Economics of Education Review 29 (2010) 364–374

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

# Peer effects, motivation, and learning

Gerald Eisenkopf

*University of Konstanz, Department of Economics, 78457 Konstanz, Germany*

ARTICLE INFO

*Article history:*
Received 5 September 2008
Received in revised form 20 July 2009
Accepted 4 August 2009

*JEL classification:*
C90
C92
I20

*Keywords:*
Learning
Peer effects
Motivation
Experiment
Economics of education

ABSTRACT

This paper confirms the existence of peer effects in a learning process with data from an experiment. The experimental setting offers an insight into the mechanisms of peer interaction and provides complementary information to empirical studies using survey or administrative data. The results show that a partner has a motivational effect even before the actual cooperation takes place. The evidence for optimal group composition is not robust. Some of the "better" students improve the performance of their partner but they induce lower motivation.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

It is a long-standing hypothesis that the performance of a student depends on the behavior and characteristics of his fellow students. The resulting benefit typically differs with the composition of the peer group. Therefore the existence and properties of peer effects influences the welfare implications for the (self-)selection of students into different learning groups. Since many policies shape selection processes in education, peer effects may have significant implications for the evaluation of these policies (Arnott & Rowse, 1987).

Econometricians invest great effort to identify quasi-experimental evidence on peer effects from administrative or survey data. The ideal data set would observe the same individual at the same time in the same environment but with a different peer. This particular type of data set is not available. Hence, researchers actually face a trade-off between internal and external validity. Almost all contributions in the literature take non-experimental data from surveys or administrations and turn them into quasi-experimental evidence as far as possible. An alternative is to generate data in an experiment with key characteristics of a real world learning process.

This paper follows this experimental path which is complementary to the econometric approach. More specifically, this experiment circumvents four problems which typically constrain the analysis of peer effects with field data: firstly, the (self-)selection of students into peer groups[1]; secondly, the impact of changes in the peer group

[1] The econometric literature provides various approaches to this problem. Gould, Lavy, and Paserman (2005), Hoxby (2000), and Ammermüller and Pischke (2006) for example rely on differences in the compositions of individual classes. Hanushek, Kain, Markman, and Rivkin (2003) and McEwan (2003) use variation within peer groups over time to identify peer effects. Stinebrickner and Stinebrickner (2006), Foster (2006), Zimmerman (2003) and Sacerdote (2001) investigate peer effects in universities in a different way, they use the random assignment of students into university dormitories as indicators for social tie formation and its subsequent impact on educational performance.

*E-mail address:* Gerald.eisenkopf@uni-konstanz.de.

0272-7757/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2009.08.005

on the learning environment,[2] thirdly, the endogeneity of peer effect measures, e.g. as characterized by the reflection problem (Manski, 1993); and, last but not least, the unclear mechanisms of peer interaction. Typically, field data do not inform *how* students influence each other and what happens if students do not cooperate at all. Survey or administrative data largely provide a 'peer group composition effect'.

The experiment in this paper covers a learning process. The subjects face a task on which they can improve over time, a logical puzzle game called Kakurasu (description see below). Some subjects can improve with a randomly assigned partner (pair treatment), others have to do it alone (single treatment). The assignment to the treatments is random as well. Environmental conditions do not change over time and there is no teacher. Hence, any difference in performance between both treatment groups at the end of the experiment derives from the differences in the treatment and constitutes a peer effect.

In the experiment each participant was tested before and after the actual treatment. This allows an insight into the mechanisms of peer interaction. The focus in this paper is on the motivational aspects of peer interaction during a learning process. More specifically, I test if the assignment of the prospective partner motivates a partner. Since the partner was presented before the first test, performance differences in this first test identify motivational differences.

The second performance test after the treatment provides information about the overall peer effect. Beside motivation, peer interaction can influence learning processes in other ways. Subjects may explain the task to each other. Such an instructional argument is a standard assumption for peer effects in models of educational production. Therefore, Rothschild and White (1995) describe education as a customer-input-technology. In the context of this experiment, one can estimate if the actual peer interaction increases the learning benefits derived from the additional motivation.

The treatment group with pair learners also provides some information about the optimal composition of learning groups. I can test if and how various characteristics of a partner influence the performance of a subject in the first and the second test.

