Evaluation of an Economics Classroom Experiment as a Teaching Method

Name:	Joris Korver
Student number:	850207469020
Date:	10 - 02 - 2012
Course Code:	ECH 80436
Supervisors:	Gerrit Antonides
	Harm Biemans
	Pierre van Mouche

Summary

Interest in experimental economics is increasing and with it the use of classroom experiments. In undergraduate economics classes, the possibility of a positive teaching effect of classroom experiments has been shown. Students taught using experiments scored at least as well on tests as students who attended conventional classes on the same subject. Students also enjoyed experiments more than conventional teaching methods. This makes classroom experiments attractive for use at secondary school as well. The teaching effect of classroom experiments on secondary school students had not been tested before, this was done in this research. Secondary school economics students either participated in a class with an experiment as a teaching method, or they attended a conventional class. Afterwards the performance on different tests of students in both conditions was compared. No significant differences in teaching effect were found. Students did enjoy the classroom experiment more than the conventional teaching method.

One test used to compare the teaching effect of the two classes was a beauty contest experiment. The choices of the secondary school students were compared to the choices of adults during other beauty contest experiments. The results showed that adults did not perform better than secondary school students during a beauty contest experiment. Instead of age and education, knowledge about game theory helped participants in the beauty contest experiment to choose lower numbers.

Table of Contents

Introduction

Behavioural economics has become an accepted part of economics (Feltovich, 2011) and a few economics textbooks already treat the subject (Bergstrom *et. al.*, 2000; Burkett, 2006; O'Sullivan *et. al.*, 2003). Before behavioural economics became accepted, students came into contact with this school of economics for the first time in high school or university. The reason for this was that behavioural economics were not part of the curriculum for secondary school economics students. Since the influence of behavioural economics has increased, students of Dutch secondary school economics are now required to know about behavioural economics as well. This way, they learn that a school of thought exists which is different from standard neoclassical economics and the students learn about the way of thinking in behavioural economics.

In behavioural economics, the behaviour of consumers is the central focus. The assumption of standard neoclassical economics that consumers always act rational and utility maximizing is generally rejected. Alternative explanations of human behaviour are developed and tested. As a method to test alternative hypotheses about human behaviour, researchers can use experiments.

Using experiments to test alternative explanations of consumer behaviour is an important part of behavioural economics. Therefore, secondary school students who learn about behavioural economics, should also learn about experiments. Behavioural economics has only been part of the Dutch curriculum of secondary school economics students for a relatively short time. As a consequence, course materials discussing behavioural economics do not all offer information about experiments or about experiments which teachers can do as a class exercise.

Although this limitation exists, the Dutch government recognizes the value of teaching secondary school economics students about behavioural economics and experiments used in behavioural economics. Therefore, the curriculum for all Dutch secondary school economics students will contain a larger amount of experiments in class (Teulings, 2005). This change in curriculum created a need for experiments that can be performed in class to teach the students about economics in a different way, while also teaching the students about conducting experiments.

The research reported here had three objectives. The first objective was to develop a classroom experiment and test it in secondary school economics classes. Furthermore, in the literature, not much is known about the effect of using an experiment as a teaching method. Some research has been done using experiments as a teaching method on undergraduate students, but to the knowledge of the authors, the effects of this teaching method have not been investigated with secondary school students. The second objective of this research was to examine the teaching effects of using an experiment as a teaching method with secondary school economics students.

The teaching effect of using an experiment as a teaching method was measured, among others, by measuring the performance of secondary school students during a beauty contest experiment. More about the beauty contest experiment and why it was used will be explained later in this paper. The third objective of this research was to test the generalizability of the outcome of a beauty contest experiment with secondary school students as participants. A lot of research has been done into the beauty contest experiment (Bosch-Domènech *et. al.*, 2010). However, the decisions made by secondary school students

during a beauty contest experiment have not been compared to those of adults, to the knowledge of the authors.

The next section treats classroom experiments. Next the research questions are presented. Then the experiment used in this research is explained and hypotheses are stated. A description of the setup of the experiment is followed by the results of testing the hypotheses. Finally, this paper is concluded with a discussion.

Classroom experiments as a teaching method

Common knowledge classroom experiment

The classroom experiment that was developed in this research teaches students about the use of common knowledge. Common knowledge is a subject that is not easily explained using only words. Therefore it was chosen as the subject for the classroom experiment. It was expected that students learned about common knowledge more easily when they experienced it first-hand during a classroom experiment.

Information is common knowledge if each person in a group has the information, and each person knows that each other person has the information, and each person knows that each other person knows that each other person has the information, and so on (Geanakoplos, 1994). An example of common knowledge is an announcement made during a mandatory company meeting that is attended by all the company's personnel. The day after the company meeting, all employees can start a conversation about the announcement with any other employee, because they know that all the other employees heard the announcement and that the other employees all know that all other employees heard the announcement.

Not only information can be common knowledge. Public actions can also be common knowledge. If an action is common knowledge, every person can observe the action and knows that all others can also observe the action. Furthermore, every person is aware that the others can see the action and know that the others are aware of this, and so on.

As an example, take five gnomes sitting in a circle, facing their king, who stands in the middle. Each gnome either wears a blue or a red pointed cap and they can see the colour of the other's caps. They cannot see the colour of their own cap and will not tell the others the colour of their caps. The king tells the five gnomes that he will ring a bell and at that moment all gnomes who know for sure they are wearing a red cap must stand up. Now suppose two of the five gnomes are wearing a red cap, three a blue cap. No matter how many times the king rings his bell, no gnome will stand up, since it is not possible for the gnomes to know the colour of their own cap, without looking at it or asking another gnome. Readers should convince themselves that it is not possible for the gnomes to know the colour of their own cap.

Now suppose the king tells the gnomes that there is *at least* one gnome wearing a red cap. The king then rings his bell for the first time, and no gnome will stand up. When the king rings the bell for the second time, both gnomes wearing a red cap stand up. This result is different from the previous situation where the king did not tell the gnomes that at least one of them wore a red cap. However, since there were two gnomes with red caps, every gnome already had that information. The difference is that in the second situation, the information that there was at least one gnome wearing a red cap became common knowledge. This made both gnomes wearing a red cap realize that the other gnome wearing a red cap saw a red cap, or they would have stood up when the bell rang the first time. This prompted both gnomes to stand up when the bell rang for the second time. Because the actions of the gnomes, remaining seated or standing up, were also common knowledge, the gnomes wearing red caps could figure out the colour of their caps.

In the example, the gnomes assumed that the others all made the right choice given the information they had. To be sure that the others all made the right choice, another kind of common knowledge is required: common knowledge of rationality. This kind of common

knowledge influences the outcomes of games and real life situations. The players may know most of the information the other player has due to common knowledge of information and common knowledge of actions, but the players do not know for sure what decisions the other player would make based on that knowledge. This is because there is usually no common knowledge of rationality.

The situation described in the example is an interesting game theoretic situation. This is because the gnomes depend on the actions of the other gnomes in order to find out the colour of their own cap. Generally in game theory, a player is better off when his or her opponents act less rational compared to him or herself. Then it is easier for the player to win the game. However, in the example of the gnomes, it is more to the advantage of each player if the others are at least as rational. It then becomes easier to trust the information obtained from the actions of the other players.

Research on classroom experiments

Interest in the use of experiments as a teaching method has increased recently. Several textbooks for economics courses come with supplements for classroom games/experiments (Dickie, 2006). With the rising interest in classroom experiments comes the question whether or not these classroom experiments actually improve learning. Do (undergraduate) students who receive classes using experiments as a teaching method perform better in tests, compared to students who attend conventional classes? Conventional classes are classes in the form of a lecture. A teacher explains the subject of the class by using a (digital) blackboard and by talking about the subject. The teacher could also use (limited) discussion with the students.

Several studies (Cardell *et. al.*, 1996; Emerson *et. al.*, 2004; Dickie, 2006) compared the teaching effect of courses using classroom experiments to the same courses using conventional teaching methods instead. All three studies used the Test of Understanding College Economics (TUCE) before and after the courses to see whether there was a significant difference in learning across courses with classroom experiments and courses using conventional teaching methods. Emerson *et. al.* (2004) and Dickie (2006) found that the courses using classroom experiments yielded better results. Cardell *et. al.* (1996) did not find a significant difference between courses with experiments and courses with conventional classes. Another study (Yandell, 1999) also compared a course using six experiments as a teaching method to the same course using only two classroom experiments. Course grades were used to compare the effectiveness of the two courses. No significant difference was found between the two courses.

Other research tested the effectiveness of a single class with an experiment against a single class with the same subject but without the experiment (Frank, 1997; Gremmen *et. al.*, 1997). The research concluded that classroom experiments led to better performance in a test compared to the conventional teaching method. Both studies measured the difference in teaching effect using multiple choice questions about the subject of the lesson.

In reaction to the interest in classroom experiments, Brauer and Delemeester (2001) created an overview of classroom games published in the literature. In their paper, they also reported the teaching effects of the few classroom games that were actually tested. They concluded that the teaching effects of classroom experiments as measured by test scores were at least as large as the teaching effects of conventional classes. In addition, students who participated in the classroom experiments mostly enjoyed those classes more than conventional classes. Evidence for the opinions of students about classroom experiments

was mainly anecdotal. Only Gremmen *et. al.* (1997) tested whether the students themselves experienced the classroom experiment as more positive than conventional classes. They found that students who participated in the classroom experiments were more positive about their class compared to students who attended conventional classes. The opinion of the students about the class is also important for the effectiveness of a classroom experiment. If students like participating in an experiment more than attending conventional classes, this is another reason to prefer experiments over conventional classes (Gremmen *et. al.*, 1997).

These results suggest that using classroom experiments when teaching economics to secondary school students is preferable to using conventional teaching methods. However, the research on classroom experiments used undergraduate students as participants throughout. Since secondary school students are younger compared to undergraduate students and are used to a different kind of teaching, the teaching effects could be different. Also, the opinion of secondary school students about classroom experiments is unknown. For these reasons, this paper aims at testing the effects of a classroom experiment on the performance of secondary school students concerning their economic knowledge and skills.

