Taylor (2004): "Comparing Student Achievement across Experimental and Lecture-Oriented Sections of a Principles of Microeconomics Course," *Southern Economic Journal*, 70(3), 672–693.

——— (2010): "Do Classroom Experiments Affect the Number of Economics Enrolments and Majors?," *International Review of Economics Education*, 9(2), 42–58.

Fels, R. (1993): "This Is What I Do, and I Like It," *The Journal of Economic Education*, 24(4), pp. 365–370.

Forrester, J. (1961): *Industrial Dynamics.* MIT Press, Cambridge, MA.

Gremmen, H., and J. Potters (1997): "Assessing the Efficacy of Gaming in Economic Education," *The Journal of Economic Education*, 28(4), pp. 291–303.

Holt, C. A. (1999): "Teaching Economics with Classroom Experiments: A Symposium," *Southern Economic Journal*, 65(3), 603–610.

Lowry, P. E. (1999): "Model GATT: A Role-Playing Simulation Course," *Journal of Economic Education*, 30(2), 119–126.

Marton, F., and R. Säljö (1976): "On qualitative differences in learning. I Outcome and Process," *British Journal of Educational Psychology*, 46(1), 4–11.

Sterman, J. (1989): "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiemnt," *Management Science*, 35(3), 321–339.

Trigwell K., P. M., and F. Waterhouse (1999): "Relations Between Teachers' Approaches to Teaching and Students' Approaches to Learning," *Higher Education*, 37, 57–70.

Watts, M., and R. Guest (2010): "Experimental Economics and Economic Education: Editorial," *International Review of Economics Education*, 9(2), 6–9.

Whitton, N. (2007a): "An investigation into the potential of collaborative computer game-based learning in Higher Education," Ph.D. thesis, Napier University.

——— (2007b): "Motivation and computer game based learning," in *ICT: Providing choices for learners and learning: Proceedings ascilite Singapore 2007.*

Wieman, C. (2009): "A new model for transforming the teaching of science in higher education," `http://www.dit.ie/media/newsdocuments/2009/ANewModel.ppt`.