Motivation has largely been ignored in the econometric literature on peer effects in education. A partner can motivate even if the basic learning technology of single learner and pairs do not differ significantly. People do not want to appear as being a lazy, for example. Falk and Ichino (2006) find experimental evidence for positive peer effects in a "real task", non-cooperative production environment. In their case the peer effect stems from the mere presence of another person in the same room doing the same job. Their study confirms the "social facilitation paradigm", a research topic in the psychological literature (e.g., Cottrell, Wack, Sekerak, & Rittle, 1968; Feinberg & Aiello, 2006; Platania & Moran, 2001; Zajonc, 1965). Quarter and Marcus (1971) call this change in behavior an audience effect. Mas and Moretti (2006) find similar results with personnel data from a grocery store chain. Social pressure apparently causes the motivation.

Note that the anticipation of this instructional advantage can act as an incentive, too. If the abilities of the partners are complementary inputs (like in De Fraja & Landeras, 2006) the marginal gain from investing effort is obviously lower in the single treatment groups, In a similar way, the subjects may also distract each other (see Lazear, 2001).

A key difference between this experiment and the one in Falk and Ichino (2006) is the timing of measurement. In their experiment, the subjects face different contexts during the measurement period. In contrast, the participants in this experiment face the same context in the measurement period but differences in the preparation period. To use educational terminology, the tests are standardized but the learning environment differs. There are some papers in education and psychology which test the performance of pair learners against that of single learners (e.g. Fleming & Alexander, 2001; Manion & Alexander, 1997). Many studies are not comparable with the experiment in this paper as they do not isolate the partners during performance measurement, therefore allowing for complementarities in the production process (e.g. Lazonder, 2005; Lui, Chan, & Nosek, 2008). None of these uses ex-ante tests to identify motivational differences. In this context the experiment is without precedent in the economic literature. I did not find anything similar in other disciplines like sociology, psychology or education.

The results show a significant peer effect. Subjects with a partner performed better in both tests. This result is in line with similar experiments (Falk & Ichino, 2006; Fleming & Alexander, 2001; Manion & Alexander, 1997) and meta-analyses of non-experimental studies on individual vs. group learning in the education literature (e.g. Lou et al., 2001). The results provide evidence for the motivational impact of peers since the subjects with a prospective partner are already better in the first test than those subjects in the single treatment group. The results provide only little evidence with respect to an optimal group composition but motivation increases with the ability of the partner. Students with an interest in logical puzzles and good performance in math improve the final test score of their partners but they induce lower motivation. This result is a possible explanation why different econometric studies come to contradicting conclusions about the properties of peer effects.

The paper is organized as follows. The next section describes the experimental design. Section 3 provides the behavioral expectations and discusses the design. Section 4 contains the results and Section 5 concludes.

[2] For example, a teacher may teach the same topic in a different way, if the average ability or the ability distribution changes in a class. Arguably, such an effect is part of a peer effect. One could distinguish between a direct peer effect, where students directly influence each other, and an indirect one, where students influence each other via the teacher.

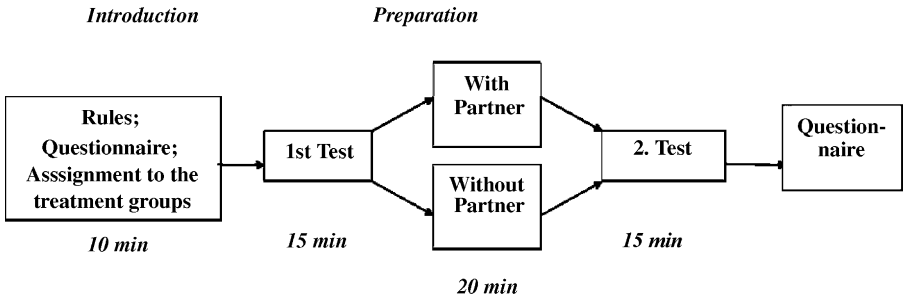



**Fig. 1.** The design of the experiment.

## 2. The experiment

### 2.1. Design of the experiment

The implementation of the peer effect design is rather straightforward (see Fig. 1). The subjects faced a learning process for a specific task. The task was a largely unknown logical puzzle called Kakurasu which will be described in greater detail below. At the beginning the participants were assigned to their treatment groups and their prospective partner. The partner was a class-mate of the same sex. This matching procedure implies that the subjects know each other and have a rather precise idea about the partner's capabilities. The subjects received some general rules (see Appendix A) and faced a first test about the task. The test measured how many puzzles the subjects solved in 15 min. The assignment to the treatment groups before the first test allows the identification of motivational differences between the treatment groups.