In this research, the teaching effect of an experiment was compared to the teaching effect of a conventional class with a related subject. In previous research on a single class using a classroom experiment, performance was better after the class with the classroom experiment compared to the conventional class (Frank, 1997; Gremmen *et. al.*, 1997). Therefore, it was expected that performance would also be better for secondary school students who participate in an experiment, compared to secondary school students who participate in a conventional class. Undergraduate students also enjoyed lessons with classroom experiments better compared to students in conventional lessons. It was therefore expected that secondary school students would enjoy a classroom experiment more than a conventional class as well.

Research Questions

The three objectives of this research were to develop a classroom experiment and test it, to examine the teaching effects of using an experiment as a teaching method for secondary school economics students and to test the generalizability of the result of a beauty contest experiment with secondary school students as participants.

The first objective, developing a classroom experiment, was of a practical nature. A subject was chosen for the classroom experiment, common knowledge, and an experiment was developed.

The second and third objectives required more research. In a classroom experiment, all students actively participate. They are required to make decisions and within the context of the experiment they also experience the consequences of those decisions. Students participating in a classroom experiment are generally more involved compared to students attending a conventional class (Gremmen *et. al.*, 1997). Therefore, it was expected that secondary school students pay more attention to a class with an experiment compared to a conventional class. The students were expected to learn more from a classroom experiment than from a conventional class. Whether or not this was true was tested by answering the following research question.

1. *"What are the differences in performance of secondary school economics students who participated in a classroom experiment compared to secondary school economics students who attended a conventional class?"*

By answering this research question, the authors wanted to find out if using an experiment in class was preferable to teaching a class using a conventional teaching method. The goal of teaching a class is to make the students understand the subject matter and enable them to apply it. In order for an experiment to be preferable to a conventional class, the students should be better able to apply what they have learned, compared to students who attended a conventional class on the subject. This meant that the difference in performance of the students after a class with or without an experiment should be measured. Therefore the following two sub research questions were asked.

1a. "How well are secondary school students able to use the subject matter presented to them in a classroom experiment?"

1b. "How well are secondary school students able to use the subject matter presented to them in a conventional class?"

Using an experiment as a teaching method would also be preferable to a conventional class if students enjoyed the classroom experiment more than the conventional class. In addition to performance, enjoyment of the students was also measured to answer the following two sub research questions.

2a.. "How much do secondary school economics students enjoy a classroom experiment?"

2b.. "How much do secondary school economics students enjoy a conventional class?"

In experimental economics literature, many experiments that use undergraduate students or other adults as participants can be found (Feltovich, 2011). No studies used secondary school students as participants. Because of this, it was unclear whether the results of studies using adults as participants could be generalized to secondary school students and vice versa. By comparing the decisions of secondary school students and adults during the same experiment, the generalizability of the findings of this research was tested. The next main research question was:

3. *"What are the differences in decision making between secondary school students and adults in the same experiment?"*

For the decisions of secondary school students to be compared to the decisions of adults participating in the same experiment, an experiment needed to be used that had been used frequently in other research before. The beauty contest experiment was chosen because it has been performed in the literature with different types of participants.

Explanation of the experiment

Experimental setting

This research compared the teaching effect of a classroom experiment to the teaching effect of a conventional class. Students who participated in the classroom experiment were in the experimental condition. Students who attended the conventional class were in the control condition. Students in both conditions attended a second class to test their acquired knowledge and skill in a related performance task.

The experiment was conducted in three economics classes of a secondary school in the Netherlands. The subject matter taught during the experiment and the conventional class was of a difficulty level appropriate for students in the final years of secondary school, preparing for university education. All students received two classes of 50 minutes each. Two classes were used, because the students received their lesson with or without an experiment during the first of those classes and their subsequent performance was tested in the second class.

All classes were taught by the same experimenter, to minimize the influence of experimenters on the results. The regular school teachers were present during the classes to make sure that the students behaved themselves and the experimenter remained in control of the groups. The classes were given according to the regular time schedule, to minimize inconvenience to the students and to have the classes take place in the regular classrooms.

In the last class prior to the experiment, the teachers informed the students that the next two classes would be used for scientific research. The subject matter would not be part of the students' curriculum. Their teachers asked them for their cooperation and the students were told that a prize could be won during the second class, to motivate the students to pay attention during the two classes.

The odd number of groups participating in this research meant that a choice had to be made between the number of groups in the experimental condition and the control condition. Because the experimental condition tested the classroom experiment that was developed for this research, two groups (with respectively 13 and 26 students) were included in the experimental condition. The remaining group (25 students) was included in the control condition.

Experimental condition

In the experimental condition, secondary school students participated in a classroom experiment about common knowledge. The experiment lasted for an entire class hour and consisted of three different rounds.

The experiment was placed in a narrative setting, to better clarify the procedure of the experiment to the students and to make the experiment less abstract. The setting is illustrated here in short. The full story of the setting as used in the classroom experiment can be found in appendix A.

Two divers, the role of the students in the experiment, are on board of a boat outfitted with blue and green diving bells. The boat floats above a recently discovered shipwreck containing items of great value. The two divers are rivals, and the payoff is highest for the diver who salvages the most valuable items.

Because of the dark, the divers are unable to see the colour of the two diving bells when they board them. The blue diving bells all turn out to have a manufacturing error

which causes the bell to rupture if the water pressure becomes to high. Divers in blue bells therefore need to stop descending and start ascending. Divers in green diving bells are in no danger. The problem, is that the divers can only see the colour of the other's diving bell due to lights on top of each diving bell.

At this moment, the students started to play their roles. They were separated into pairs. All students received a card, which they themselves could look at, but they were not allowed to share the information on their card with their classmates. On the card the students could see the colour of the diving bell of their classmate. The students did not know the colour of their own diving bell. The students were then simultaneously given the opportunity to stop descending and start ascending. Students who made this choice had to stand up, so the other students could see which students chose to ascend. Students who wanted to keep descending simply needed to remain seated. Once every student had made a choice, a final possibility was offered to start ascending to those students who were still seated. Ascending students could not reverse their choice.

After two opportunities to make a choice, the first round was over. Students were asked to reveal to each other the colour of their diving bells. Every student then knew whether or not he or she had made the right decision. The experimenter asked a few students about their choice and their motivation for that choice. The aim of these questions was to alert the students to the fact that there was not enough information to make the right choice with certainty. During the first round of the experiment, students could only guess the colour of their own diving bell.

At the start of the second round, the students were divided into different pairs. The second round was largely the same as the first round, with one exception. When the captain warned the students that all blue diving bells were damaged, he also told both divers that at least one of them was inside a blue diving bell. Both divers heard the same announcement and were aware that the other diver heard the captain as well. The students were therefore aware that the other students also knew that at least one diving bell was blue. It was common knowledge that at least one diving bell was blue. Compared to the first round, in the second round students had more information.

The rest of the procedure for the second round was the same as the first round. When all students had made their choice, the examiner asked a few of the students about the reasoning behind their decision. By asking students who made the right decision about their reasoning, students who made the wrong choice heard how they should have reasoned to make the right decision. When a few students told the rest about their reasoning and at least one student gave a sound reasoning for either remaining seated or standing up, the examiner explained to all students how they could be able to find out the colour of their own diving bell. A correct line of reasoning is presented appendix B.

The third and final round of the experiment put the students in a more challenging situation. In this round the students were arranged into groups of three. The situation was the same as in the second round, with the difference that this time three divers were descending. Again, students could either be in a blue diving bell or in a green diving bell and the blue diving bells were broken. As in the second round, at least one of the descending diving bells was blue and this fact was common knowledge. The students knew the colour of the other two diving bells.

Again, the students had to find out the colour of their own diving bell and choose whether they wanted to keep descending or start ascending. To this end all students received a card indicating the colours of the diving bells of the other two students. Unlike in

the previous rounds, the cards contained information about two diving bells and the order in which the cards were handed out mattered. The cards were therefore handed out in such a manner that each student received information corresponding with the cards of the other two students.

In the third round, students had three opportunities to choose whether they wanted to start ascending. The students made their first choice simultaneously. After the first opportunity, the students could observe the choice of the other two students. The students who had remained seated then got the opportunity to change their mind and stand up. Finally, students who had remained seated during the first two rounds got one final chance to observe the choices of the other two students and decide if they wanted to stand up the third time. During all opportunities to stand up, students who had stood up during previous rounds were not allowed to change their mind.

After the third opportunity to make a choice, the final round of the classroom experiment was finished. All students were asked to show the other two students the colour of the diving bells. Students could then see whether they had made the right decision.

Excluding groups of three green diving bells left three possible combinations. Students could be in a group with two green diving bell and a blue diving bell, they could be in a group with one green diving bell and two blue ones or their group could consist of three blue diving bells. During the classroom experiment, groups containing all three combinations were present. The experimenter asked members of three groups with different combinations of diving bell colours about their choices and the colour of the diving bell they were in. This way, students heard about the combinations of colours in the other groups as well as the reasoning others used to make a decision.

After the experimenter questioned some students about the decision they made, a line of reasoning that the students could have used to find out the colour of their diving bell was explained. The explanation can be found in appendix C.

At the end of the classroom experiment, the experimenter announced that during the next class, another experiment would be held in which all students were supposed to participate. The examiner also told the students that a prize of 10 Euros could be earned by the student making the best decision and that the experiment they just participated in would help the students with the decision.

Control condition

Students in the control condition did not participate in a classroom experiment. Instead, they attended a conventional class on a subject related to the subject of the classroom experiment. In the conventional class, the students were taught about a subject by the experimenter and were asked to make exercises about the subject afterwards. No experiment was done in class.

The subject of the conventional class was different from the subject of the classroom experiment. The reason for this was the difficulty to clearly explain common knowledge to secondary school students without the use of an experiment. The concept of common knowledge could be explained to the secondary school students. However, the implications of common knowledge for decision making would could not be explained as well. Examples could be used, but secondary school students could understand common knowledge better after experiencing the implications themselves. Therefore, the subject of the lesson in the control condition was the prisoner's dilemma.

The prisoner's dilemma was chosen as the subject of the conventional class because the situation of the prisoner resembled the situation of the divers. A choice had to be made between two conflicting possibilities. Someone else had to make the same choice, and the outcome of the choice of one student affected the outcome for the other student. In both situations, students had to take into account the choices the other students made.

The difference between the subjects of the experimental class and the control class could influence the performance of the students during the test in the second class. Care was taken to make sure that what the students learned from both classes was comparable. Despite this effort, it is possible the differences between the subject of the two classes were significant.

A full script of the class (in Dutch), including exercises and slides can be received from the authors at request. The conventional class was taught using slides projected by a beamer. After a short introduction by the experimenter, the class started with the question whether any of the students was familiar with the prisoner's dilemma. This question was asked because the students had been taught about game theory in the past. The students were somewhat familiar with the prisoner's dilemma. After recollection from the students and some explanation from the experimenter, the students were taught to write the prisoner's dilemma in the form of a matrix.

Next, the students watched a scene from the movie 'A Beautiful Mind'. Afterwards, the experimenter constructed a matrix for a simplified version of the scene. Two male friends in a bar had the choice between hitting on a blond girl or on one of two dark-haired girls. Both boys liked the blond girl the most, but if they both hit on her, neither could go out with her. So the boys agree to both go talk to a dark-haired girl. That way their chances on a date improve dramatically. If one of the boys were to break the agreement, he could be the only one approaching the blond girl and might end up dating the most preferable girl. So both boys must choose whether they want to stick to the agreement or not. Together with the students the experimenter filled in the consequences of the possible choices for the boys, using grades. The resulting matrix resembled a prisoner's dilemma, but the experimenter explained why the matrix was not actually a prisoner's dilemma. The matrix was not actually a prisoner's dilemma, because there was no dominated strategy.

When the matrix was completed, the topic of dominated strategies was discussed. It was explained to the students that in the matrices, sometimes strategies can be found that are always inferior to the other and should therefore never be chosen. This was then shown for the two matrices constructed in the lesson.

When the students had been instructed about the way to find dominated strategies, they received two assignments about prisoner's dilemmas and dominated strategies. They were told to complete the assignments right away within a time limit of ten minutes. When the time limit was passed, the experimenter explained the right answers of the students.

After the conventional class was over, the experimenter told the students that the next class would be part of the same research. The experimenter explained that the next class would have an experiment in which all students could participate and the student who made the best decision would earn 10 Euros.

Performance test

During the second class the ability of the secondary school economics students to apply what they had learned in the first class was tested by means of a beauty contest

experiment and a questionnaire. The class was the same for students in the experimental and the control condition.

During the beauty contest experiment, all participants chose one number between 0 and 100. Once all participants made their choice, the average of all chosen numbers was multiplied by a factor p ($0 < p < 1$). In this research, p equalled 2/3. The participant who chose the number closest to 2/3 of the average of all chosen numbers was the winner and won 10 Euros. The prize did not depend on the value of the winning number. In case of a tie, the prize was shared among the winners.

When choosing a number during the beauty contest experiment, it was important to take into account the possible choices made by the other students. In the experimental condition and the control condition, students were taught to take into account the possible choice of others when making their own choice. The beauty contest experiment was specifically chosen to compare students in the experimental and the control conditions, because choosing the winning number during the beauty contest experiment requires students to correctly predicting the reasoning used by the other students. The students would only find a solution for the problem presented to them during the beauty contest experiment if they were able to use an adequate level of reasoning.

In the beauty contest experiment, participants who expect most of the choices of the others end up choosing the lowest numbers. If the teaching effect of a classroom experiment is larger than the teaching effect of a conventional class, the numbers chosen during a beauty contest experiment by students in the experimental condition should, on average, be lower than the numbers chosen by students in the control condition. This is because students who are better able to use the reasoning of others in their own decision, chose lower numbers. The first hypothesis expresses this presumption.

Hypothesis 1: *Students in the experimental condition will on average choose a lower number in the beauty contest experiment compared to students in the control condition.*

The reasoning used by participants during the beauty contest experiment can be described using different models (Bosch-Domènech *et. al.*, 2002). The four models discussed here include the fixed-point argument, iterated elimination of weakly dominated strategies, iterated best reply, and iterated best reply to the non-degenerate beliefs.

The first reasoning model is based on the fact that the lowest number of the interval is the unique (Nash) equilibrium. All participants want to choose a number lower than the numbers chosen by the others and assume that each of them has the same objective. When this assumption is true, no matter how low the chosen number, the other participants can still choose even lower numbers, unless one chooses the lowest number possible. If all participants follow this line of reasoning, all participants should choose the lowest possible number. Any participant deviating from this deviates from the winning number. Therefore, all participants will choose the lowest number, i.e. 0. This reasoning model is called the fixed point argument (Bosch-Domènech *et. al.*, 2002).

The second reasoning model is based on avoiding dominated strategies. Since the winning number is equal to 2/3 of the average of the chosen numbers, even if all participants choose the highest possible number (100), the winning number cannot be higher than 67. This means that 67 weakly dominates all higher numbers. However, if participants realize this and expect the others to realize this as well, 44 (which roughly equals 2/3 of 67) weakly dominates all higher numbers. If participants expect the others to

realize the second step as well, even more numbers become weakly dominated. Iteration of this strategy will lead to choosing lower numbers with every iteration, up to choosing 0. This reasoning model is called iterated elimination of weakly dominated strategies (Bosch-Domènech *et. al.*, 2002).

The first two reasoning models are theoretical and based on game theory. The next two models are based on experimental results based on actual human behaviour.

The third reasoning model predicts the numbers chosen by participants based on iterations of the best reply to the choices of the other participants. A participant who does not take into account the numbers chosen by the others uses level 0 reasoning. That participant chooses a number randomly. The expected average of the numbers chosen by participants using level 0 reasoning equals 50. A participant who does consider the choices of the others, but thinks that all the others use level 0 reasoning, is using level 1 reasoning. Participants using level 1 reasoning will therefore chose 2/3 of 50 which roughly equals 33. Participants using level 2 reasoning expect the other players to use level 1 reasoning and choose 22 (2/3 of 33). In general, participants use level *n* reasoning if they expect the others to be using level $n - 1$ reasoning. The number chosen then equals $50p^n$ (Nagel, 1995). Notice that the main difference between the iterated best reply model and the iterated elimination of weakly dominated strategies lies in the different starting point (50 vs. 100).

The last reasoning model resembles the iterated best reply model. However, the assumption that participants expect all other participants to use the same level of reasoning is questioned. According to the iterated best reply to the non-degenerate beliefs model each participant has a belief about the level of reasoning of the others, where the others can use different levels of reasoning. For example, a participant can expect half of the others to randomly choose a number (level 0 reasoning) and the other half to use level 1 reasoning. The participant him- or herself then chooses his or her number using level 1.5 reasoning.

Experimental results from beauty contest experiments suggest that the first two reasoning models do not adequately describe behaviour of participants during the beauty contest experiment (Bosch-Domènech *et. al.*, 2002). The remaining two models give better predictions of the numbers chosen. The iterated best reply to the non-degenerate beliefs model has been compared to the iterated best reply model, to test whether participants expect all others to use the same level of reasoning or if they expect the others to use different levels of reasoning (Stahl, 1998). No significant difference was found between the predictions of the two models. In this research, the iterated best reply model was used to compare levels of reasoning used by the secondary school students, because the model is easier to work with.

If the teaching effect of the experimental condition is larger compared to the teaching effect of the control condition, students in the experimental condition are better at taking into account the choices of others and expect the others to consider their possibilities more compared to students in the control condition. This means that the level of reasoning used by students in the experimental condition would be higher compared to the level of reasoning used by students in the control condition. This was the second hypothesis.

Hypothesis 2: *The level of reasoning used during the beauty contest experiment will be higher for students in the experimental condition compared to students in the control condition.*

The procedure used for the beauty contest experiment in the performance class was as follows. The students were asked to write down any number between 0 and 100 and their name on a small piece of paper. The names were necessary to determine the student who chose the number closest to the 'most beautiful number'. The students were not allowed to communicate, they had to choose their number on their own. Once every student had written down a number, the pieces of paper were collected by the experimenter. After all numbers were collected, two-thirds of the average of all numbers that were written down was calculated. The number found this way was the 'most beautiful number'. The student closest to this number was the winner of the 'beauty contest game' and received 10 Euros. After the explanation of the experiment, the experimenter summarized the procedure and explicitly told the students they had to try to write down the number they thought would be equal to two-thirds of the average of all numbers written down. The experimenter asked the students whether they understood the experiment. Because no communication was allowed, this was the last moment for the students to ask questions.

When all chosen numbers had been handed in, the winner of the experiment was not immediately announced. First, a questionnaire was handed out. The questionnaire can be found in appendix D. The questionnaire was used to test the performance of the students on a number of game theoretic problems. The questionnaire also contained questions that measured the opinion of the students about the treatment classes and questions that collected demographic data about the students.

The questionnaire first asked the students what number they filled in during the beauty contest experiment. They were also asked what reasoning they used to choose their number and what reasoning they thought the other students had used. The students were asked for their reasoning and their expectations of the reasoning of their classmates to identify their level of reasoning.

The next question asked the students whether they expected a group of thirty economics teachers to have a different result in the beauty contest experiment, compared to the students themselves. If a student indicated that he or she expected a difference, he or she was asked to write down why there was a difference. This question was used to check whether students understood that the number they chose depended on the choices of the other students and thereby the knowledge of the other students.

Next, a question about a water lily patch was asked (Frederick, 2005). The question measured whether students were able to use a higher level of reasoning. The question was about part of a lake that was covered with water lilies. Each day the area covered doubled. It took 48 days for the water lilies to cover the whole lake. The question for the students was how long it took the lilies to cover half the lake. The right answer to the question was 47 days.