After the first test, subjects could prepare for 20 min for a second test. In this preparation period the treatment took place. Some subjects had to prepare alone (single treatment), others with a partner (pair treatment). All subjects got further puzzles, pens and writing paper. The subjects in the pair treatment were allowed to talk to their assigned partner during this preparation period. The subjects learned at the beginning of the experiment about their treatment and the prospective partner.

After the preparation period the second test started. Again, the test measured how many puzzles the subjects solved in 15 min. However, the puzzles were more difficult in this test. Questionnaires were handed out before and after the experiment.

### 2.2. The task

In the experiment the participants can learn a solution strategy for the logical puzzle (see Appendix B[3]). This type of puzzle consists of a matrix in which the correct fields have to be marked. Numbers at each line and column of the matrix allow the subject to derive the logically correct solution.

The task satisfies several criteria. Subjects can learn the logic of the puzzle within a reasonable time. The puzzle can vary in difficulty by increasing or decreasing the size of the matrix. Unlike the famous Sudoku, the puzzle used here is largely unknown. Hence many subjects start from zero.[4]

The matrices in the first and second tests are of different sizes ($4 \times 4$ and $5 \times 5$, respectively). This variation is used to avoid floor or ceiling effects in the learning process. The experiment has been pre-tested with first year university students. Here, the matrices were of identical size in both tests, to show if subjects actually improve. Using $5 \times 5$ matrices for puzzles in the first test caused a floor effect. Only a few people actually managed to solve a single puzzle. On the other hand, $4 \times 4$ matrices for puzzles in the second test are too easy to solve. Many participants did not improve from test 1 to test 2 (a ceiling effect). In the preparation period, the subjects got $4 \times 4$, $5 \times 5$ and $6 \times 6$ puzzles.

### 2.3. The procedure

The first experiment was conducted on the 5th of December in 2006 with 85 Swiss students at a public high school (Kantonsschule) in Kreuzlingen in the Canton of Thurgau in Switzerland. A second experiment took place on 26th February, 2007 with 61 students from a similar school in Romanshorn in the same canton. Both schools roughly capture the top 15% of the students in their hometown.

The students were recruited from the top three grades (age 15–18) through the school's intranet. Students responded via E-Mail and provided information about their grade and sex. Each participant got 20 Swiss Francs (about 12.40€ or 16.25 US$ in December 2006) for their participation. The single treatment included 48 participants and 96 were assigned to the pair treatment group. The subjects were assigned randomly to the different groups. Each subject in the pair treatment group got a randomly assigned partner, but only from the same class level and sex. Due to missing partners three pairs were formed *ad hoc* with subjects from different grades. All partners learned at the beginning of the experiment about their prospective partner. Table 1 shows the composition of single treatment and pair treatment groups.

[3] A detailed description of the puzzle can be found at www.janko.at (in German). The owners of the website created the meaningless name for the puzzle.

[4] Similar puzzles exist on dedicated homepages on the internet. Hence, it is likely that some participants have an advantage even though they do not know the puzzle.

**Table 1**
The distribution of the subjects into single treatment and pair treatment groups.

| Grade | Single treatment group | | | Pair treatment group | | |
|---|---|---|---|---|---|---|
| | Male | Female | Sum | Male | Female | Sum |
| 2 | 11 | 10 | 21 | 27 | 18 | 45 |
| 3 | 8 | 7 | 15 | 19 | 10 | 29 |
| 4 | 7 | 6 | 13 | 16 | 6 | 22 |
| Sum | 26 | 23 | 49 | 62 | 34 | 96 |

All subjects of a school did the experiment at the same time to ensure that students could not communicate solution hints to following students. In Kreuzlingen, the subjects did the experiment in five different rooms in the school. Two rooms were filled with single learners, three rooms with the pair treatment group. The differences across rooms within a specific treatment group are insignificant. In Romanshorn, the students were separated in two different rooms according to their treatment. All the students received their instructions in standardized oral and written form from the author of this paper. Each room contained either 20 or 40 persons. One person per 20 participants was in charge of the technical details (i.e. two persons were in each large room). These supervisors received instructions about the procedure of the experiment but not about the puzzle. The participants were explicitly told that the overseer could not answer questions with respect to the puzzle.