An answer that intuitively seems right is 24 days. Students who do not take the time to consider the problem will probably give this answer, whereas students who consider the problem more carefully should realize to think back one step in time to find the correct answer. If students who participated in the classroom experiment during the first class were learned more to take all available information into account, compared to students who attended the conventional class, students in the experimental condition would perform better on this task. That was the third hypothesis:

Hypothesis 3: *A larger part of the students in the experimental condition will solve the 'water lily problem' correctly compared to students in the control condition.*

The next part of the questionnaire was an economics assignment that tested the ability of the students to apply what they had learned. The students were given a matrix which displayed possible revenues of two competing companies that had to choose between advertising and not advertising (see appendix D). The students were asked three questions about the possibilities of the companies. The first question was whether company A had a strategy that was always better than the other and if so, what that strategy was. The second question was the same, only this time it was about company B. The third question asked what the student would choose if he or she had to make the decision for company A and why. This assignment resembled two tasks given to students in the first lesson of the control condition. Since those students had more experience with similar tasks, it was expected that they would score better on the assignment. That was the fourth hypothesis of this research.

Hypothesis 4: *On average, students in the control condition will score better on the business dilemma task, compared to students in the experimental condition.*

The next part of the questionnaire measured the opinions of the students about the treatment classes. Students were presented with five statements, which can be found in appendix D. Students were asked to indicate how much they agreed with the five statements on a 5-point Likert scale. The statements measured whether students in the experimental condition enjoyed the first class more, compared to students in the control condition. The statements also measured the opinion of the students on the interestingness, difficulty and usefulness of the treatment class. Students who participated in the classroom experiment got to experience the subject matter firsthand. Since the classroom experiment was different from the way students were usually taught, it could also have been more fun compared to their regular classes. Therefore, it was expected that students in the experimental condition agreed more strongly with the statements that the treatment class was fun, interesting and useful. Because students in the experimental condition participated in the classroom experiment and had to make decisions, the treatment class was expected to be more difficult for the students, compared to students in the control condition. These expectations result in the following hypothesis.

Hypothesis 5: *Students in the experimental condition will agree more strongly with the statements that the treatment class was fun, interesting, difficult and useful, compared to students in the control condition reporting about the regular class.*

The final questions asked the students their sex and their grades for both economics and mathematics. These questions were asked because it was expected that the grades for those subjects could influence the rest of the answers in the questionnaire.

By the time the students had all filled out the questionnaire, the experimenter had calculated two-thirds of the average number chosen during the beauty contest experiment at the first part of the performance class. After the experimenter explained why it was difficult to choose the right number in the experiment and showed this using a few slides, the winner of the experiment was announced and the winning student received the 10 Euros.

Results

The results are split into four parts. The first part offers a summary of the sample. In the second part, a report is given of the results of the test of the common knowledge experiment with secondary school students as participants. The third part analyzes the differences and similarities between students in the experimental condition and students in the control condition. The last part compares the decisions of secondary school students made during the beauty contest experiment to the decisions of adults made during similar experiments in the literature.

Sample data

In total, 62 secondary school economics students participated in the beauty contest experiment and filled in the questionnaire afterwards. Of all participants, 39 were boys and 23 were girls. In the experimental condition, 10 of the 37 students were girls, the remaining 27 students were boys. In the control condition, 13 of the 25 students were girls and 12 were boys. The difference in the number of girls in the two conditions was significant ($p < .05$) and could influence the dependent variables. Hence, sex was considered as a covariate during the analysis. The questionnaire also asked the students their mathematics and economics grades. The students were first asked whether they studied mathematics A or mathematics B[1]. In the experimental condition, 20 students studied mathematics A and the remaining 17 students studied mathematics B. In the control condition 16 students studied mathematics A and 9 mathematics B. The difference across conditions was significant ($p < .10$) and could influence the results of the experiment. Therefore, type of mathematics attended by the students was also considered as a covariate during the analysis.

In the experimental condition, students who studied mathematics A had an average grade of 6.82 (out of 10). Students in the experimental condition who study mathematics B had an average grade of 6.26. In the control condition, mathematics A students had an average grade of 6.61 and students who studied mathematics B had an average grade of 6.59. The average grades for mathematics were not significantly different across the different types and across the different conditions. The grades were therefore not expected to influence the results of the experiment.

For the subject economics, the average grade of students in the experimental condition was 6.23. The average grade of students in the control condition for economics was 6.34. The economics grades were not significantly different across the different conditions and were therefore also not expected to influence the results.

In this research, the variables sex and type of mathematics studied were controlled for using two different methods. One method used was analysis of covariance (ANCOVA). When using ANCOVA, values of the dependent variable are compared between the two different conditions, after the relation between the covariates and the dependent variable has been controlled for. An advantage of using ANCOVA is that no students had to be excluded from the analysis. A disadvantage of ANCOVA is that it could only be used to analyse variables that were of at least the interval scale. In this research, that meant that not all variables could be analysed using ANCOVA.

[1] In the Netherlands, secondary school students can choose different combinations of subjects. Depending on the combination, the students attend different types of mathematics. Mathematics A treats less complex mathematics compared to mathematics B and mathematics A is aimed more at application compared to mathematics B.

The second method used to control for the covariates was using matched samples. For every student in the control condition (which had the lowest number of participants), a student in the experimental condition was selected that had the same sex and studied the same type of mathematics. Using this method, the possible influence of sex and type of mathematics studied on the other variables was prevented. Even if sex and/or type of mathematics studied influenced any other variables, the influence was the same in both conditions. The advantage of this method of controlling for sex and type of mathematics studied was that all types of analyses that were possible before matching samples were still useable after matching samples. Using this method, variables that could not be analysed using ANCOVA could still be analysed. A disadvantage of matching samples was that not all students who participated in the experiment could be used in the analysis, since not all students could be matched. The procedure used to match students from different conditions and the resulting matched samples can be found in appendix E. Only 50 of the 62 students could be included in the analysis utilizing matched samples, because 12 students from the experimental condition remained who could not be matched.

In the next section, the experimental condition is compared to the control condition. Where possible, ANCOVA was used in preference to matched samples, because that way the analysis could be done using more participants and the power of the results would be greater.

Test of the common knowledge experiment

One of the objectives of this research was to develop a classroom experiment and test it in secondary school economics classes. The common knowledge experiment that was developed was used in the treatment class attended by students in the experimental condition. In this section, a report is given of the results of using the common knowledge experiment in secondary school economics classes.

The classroom experiment started off by explain the narrative setting to the students. The narrative setting was used to increase the involvement of the students and to make the experiment more interesting for the students. The setting was also to used to make the experiment easier to understand for the students. After the explanation of the narrative setting, the students seemed to understand their role in the experiment and they were paying attention to hear about their part in the experiment. Throughout the whole experiment, students participated without losing interest. Even though the different rounds of the experiment were quite similar.

The observed choices of the students in first part of the experiment were either to guess and hope continuing descending was the best choice, or to choose the safest option and begin ascending. After the first round, the students understood that they did not have enough information to determine the best option.

In the second round, the students were provided with additional information in the form of common knowledge. Most students, especially those in pairs of two blue diving bells, still used either the guessing strategy or the strategy of choosing the safest option. However, in the second round it was possible to use reasoning to determine the right decision. Some of the students used this new strategy. They were mostly students who saw a green diving bell and knew they had to be in a blue one themselves. A few students who saw a blue diving bell also used the strategy. At the end of the second round, the reasoning strategy was explained to all students. No student indicated not to understand it, which presumably meant that most of them really did understand it.

The third round put the students in the most difficult situation and the results reflected the difficulty. Students in groups with one blue diving bell were mostly able to use reasoning to find their best choice. Students in groups with two blue diving bells had more trouble to determine the colour of their own diving bell, and students in groups with only blue diving bells all resorted to guessing or choosing safely. After an explanation of how reason could be used to determine the colour of the diving bell students were in, some students struggled to understand the reasoning. With the increase of the number of diving bells, the difficulty level increased rapidly.

Comparing the third round to the first round showed that the secondary school economics students began to use a strategy utilizing reasoning more as the experiment progressed. They learned how to make use of common knowledge to determine the optimal choice.

The experiment seemed to be challenging enough for the students. When they were in one of the easier groups, they understood how to find out their best choice, but students in the most difficult groups had trouble determining the right choice.

For the experiment, chairs and tables had to be moved around the classroom. This created some unrest, but time was reserved for this. The experiment still fitted within one hour. Students had to be specifically told not to look at their cards when they received them and not to share their information with the other students. This method worked, ensuring that the students 'played fair'. After revealing their cards to the others at the end of each round, the students spontaneously started discussing their results. They showed interest in the choices of others and were interested in whether or not they made the right choice.

Overall, the experiment was suitable to use in the secondary school economics classes. The students paid attention, were involved, were challenged and showed progress.

Performance across conditions

All questions of the questionnaire used to measure performance of the secondary school students and the answers of the students are discussed here.

Beauty contest experiment

It was stated in the first hypothesis that students in the experimental condition would choose, on average, lower numbers in the beauty contest experiment compared to students in the control condition.

To test the first hypothesis, the average numbers chosen in the two conditions were compared, using ANCOVA with sex and type of mathematics studied as covariates. The average number chosen in the experimental condition was 21.48. The average in the control condition was 20.35. The difference was not significant. Also, the relation of the covariate sex with the number chosen and the relation between the second covariate, type of mathematics studied and the chosen number were not significant.

The first hypothesis was rejected. No significant difference was found between the numbers chosen by students in either condition during the beauty contest experiment. This indicated that there was no difference in teaching effect between a classroom experiment and a conventional class.

Level of reasoning

The second and the third questions of the questionnaire were (translated from Dutch): *"Explain why you chose this number, instead of a different number"* and *"how do you*

expect that the other students reasoned to choose their numbers?" These two questions were asked to gain more insight into the thought process of the students while choosing a number. These open questions were used, together with the number chosen in the experiment to identify the level of reasoning the students used when choosing their number. An experimenter analysed the answers to the two open questions and assessed for each student a corresponding level of reasoning.

Based on their answers to the first three questions, the students were divided into five different categories. Students who used level 0 reasoning just chose a number without taking the choice of the other students into account. Students using level 1 reasoning expected the rest of the students to choose a number at random and they themselves choose two-thirds of what they expected to be the average number chosen by the others. Students divided into the level 2 reasoning category reasoned that the other students would expect that the rest would choose a number at random. The category of students who used level 3 or higher reasoning expected the other students to use at least level 2 reasoning. Students in the fifth category could not be included in the other categories because of the divergence of their answers to the first three questions. The answers contradicted each other and it was not clear which level of reasoning the students had used to decide on a number.

Matched samples were used to analyze the levels of reasoning of the students. Table 1 shows the number of students using each level of reasoning across the conditions.

Table 1: Number of students using each level of reasoning across conditions.

Reasoning level	Experimental Condition	Control Condition
0	5	5
1	10	8
2	6	9
3+	4	2

The second hypothesis stated that students in the experimental condition would on average use a higher level of reasoning during the beauty contest experiment compared to students in the control condition. However, no significant difference was found between level of reasoning of students in the experimental condition and students in the control condition.

The second hypothesis was also rejected. No significant teaching effect was found on the level of reasoning used by secondary school students during the beauty contest experiment.

Expectation of the behaviour of economics teachers

After the students explained their reasoning, they were asked to read the following statement (translated from Dutch). "*Assume that your teacher would participate in this experiment together with 29 other economics teachers."* The students were then asked whether they expected the outcome of the experiment to be different from the outcome of the experiment in which they participated themselves. If a student did not think there would be a difference, he or she did not have to give an argument for his or her answer. If a student did expect a difference, the student was asked to explain what the difference would be.

This question measured whether the students realized that the number they chose during the experiment depended on the choice of the other participants. If a student

expected economics teachers to be better able to chose the right number, he or she should expect the outcome in the experiment with the teachers to be lower. If a student answered yes to this question, the student understood that there was more to the beauty experiment than just choosing a number. If the student also expected the chosen numbers to be lower when the teachers choose them, they understood the experiment best of all.

First, students were divided based on whether they believed that there would be a difference between students and teachers in the experiment. Second, the argument given by the students who expected a difference was used to decide whether they thought the mean number chosen would be lower in the experiment with teachers compared to the experiment with students. Table 2 shows how many students expected a difference and the amount of students among them who also expected a lower number.

Table 2: Number of students who thought there would be no difference between teachers and students in the beauty contest experiment and students who thought there would be a difference. Students who thought there would be a difference were divided by the explanation that the numbers would be lower and other explanations.

	Difference with economics teachers?		
Condition	**No**	**Yes, other argument**	**Yes, lower number**
Experimental	13	4	8
Control	12	8	5
Total	25	12	13

The table shows that in both conditions the number of students who thought there would be no difference was approximately equal to the amount of students who did expect a difference. No significant difference was found.

Growth of the lily patch

To correctly answer the question about the lilies, students had to think back one step. Students in the experimental condition had more applied experience with this way of thinking. It was therefore hypothesized that more students in the control condition would give a wrong answer to this question.

The students gave a few different answers to the open question. Those answers were used to divide students into students with right answers and students with wrong answers. Table 3 shows the number of students in both conditions who gave the right or a wrong answer.

Table 3: Number of students with the right or a wrong answer to the question about the water lilies.

	Experimental condition	Control condition
Right answer	23	19
Wrong answer	2	6

Table 3 shows that in both conditions a large number of students gave the right answer. Judging by the frequencies displayed in the table, there was no clear difference between the two conditions.

Pearson's χ^2-test was performed on the 50 matched students. However, the expected count of the two cells with the number of students who were wrong in table 3 was too low for the test to be reliable. Fisher's Exact Probability Test was therefore used. No significant difference was found across the conditions.

Hypothesis 3 was rejected. Students who attended the conventional class were not less able to solve the water lily problem than students who participated in the classroom experiment.

Strategies for the two companies A and B

The last question that measured what the students learned during the treatment class was about the two companies, A and B. The companies had to choose between advertising and not advertising. Company A did not have a dominant strategy. Company B did have a dominant strategy. It always yielded the biggest profits for company B to advertise.

The question about the two companies was split into three parts. The first part asked whether company A had a choice that always yielded a better result than the other choice and why. The second part asked the same question for company B. The last part asked the students what their own choice would be if they had to make the decision for company A.

The answers to the question about the two companies were analysed by grading the answers of the students. The first two questions were each worth 2 points and the third question was worth 3 points. The procedure used to score the answers for each part of the question is found in appendix F.

The scores of all students for the three parts were added up to create a final score for the question about the two companies. Table 4 shows the distribution of the students across the possible amounts of points.

Table 4: Distribution of the students across the possible amounts of points. Shown for students in each condition and for all students together.

Total points scored	Experimental condition	Control condition	All students
0	0	1	1
1	0	0	0
2	2	0	2
3	2	3	5
4	6	7	13
5	11	3	14
6	5	5	10
7	11	6	17

Students in the control condition learned to solve similar problems in their first lesson, whereas students in the experimental condition did not receive practice in reading matrices and solving similar problems. It was therefore hypothesized that students in the control condition would, on average, score better on the question about the two companies compared to students in the experimental condition.

The average score for the assignment of students in the experimental condition was 5.206 out of 7. The average score of students in the control condition was 5.095. A significant relation between sex and the score was found. The average score of boys was 5.49, while the average score of girls was 4.61. After controlling for the covariate sex, no significant difference was found between the scores across conditions.

The fourth hypothesis was rejected. The score of students who participated in the classroom experiment was not significantly different from the score of students who attended the conventional class.

Opinions about the treatments

For secondary school students, a classroom experiment was an unusual method of teaching. The research was therefore interested in the opinions of the secondary school students about the treatment classes.

To measure the opinion of the students about the classes with and without an experiment, they were presented with five statements. The students were asked to indicate how much they agreed or disagreed with the statements. The statements were different across the two conditions, because they depended on the treatment class the students had attended. The statements presented to students in the experimental condition are used in this section.

For the analysis of the opinions of the students about the five statements, the answer options were each coded with a number. 'Completely disagree' was coded with a 1, 'somewhat disagree' with a 2, and this was continued until 'completely agree', which was coded with a 5. So the higher the number, the more the student agreed with the statement.

Using the scores obtained from the answers of the students, the opinions of the students in the two different conditions could be compared. This was done by comparing the average answers for each statement across the two conditions using independent t-tests.

The opinions of the students about the statements "the previous class with an experiment was fun", "the previous class with an experiment was interesting" and "the previous class with an experiment was useful as a preparation for the experiment" were not significantly different across conditions.

The opinions of the students about the statement "the previous class with an experiment was more fun than other economics lessons" were significantly different across conditions. The average agreement of students in the experimental condition was higher (M = 4.22), compared to students in the control condition (M = 3.76).

The opinions of the students about the statement "the previous class with an experiment was difficult" were also significantly different across conditions. Students in the experimental condition considered the treatment class to be less easy (M = 2.73) compared to students in the control condition (M = 2.16).

Hypothesis 5 was partially rejected. Students who participated in the classroom experiment did not agree more with the statements that the treatment class was fun, interesting or useful. However, students in the experimental condition did think significantly more than students in the control condition that the treatment class was more fun compared to regular economics classes. Compared to regular economics classes, secondary school students liked the classroom experiment more than the conventional class. Students in the experimental condition also agreed significantly stronger with the statement that the treatment class was difficult compared to students in the control condition. Classroom experiments challenge secondary school students more than conventional classes, but classroom experiments are also enjoyed more.

Beauty contest experiment: secondary school students compared to adults

This research was not the first research to study the beauty contest experiment. Other research has studied the choices of participants during a beauty contest experiment (for example: Alba-Fernández *et. al.*, 2006; Bosch-Domènech *et. al.,* 1997, 2002; Selten *et. al.* 1997; Thaler, 1997). However, unlike other research studying beauty contest experiments, this research was the first to perform the beauty contest experiment with participants who

were non-adult students. The participants in other beauty contest studies were predominantly adults. A few beauty contest experiments held in newspapers may have had participants who were younger than 18 years old. Even so, the participants were not all under 18 years old.

Because no results are published of beauty contest experiments with secondary school students as participants, the choices made by the participants in this study were compared to the choices of adults during beauty contest experiments. This comparison was made to test the generalizability of results of beauty contest experiments with adult participants to results of beauty contest experiments with secondary school students as participants. Another reason to compare the choices of secondary school students and adults during a beauty contest experiment was to check whether the beauty contest experiment was well suited to compare the experimental to the control condition in this study. In the case of no significant differences between the choices of secondary school students and those of adults during a beauty contest experiment, the lack of difference would imply that no attributes of the participants influenced the choices made during the beauty contest experiment. If that were true, no differences between the two conditions in this study would influence the numbers chosen during the beauty contest experiment. The fact that no significant differences were found between the choices of the two conditions in this study could then be explained by properties of the beauty contest experiment, instead of explained by differences in the experimental and the control condition.

The numbers chosen by the secondary school students during the beauty contest experiment were compared to numbers chosen by adults, collected from the literature. Other research using the beauty contest experiment with the same parameter (2/3 of the mean) was consulted. Numbers chosen in the first round of the other experiments were used.

The experiments included in the comparison used the following groups of adults as participants: business undergraduate students, undergraduate students from various departments, economics undergraduate students (with and without the possibility to take the choice of the number in the beauty contest home to think about it), theorists and newsgroup participants. An overview of the different subject groups is found in appendix G.

The subject groups were compared using ANOVA, with the subject group the participants belonged to as the independent variable and the chosen numbers as the dependent variable. A significant ($p < .05$) effect of subject group on the numbers chosen was found. Post-hoc tests were used to find contrasts between the different subject groups. The average number chosen by theorists ($M = 16.95$), secondary school students ($M = 20.75$) and newsgroup participants ($M = 21.74$) was significantly lower compared to the average number chosen by participants in the other subject groups. The average number chosen by business undergraduate students ($M = 32.95$) and undergraduate students from various departments ($M = 35.16$) were significantly higher compared to the numbers chosen by participants in the other subject groups. Based on the contrasts that were found, participants in the beauty contest experiments could be divided into three subsets. The average number chosen by subject groups in the same subset were not significantly different. The theorists, secondary school students and newsgroup participants belonged to one subset. The economics undergraduate students, who had limited knowledge in game theory, belonged to their own subset. The third subset contained the business undergraduate students and the undergraduate students from various departments. The subsets are summarized in table 5.