### 2.4. Behavioral expectations

Differences between the treatment groups in the first test, i.e. *before* the learning process starts, reflect differences in motivation. At that point of time, the partners had no time for cooperation. Any difference between the treatment groups is caused by this anticipation of interaction. From the beginning subjects in the pair treatment knew about their prospective partner and that they will cooperate face to face. As in Falk and Ichino (2006), this setup should support social facilitation mechanisms and induce motivation.

**Prediction 1.** Subjects in the pair treatment perform better in the first test than single learners.

Differences in the second test confirm the existence of peer effects. It is reasonable to assume that motivation has an important impact on the learning process and the final performance. Furthermore, the subjects in the pair treatment can support each other in the learning process.

**Prediction 2.** Subjects in the pair treatment also perform better in the second test than single learners.

The econometric as well as the theoretic literature on peer effects has focused on the cognitive aspects of peer interaction, e.g. the ability of the learning partners. The characteristics of a learning group matters if the performance of a subject in the pair treatment group increases or decreases with some exogenous characteristics of the partner. Most theoretical models on educational production assume that learning processes improve with the ability of the relevant peers. (e.g. De Fraja & Landeras, 2006; Lazear, 2001). We base the following prediction on this assumption but the empirical evidence for it is unclear. The econometric literature has investigated this assumption in various ways and with different conclusions.[5]

**Prediction 3.** The performance of a subjects in the pair treatment perform increases in the ability of the learning partner.

## 3. Discussion of the design

The random assignment of subjects into different treatment groups ensures that self-selection does not play a role. All subjects face the same problem. Within the pair treatment, the groups are of identical size. While verbal interaction between partners was allowed during the preparation period, all subjects faced the same conditions during the tests and the analysis is based on the results of these tests. As stated above, this is a key difference to Falk and Ichino (2006).

The results of the experiment are likely to change with the task, the subject pool or the pay-out formula. The floor and ceiling effects identified in the pre-tests support this presumption. A change in the result does not imply that the design is inappropriate. The fixed payment keeps the setup as simple as possible. If financial incentives were the only source of motivation the results in the first test should not differ within and across treatment groups. Since students know each other and cooperate face to face in the preparation period, financial incentives could induce some of them to promise post-experimental rewards or punishments to elicit the desired level of cooperation. Furthermore, the results from Gneezy and Rustichini (2000) show that the impact of financial incentives on performance in tasks with a great cognitive demand is not trivial. Hence, a much larger experimental setup would be necessary which would extend beyond the objectives of this paper. Finally, Deci and Ryan (1985) claim that human beings' "need for competence" is a source of intrinsic motivation equivalent to basic needs for autonomy and social relatedness. This assumption implies that subjects with different marginal productivity will provide different test results. The experiment allows for investigating if this premise is too simplistic.

The identification of motivational differences between the treatment groups requires that the subjects know about their assignment before the first test. Otherwise no behavioral variable would distinguish between motivational and other aspects of cooperation. This setup does not allow a distinction between different sources of motivation nor exactly quantify the motivational impact.

## 4. Results

Table 2 provides information about the descriptive statistics. Despite random assignment into the treatment groups, it turns out that the subjects in the pair treatment

[5] Take, for example, Table 1 in Ammermüller and Pischke (2006).

**Table 2**
Descriptive statistics.

| | Single treatment | Pair treatment |
|---|---|---|
| Number of subjects | 49 | 96 |
| Test 1: average number of solved puzzles ($4 \times 4$ puzzles) | 3.408 (St.Dev.: 2.806) | 4.479 (St.Dev.: 2.546) |
| Test 2: average number of solved puzzles ($5 \times 5$ puzzles) | 2429 (St.Dev.: 1.860) | 3.375 (St.Dev.: 1.991) |
| Math performance in the last year (variable name: *Mathmark*; from 1 (very bad) to 6 (very good)) | 4.44 (St.Dev.: 0.79) | 4.43 (St.Dev.: 0.78) |
| Share of students with interest in logical puzzles (variable name: *Sudoku*) | 65% | 80% |

express significantly[6] more often an interest in logical puzzles (Variable *sudoku*).