Table 5: Division of the different subject groups into subsets

Subject group	Subset		
	1	2	3
Theorists	X		
Secondary school students	X		
Newsgroup participants	X		
Economics undergraduates		X	
Economics undergraduates (take home)		X	
Business undergraduates			X
Undergraduates from various departments			X

Secondary school students did not belong to the same subset as any of the undergraduate students included in the analysis. The secondary school students choose significantly lower numbers during the beauty contest experiment, compared to the undergraduate students. The difference indicates that neither age, nor life experience or education influenced the numbers chosen during the experiment. The economics undergraduate students, who had limited knowledge of game theory, choose significantly lower numbers compared to all other undergraduate students, who had no knowledge of game theory. This result implies that knowledge of game theory lowers the numbers chosen during the beauty contest experiment. The relation between knowledge of game theory and the numbers chosen could explain why the secondary school students, who received less education compared to the undergraduate students, chose significantly lower numbers compared to the undergraduate students. The secondary school students received a lesson in game theory, prior to participating in the beauty contest experiment. The theorists also choose significantly lower numbers compared to undergraduate students. Their knowledge of game theory is more extensive than the knowledge of undergraduate students, confirming the possibility of a relation between knowledge of game theory and the numbers chosen during a beauty contest experiment. Because the group of newsgroup participants could contain any type of respondent, it was not possible to explain why these participants belonged to the same subset as the theorists and secondary school students.

The comparison of choices of secondary school students to the choices of adults during the beauty contest game demonstrated that age and education did not influence the numbers chosen. This outcome showed that results from beauty contest experiments with adult participants are generalizable to beauty contest with secondary school students as participants.

Knowledge of game theory appeared to lower the average of the chosen numbers. When performing a beauty contest experiment, the type of participants influenced the results less than their knowledge of game theory. This result confirmed that the beauty contest experiment could be used to measure the knowledge of the secondary school students about game theory.

Discussion

After comparing the performance of students in the experimental and the control condition, it was found that there was no significant difference between the results of a classroom experiment and a conventional class. That finding contradicted the hypotheses. The only significant difference between the teaching effect of the two classes was the extent to which the students enjoyed the classes. Students in the experimental condition experienced their treatment class as more fun than a regular economics class and also as more difficult, compared to students in the control condition. On the one hand, the result that there was no significant difference in performance across conditions contradicted the results of previous research on the teaching effect of classroom experiments (Frank, 1997; Gremmen *et. al.*, 1997; Dickie, 2006). On the other hand, the result agrees with research that found no significant difference in teaching effect of classroom experiments and conventional classes (Cardell *et. al.*, 1996; Emerson *et. al.*, 2004). The finding that the students who participated in the classroom experiment liked the difference with their regular classes more than students who attended the conventional class confirmed the only other study that has measured whether students liked classroom experiments more than conventional classes (Gremmen *et. al.*, 1997).

Neither this research nor previous research asked the participants why they enjoyed the classroom experiments. It is possible that the change of teaching method is enjoyable for the students, since most of the regular classes are all taught the same way. Students may also like the applied approach of classroom experiments more than the theoretical approach of conventional classes. Student interactions and an increase in freedom for the students could also be why they like classroom experiments better. No previous research reported that students thought that classroom experiments were difficult compared to conventional classes. Students may find classroom experiments to be more difficult because they have a more active role compared to conventional classes.

The lack of difference between the teaching across conditions could be explained in several ways. One explanation could be that the teaching effect of a classroom experiment was indeed the same as the teaching effect of a conventional class. If that were true, teachers could freely choose between the two teaching methods. Using classroom experiments would even be preferable, because classes with an experiment would be more fun and challenging for the students. When teachers have students participate in classroom experiments, they fulfil a different role in class compared to when they teach a conventional class. The teachers explain the rules of the experiment and see to it that the students abide by those rules. The teacher guides the students through the experiment, instead of directly explaining the subject matter to them. Because of the different role of teachers during classroom experiments, teachers may not be able to execute classroom experiments without training. They have to become familiar with classroom experiments and learn to supervise them.

There are other possible explanations for the lack of difference. All students received just one class on game theory, either a conventional class or a class with an experiment. It is possible that one class was not enough to make a difference. If that were true, one of the teaching methods could yield better results compared to the other, if more classes using the superior method were attended. Previous research already showed that the teaching effect of classroom experiments was higher compared to conventional classes when not just one but multiple classes were attended (Dickie, 2006). If undergraduate students score better on

tests because they participated in classroom experiments, the same could be possible for secondary school students. Based on one classroom experiment, students in the experimental condition enjoyed the experiment more compared to their regular classes. This research did not test whether secondary school students would still enjoy classroom experiments more if they were required to attend a series of experiments. Undergraduate students did enjoy a series of classroom experiments more than a series of conventional classes. The same could be true for secondary school students.

The results of this research can also have been influenced by the difference in subject matter across the two treatment classes. The subject of the classroom experiment was common knowledge, while the subject of the conventional class was the prisoner's dilemma. In both classes, the main message was that students had to take into account the choices of the other students. Although the two subjects were matched as much as possible, the difference could still have influenced the results. In order to compare teaching using classroom experiments to teaching in a conventional class, previous research also developed experiments and conventional classes that matched as much as possible. The subject was held constant among treatments. Still, the classes were not the same and the subject matter was not treated exactly the same during different classes. Previous research did not report influence of difference in subject among treatment classes. It is therefore unclear whether a difference in subject matter among treatment classes influenced the results or not.

It is also possible that no differences were found across conditions, because the wrong questions were asked. One test used in order to find a difference between the two conditions was the 'water lily' problem. Only a small part of the students did not solve that problem correctly. Either the students were already familiar with the problem, or it was too easy. The choices the students made during the beauty contest experiment would only be useful for a comparison of the two conditions if the motivations for the choices were clear. This was not always true, because some students were unable to express their motivation for the number they had chosen. The problem that resembled the regular method of testing knowledge in secondary school most was the question about the two businesses.

The classes were attended by the students in different groups at different moments. This was because the classes took place during the regular economics classes. In between the classes, the students in the three groups attended other courses together. Some of the students were therefore able to hear about the performance test before they participated in it themselves. The information they received from their fellow students could have influenced their answers, since a prize was handed out. Although students were asked at the end of each session not to share what had happened during that session with their fellow students, it was very well possible that they did so anyway. To be sure, the experimenter asked at the end of the last session whether any of the students had information about the performance test, prior to taking the test. This was confirmed by 8 students. Part of the students in the final two sessions had more information available while make their decisions compared to students in the first session. The second session of the performance test was attended by students in the control condition and the third session by students in the experimental condition. Students in both conditions could have been influenced by information they received from students who attended earlier performance tests. The results may have been influenced by this difference. In future research, this possibility should be taken into account. This could be done by letting the students take the tests all at the same time or by postponing the announcement of the winners until after all students in all conditions have taken the tests.

The number of students in the control condition was small. Since the power of the statistical tests used in this paper depend on the number of cases, the power might not have been sufficient to detect differences when there were actually differences. Also, the difference in distribution of girls over the conditions was significant. When comparing some variables, part of the participants had to be excluded from the analysis to enable controlled comparison. If more classes of students were included in the experiment, the division of girls and boys in the conditions could have been more equal and no students would have had to be excluded. In future research, more participants should be used, to be better able to check whether there are significant differences between the conditions.

Multiple explanations for a lack of difference between the teaching effect of the two conditions are possible. It is therefore too early to conclude whether teaching using classroom experiments is effective or not. First, the alternatives, using more classes, using the same subject matter in the treatment classes, measure effects using different questions, preventing cheating and using more participants should be explored and perhaps eliminated.

During the analysis of the open questions of the questionnaire, it seemed that the questions about the motivation of the students to choose their number were not interpreted the same way by all the students. Although several researchers looked at the questions and approved them, no pilot study was conducted to test the questionnaire. If some of the questions were unclear or open to interpretation by the students, the questions might not have measured what they were supposed to. It was therefore possible that the quality of the questionnaire influenced the results. To prevent this from happening during future research, the questionnaire should be tested before it is used to collect data.

The distribution of the students across the different levels of reasoning could have been influenced by the experimenter who assessed the level of reasoning. This might have influenced the results. It would have been preferable to have multiple persons assign the students a level of reasoning, in order to increase the reliability.

Students in the experimental and the control conditions did not perform significantly different during the beauty contest experiment. This could indicate that the teaching effect was not significantly different across conditions. However, it was also possible that the beauty contest experiment was not well suited to measure a difference in performance. To test whether the beauty contest experiment could be used to find a difference in teaching effect, the choices made by different types of participants during beauty contest experiments were compared. Age and education did not influence the choices of the participants, knowledge of game theory did. Increased knowledge of game theory lowered the average numbers chosen during the experiment. The beauty contest experiment could be used to measure a difference in performance across conditions. The result that age and education did not influence the choices of the participants showed that results from beauty contest experiments with one type of participants can be generalized to other types of participants.

Recommendations

This research did not confirm whether teaching using a classroom experiment was more effective compared to teaching using a conventional method. No significant difference in performance was found across conditions. A possible conclusion was that a classroom experiment makes students perform as well as a conventional class does. This was not the only possible explanation for the lack of difference between the two methods. To confirm whether the result of this research is true, more research is necessary.

Students who participated in the classroom experiment did enjoy it significantly more than regular classes, compared to students who attended a conventional class. Students in the experimental condition also judged their treatment class as significantly more difficult compared to how students in the control condition judged their treatment class. Neither this research, nor previous research with similar findings asked the students why they liked classroom experiments better or why they judged classroom experiments to be more difficult than their regular classes. Future research into classroom experiments should also ask the students why they liked or disliked the classes. The opinion of participants can then be used to improve the appeal of classroom experiments even more.

In future research, secondary school students in both conditions should not receive just one game theory class prior to participating in the performance tests. Instead, the students should receive multiple classes, either only classroom experiments, only conventional classes or possibly a combination of both. By adding a condition in which students attend classroom experiments as well as conventional classes, it could be tested whether a combination of the two teaching methods is even more effective compared to using only one method. When a series of classroom experiments is compared to a series of conventional classes, differences in teaching effect across conditions should become more pronounced. The opinion of the students about classroom experiments after participating in multiple experiments can then also be measured. This way, it can be tested whether students will still enjoy classroom experiments more than conventional classes after they become accustomed to classroom experiments.

If, as in this research, students in different groups participate in the performance tests at different times, they can talk to each other about the tests. This could influence the results. For future research, it is therefore recommended that all students participate in the performance tests at the same time. That way, students are not able to communicate about the performance tests. A possible way to have all performance tests take place at the same time is by adjusting the schedule for some classes to make the economics classes all fall at the same time. Another possibility is one (or multiple and simultaneous) session later on the day, after all other classes have finished, in which the students all take the tests at the same time.

The questions used to measure differences between students across conditions were probably not all equally well chosen. The question about the lilies on the pond should not be used in future research, since it was too simple. The beauty contest experiment can be used to measure performance. The measure can be more effective if the questions about the motivation for the choices are improved. The questions should resemble questions of end of period tests more.

The subject of the classroom experiment was common knowledge. The experiment used to teach the students about common knowledge was used before in the literature a few times. As with the beauty contest experiment, the participants in experiments about

common knowledge in other research were adults. It was therefore interesting to observe the behaviour of secondary school students during a common knowledge experiment. The experimenter who oversaw the classroom experiment was also tasked with observing the behaviour of the students. It proved to be too complex to guide the students and at the same time carefully observe their behaviour and their choices. For this reason, the behaviour of the students during the common knowledge experiment was not mentioned in this paper. In future research using common knowledge experiments with secondary school participants, a second experimenter or a video camera for observing the students is recommended.

The comparison of the choices of secondary students and adults during a beauty contest experiment revealed a possible influence of knowledge of game theory with the numbers chosen during the experiment. Previous research shows a similar result, that level of training in game theory influences level of reasoning used during a beauty contest experiment (Bosch-Domènech *et. al.*, 2002). However, the relation found between knowledge of game theory and choices during a beauty contest experiment was circumstantial in both studies. In order to learn more about this relation, more thorough research of knowledge of game theory and choices during a beauty contest experiment is necessary.

References

Alba-Fernández, V., Brañas-Garza, P., Jiménez-Jiménez, F., Rodero-Cosano, J. (2006) Teaching nash equilibrium and dominance: A classroom experiment on the beauty contest, *Journal of Economic Education* 37 (3), pp. 305-322

Bosch-Domènech, A., Montalvo, J.G., Nagel, R., Satorra, A. (2010) A finite mixture analysis of beauty-contest data using generalized beta distributions, *Experimental Economics* 13 (4), pp. 461-475

Bosch-Domènech, A., Montalvo, J.G., Nagel, R., Satorra, A. (2002) One, two, (three), infinity, ... : Newspaper and lab beauty-contest experiments, *American Economic Review* 92 (5), pp. 1687-1701

Brauer, J., Delemeester, G. (2001) Games economists play: A survey of non-computerized classroom-games for college economics, *Journal of Economic Surveys* 15 (2), pp. 221-236

Cardell, N.S., Fort, R., Joerding, W., Inaba, F., Lamoreaux, D., Rosenman, R., Stromsdorfer, E., Bartlett, R. (1996) Laboratory-Based Experimental and Demonstration Initiatives in Teaching Undergraduate Economics, *American Economic Review* 86 (2), pp. 454-459

Dickie, M. (2006) Do classroom experiments increase learning in introductory microeconomics?, *Journal of Economic Education* 37 (3), pp. 267-288

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

Feltovich, N. (2011) What's to know about laboratory experimentation in economics?, *Journal of Economic Surveys* 25 (2), pp. 371-379

Frank, B. (1997) The impact of classroom experiments on the learning of economics: An empirical investigation, *Economic Inquiry* 35 (4), pp. 763-769

Frederick, S. (2005) Cognitive Reflection and Decision Making, *Journal of Economic Perspectives* 19 (4), pp. 25-42

Gremmen, H., Potters, J. (1997) Assessing the efficacy of gaming in economic education, *Journal of Economic Education* 28 (4), pp. 291-303

Nagel, R. (1995) Unraveling in guessing games: An experimental study, *American Economic Review* 85 (5), pp. 1313-1326

Selten, R., Nagel, R. (1997) 1000 DM zu Gewinnen, *Spektrum der Wissenschaft* 3 (11), pp. 10

Stahl, D.O. (1998) Is step-j thinking an arbitrary modelling restriction or a fact of human nature?, *Journal of Economic Behavior and Organization* 37 (1), pp. 33-51

Bergstrom, T.C., Miller, J.H. (2000) *Experiments with economic principles: Microeconomics*, Irwin McGraw-Hill, New York

Burkett, J.P. (2006) *Microeconomics: Optimization, Experiments, and Behaviour*, Oxford University Press, Oxford

Geanakoplos, J. (1994) *Handbook of Game Theory*, Vol. 2 Elsevier, Amsterdam

Littlewood, J. E. (1953) *A Mathematician's Miscellany*, Meuthen & Co. Ltd, London

O'Sullivan, A., Sheffrin, S.M. (2003) *Economics*, Prentice-Hall, New York

Teulings, C.N. (2005) *The Wealth of Education*, SLO, Enschede

Yandell, D. (1999) *Using economic experiments in the classroom*, Prentice-Hall, Upper Saddle River, NJ

Bosch-Domènech, A., Nagel, R. (1997) Guess the number: comparing the Financial Times and Expansion's results, *Financial Times*, June 30, Sec. 8, pp. 14

Thaler, R. (1997) Giving markets a human dimension, *Financial Times*, June 16, Sec. 6, pp. 2-5

Appendix

A: Scenario common knowledge experiment

"During the 18th century a Spanish ship loaded with gold sank to the bottom of the Atlantic Ocean during a storm. The site of the shipwreck was unknown for many years. Recently, a billionaire who collects lost treasures discovered the site of the shipwreck. The billionaire is not the only one in possession of the location of the shipwreck. Others also want the treasures within the shipwreck. The billionaire wants to salvage the gold as soon as possible, to prevent others from finding the gold before him."

"The billionaire has hired two divers to salvage the lost gold using a boat equipped especially for deep sea diving. On board of the boat are a number of diving bells for descending into deep waters. The diving bells are either green or blue, manufactured by two different manufacturers. The boat arrives on the site of the shipwreck on a clouded, dark night. Because there is no time to lose, the divers immediately descent into the water, each in their own diving bell."

"Although the divers were hired by the billionaire and so have the same goal, they are also competitors. This is because the billionaire offered a large reward to the diver who salvages a unique golden cross which is part of the sunken treasure."

"An hour after the divers started descending they are nearing the bottom of the Ocean and the sunken ship. As part of the interior of the diving bell, a diver can see a speaker box. The captain of the boat told the divers that the speaker boxes can be used by him to speak to both divers simultaneously. The divers do not have any means to talk back, however. Another part of the interior of the diving bell is an emergency button which a diver can press to let the captain know that he is in danger and wants to be lifted up. The diving bells have windows and the divers can see each other's diving bells, thanks to lights on top of each diving bell. The divers can see the colour of the other diving bell, but not the colour of their own diving bell. Each diver sees the other diver looking through the window, watching him."

"Suddenly the voice of the captain sounds through the speaker boxes. The captain has an alarming announcement for the divers. All blue diving bells are faulty. When the blue diving bells descent to deep they will rupture and flood. The captain does not know the colour of the diving bells. Divers in a blue diving bell have to press the emergency button to save themselves from drowning. Divers in a green diving bell are save and can continue descending."

"Due to the dark night, the divers themselves were also unable to see the colour of their own diving bells. A diver in a blue diving bell has to press the emergency button to save himself from drowning. A diver in a green diving bell can continue descending to salvage the gold and try and finding the unique golden cross. If a diver presses the emergency button while he is in no danger of drowning, odds are that someone else has salvaged the gold before the diver returns to the shipwreck in a different diving bell."

What would you do if the choice was yours?"

B: Possible line of reasoning with two diving bells

When the captain announces to the divers that all blue diving bells are broken and that at least one blue diving bell is being used, the students only know the colour of the diving bell of the other student. That diving bell can be green, or it can be blue. If it is green, the student should conclude that his or her own diving bell is blue, since he or she knows that at least one diving bell is blue. Having come to that conclusion, the student should start ascending to prevent him- or herself from drowning. If the colour of the other diving bell is blue, the student cannot know for sure the colour of his or her own diving bell. He or she then needs to take into consideration the fact that the other student can see the colour of the diving bell the first student is in. In addition, since it is common knowledge that there is at least one blue diving bell in use, the first student also knows that the second student knows this. If the second student sees a green diving bell, he or she will know that he or she is in danger of drowning and will start ascending. If the second student sees a blue diving bell, he or she cannot yet know for sure the colour of his or her own diving bell.

When a diver sees a blue diving bell, the best course to take is to remain seated. That way, the diver can see what the other diver does and base his or her decision on that additional piece of information. If the other student stands up and thus starts ascending, he or she must see a green diving bell and the first student can continue descending. If the second diver also remained seated, both divers see a blue diving bell and should start descending.

C: Possible line of reasoning with three diving bells

All students could use the following line of reasoning. The first thing students should have checked was the colour of the other two diving bells. Three combination were possible. The other diving bells could both be green, there could be one green and one blue diving bell, or they could both be blue. If a student saw two green diving bells, that meant that the student was him- or herself in the blue diving bell, since there was at least one blue one. The student who saw two green diving bells should therefore stand up during the first opportunity to do so. If a student did not see two green diving bells, the first decision was not as easy. The student could not know for sure the colour of his or her own diving bell. The best choice to make in that case was to remain seated and see what the other two students would do. If one of the others stood up when the opportunity to do so was first presented, that student signalled that the other two were inside green diving bells and they were save to keep descending.

If no student stood up, the other two students both see at least one blue diving bell and could not make up their mind during the first decision moment. If the student saw a green and a blue diving bell and both other students remained seated, the student could conclude that he or she was in a blue diving bell and should have stood up. The student knew this, because the other student in a blue diving bell saw the first students colour and the green bell which they both could see. Since the second student did not stand up, he or she saw a blue diving bell and the first student had to be in it.

The last possibility was that the student saw two blue diving bells. In that scenario, it was not possible that another student, if he or she did not make a mistake, has stood up at the first possibility to do so. During the second possibility to stand up, the student who saw

two blue diving bells was still not able to know the colour of his or her own diving bell. To find out, the student should have remained seated and watched what the other two students would do. If the student was in a green diving bell, the others would stand up based on the logic explained earlier. So if that happened, the student could remain seated, because he or she knew that his or he diving bell was green and save. If the student was in a blue diving bell, the others will also remained seated. If after the second opportunity to stand up all three students were still seated, they should have all stood up at the last opportunity to do so, because they all saw two blue diving bells and therefore all three were inside a blue diving bell.

D: Questionnaire (experimental condition)

1. What number did you just write down on your piece of paper?

…….....

2. Please explain why you chose the number you did, instead of any other number.

…….....

…….....

…….....

3. How do you hope that the other students reasoned while choosing a number?

…….....

…….....

…….....

Suppose that your teacher participates in the beauty contest experiment together with 29 other economics teachers.

4. Will the outcome of the experiment be different from the outcome of the experiment in which you participated?

- **No** *(skip question 5 if this is your answer)*
- **Yes**

5. What kind of difference do you think there will be with the outcome of the experiment with a group of economics teachers?

…….....

…….....

…….....

The next two questions will be about reasoning, just like the experiment.

6. A small part of a lake is covered in water lilies. Each day, the part of the lake that is covered by lilies becomes twice as large. If it takes 48 days for the lilies to cover the whole lake, how long did it take for the lilies to cover half the lake?

……

7. Two companies from the same industry have the choice whether they want to advertise or not. Company A is a smaller company compared to company B. If company B advertises, almost all interest goes out to the company and company A will earn less compared to when company B does not advertise. When company B does not advertise, company A will have good profits and advertising will cost more money than it makes.

Below, the profits (in thousands of Euros) of both companies are given when they advertise or when they do not advertise. The box in the lower right corner for example shows that both companies will have a profit of 400 Euros if they both do not advertise. The box above it shows that company A will have a profit of 300 when it advertises while company B does not advertise. Company B has a profit of 250 in that case.

		Company B	
		Advertise	Not advertise
Company A	Advertise	Profits company A: 200 Profits company B: 500	Profits company A: 300 Profits company B: 250
	Not advertise	Profits company A: 100 Profits company B: 750	Profits company A: 400 Profits company B: 400

a) Does company A have a strategy which is always superior compared to the other strategy? If yes, which strategy is that?

……

b) Does company B have a strategy which is always superior compared to the other strategy? If yes, which strategy is that?

……

c) If you were an employee of the marketing department of company A, would you advertise or not?

……

……

In the table on this page you can see 5 statements about the previous class you attended, the one with the experiment. For each statement, please indicate the extend to which you agree with it. You can do so by ticking the appropriate box in each row.

	Completely agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Completely disagree
The previous class with an experiment was fun	☐	☐	☐	☐	☐
The previous class with an experiment was more fun than other economics classes	☐	☐	☐	☐	☐
The previous class with an experiment was interesting	☐	☐	☐	☐	☐
The previous class with an experiment was difficult	☐	☐	☐	☐	☐
The previous class with an experiment was a useful preparation for the experiment	☐	☐	☐	☐	☐

8. Are you a boy or a girl?

- ☐ **Boy**
- ☐ **Girl**

9. What is your economics grade?

……

10. What kind of mathematics do you attend?

- ☐ **Mathematics A**
- ☐ **Mathematics B**

11. What is your mathematics grade?

……

Thank you!

E: Procedure matched samples

The students in the experimental condition were matched to students in the control condition as follows. In both conditions the students were separated into boys and girls. Additionally, they were separated in students studying mathematics A and students studying mathematics B. This meant that the students in the two conditions were divided into four groups. Afterwards, a student was matched to a student in the corresponding group in the other condition, based on their mathematics grade. If possible[2], the students with the grades closest to each other were matched. Table 2 shows all students in the control condition and the students they were matched with.

Table 6: Students in the control condition matched with similar students in the experimental condition. The students were matched on sex, type of mathematics and mathematics grade, in that order. Students in *italics* could not be matched on sex, because there were more girls in the control condition than in the experimental condition.

Control condition			Experimental condition		
Sex	**Type of mathematics**	**Grade**	**Sex**	**Type of mathematics**	**Grade**
Girl	B	7.3	Girl	B	8.1
Girl	B	6.2	Girl	B	8
Girl	B	5.9	Girl	B	7
Girl	B	5.8	Girl	B	6.5
Girl	B	4.4	Girl	B	6.5
Girl	A	7	Girl	A	7
Girl	A	6.4	Girl	A	6.4
Girl	A	6.3	Girl	A	6.3
Girl	A	5.8	Girl	A	6
Girl	A	5.6	Girl	A	5.3
Boy	B	7.5	Boy	B	7.7
Boy	B	6.5	Boy	B	6.6
Boy	B	6.4	Boy	B	6.5
Boy	B	6.3	Boy	B	6
Boy	A	7.5	Boy	A	7.5
Boy	A	7.4	Boy	A	7.4
Boy	A	7.2	Boy	A	7
Boy	A	6.9	Boy	A	7
Boy	A	6.9	Boy	A	6.8
Boy	A	6.1	Boy	A	6.1
Boy	A	5.6	Boy	A	6
Boy	A	5.4	Boy	A	5.6
Girl	*A*	*9*	*Boy*	*A*	*8.4*
Girl	*A*	*8*	*Boy*	*A*	*7.4*
Girl	*A*	*8*	*Boy*	*A*	*6.5*

The last three rows of table 2 show that 3 girls were matched to boys. The reason for this is that there were more girls in the control condition compared to the experimental

[2] Sometimes students had similar grades and a student was chosen at random to be matched with a student in the other condition. Other times a student with a mathematics grade second closest to that of a student in the other condition had to be matched, to facilitate a reasonable match between two other students.

condition. These 6 students were included in the analysis to not further reduce the number of participants and because the students were matched on the other control variable.

Of the 50 matched students, 18 study mathematics B and 32 study mathematics A. 20 of the 50 students are girls, 30 are boys. All students who were not matched and as a consequence not included in the analysis were boys. Among those 12 boys, 8 studied mathematics B and 4 mathematics A. The students who were not included in the analysis are shown in table 3.

Table 7: Students who were not included in the analysis.

Sex	Type of mathematics	Grade
Boy	B	7
Boy	B	7
Boy	B	6.9
Boy	B	6.8
Boy	B	5.6
Boy	B	5.6
Boy	B	5.2
Boy	B	5
Boy	A	6.5
Boy	A	6.4
Boy	A	6.3
Boy	A	6.2

F: Score rules question 7

Part a)

The question of this part is: “Does company A have a strategy which is always superior compared to the other strategy? If yes, which strategy is that?”

This part is worth 2 points. A student will only receive full points for answering that there is no superior strategy. 1 point can still be earned when the student says one strategy is better than the other and gives a plausible reason why. Because the result of the choice of company A depends on the choice of company B, it is never completely right to say one strategy is better than the other.

Table 8: Number of points received for the first part of question 7 based on the answer and the motivation of the student.

Yes or no	Motivation	Points
No	-	2
No	Valid reasoning	2
Yes	-	0
Yes	Advertise	0
Yes	Not advertise	0
Yes	So the other company will earn less	1
Yes	The average profits will be higher	1

Part b)

The question of this part is: “Does company B have a strategy which is always superior compared to the other strategy? If yes, which strategy is that?”

This part is worth 2 points. A student receives full points if he or she answers that company B has a superior strategy. That strategy is to advertise. If a student answers that the superior strategy for company B is not to advertise, that answer is worth 0 points. No plausible reasoning is possible for that answer.

Table 9: Number of points received for the second part of question 7 based on the answer and the motivation of the student.

Yes or no	Motivation	Points
Yes	Advertising always leads to the best outcome	2
Yes	-	2
Yes	When company B advertises, company A will always make the least amount of money	1
Yes	Average profits are highest	1
Yes	Profits will be higher than profits of company A	1
No	-	0
No	Every line of reasoning	0

Part c)

The question of this part is: "If you were an employee of the marketing department of company A, would you advertise or not?"

This part is worth 3 points. The third part is worth more points, since the best choice for company A depends on the best choice for company B. Because the students can reason what choice company B almost certainly makes, they can now choose the corresponding best strategy for company A. To realize this, the choice of company B should be taken into account. Because of this, part c is more complicated than the other two parts and worth more points.

The correct answer is for company A to advertise, since the students can assume that company B would advertise. This answer is worth the full 3 points. Students who answer that company A should advertise without explaining why receive 2 points, *if they answered that company A has no superior strategy in part a.* Answering yes with another reasoning than the one given above can be worth between 0 and 2 points. Every student who answers that company A should not advertise receives 0 points, unless they have a very convincing reason for their answer, in which case they can receive at most 1 point. Judging this part is the most difficult, since not all possible answers and lines of reasoning can be prepared for.

Table 10: Number of points received for the third part of question 7 based on the answer and the motivation of the student.

Yes or no	Motivation	Points
Yes	Company B will advertise	3
Yes	- ; part a was answered with no	2
Yes	Highest average profits	1
Yes	Highest minimal profits	1
Yes	Decreases the profits of company B	1
Yes	Advertising is necessary for competition	0
No	Advertising is expensive, if not company A should advertise	1
No	-	0
No	A convincing motivation	1
-	Only when company B also advertises	1

G: Overview of subject groups beauty contest experiments

Table 11: Number of participants, mean and standard deviation of the numbers chosen by different subject groups during beauty contest experiments.

Subject group	Number of participants	Mean of the chosen numbers	Standard deviation of the chosen numbers
Secondary school students	59	20.75	9.977
Business undergraduates	128	32.95	20.460
Undergraduates from various departments	85	35.16	19.662
Economics undergraduates	138	26.78	17.716
Economics undergraduates (take home)	119	25.15	16.979
Theorists	145	16.95	21.634
Newsgroup participants	149	21.74	20.143