I present OLS estimations in the subsequent econometric analysis. The performance is measured by counting the number of solved puzzles. Therefore, Poisson regressions or negative binomial regressions seem to be plausible alternatives, depending on the dispersion. The analysis has been conducted with all three methods but the alternatives do not deliver qualitatively different results.

### 4.1. Motivation effects (Test 1)

A first estimation model shows a difference in the first test between the treatment groups (Table 3). I estimate output on a treatment dummy for the pair treatment (Model 1 in table). The results show a significant treatment effect even after controlling for heterogeneity between the groups with respect to sex (0 = female, 1 = male), the school grade (*Grade*: 2, 3, or 4) and a school dummy (all in Model 2).[7] Further controls for marks in math and general preferences for logical puzzles support this evidence (Model 3). Hence, prediction 1 is confirmed

There is some evidence about the source of motivation. I estimate the performance in the first test with exogenous characteristics of the partner. Any significant characteristic is a source of motivation. In this experiment a partner with a better math score induces lower motivation (Table 4).

### 4.2. Peer effects (Test 2)

The performance differences in the second test are also significant. Model 1 in Table 5 provides the overall peer effect. To some extent the results in the first test drive this significant effect (Model 2). The resulting treatment effect is hardly significant at the 10% level. It is insignificant once differences in group composition are accounted for (Model 3). Model 4 finds a significant treatment effect once a non-linear impact of the first test (*first test*$^2$) is taken into account[8]. This latter result holds even if the math mark and the interest in logical puzzles is taken into account (Model 5).

These results confirm prediction 2. However, they do not provide definite evidence that the actual interaction between the subjects in the pair treatment matters once motivational differences are taken into account.

Differences between the schools are also interesting. The students in one school perform worse in the first test than their colleagues in the other one, i.e. the task is initially more difficult for them[9]. Estimating the first model in Table 5 separately for each school documents a significant treatment effect in the school with the lower initial performance ($p$-value = .033) and an insignificant one in the better school. No differences across the other subgroups (i.e. gender and grade) are observable.

### 4.3. Optimal peer group composition

The final focus is on the interaction of subjects in a group. At first I test if performance increases in the performance of the partner. The following analysis uses data from the 96 subjects in the peer treatment group. I took several performance measures as independent variables in OLS regressions: The score of the partner (*partnerscore*), the partner's marks in math (*mathpeer*) and her interest in similar logical puzzles (*sudokupeer*). Apart from the test score, these variables have been taken from the questionnaires.

The score of the partner in the first test does not show any significant effect (Model 1 in Table 6), even as a non-linear measure.[10] The partner's interest in logical puzzle has a significant effect (Model 2) but it declines once it is controlled for the math mark ($p$-value: .100). The partner's performance in math is always an insignificant contributor to the performance of a subject.

The final questionnaire included questions about the quality of cooperation, how subjects liked their partner, if they were in the same class, etc. None of these variables has a significant impact. Overall, the results do not allow any definite conclusion about the dominance of any particular group composition policy. Therefore, the empirical support for prediction 3 is very weak. Further research with a larger subject pool is required.

## 5. Summary and discussion

The experiment introduced in this paper confirmed the benefits from cooperation in a learning context with an

[6] On the 5% level, according to the $t$-test.

[7] No incident of cheating was observed during the test, which could have driven the result. The correlation between the first test score of a subject in the pair treatment and the test score of his prospective partner is .01, which underlines this observation.

[8] The $p$-value for the treatment effect is .093 if model 4 is estimated with both *first test* and *first test*$^2$.

[9] The $p$-value in a two-sample $t$-test is .0438.

[10] Even if it was significant, it would be affected by Manski's (1993) reflection problem.

**Table 3**
Estimation of differences between the treatment groups in the first test.

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| OLS: $N = 145$, dependent variable: *first test*; coefficients (robust standard error) | | | |
| *Treatment* | 1.071 (.476)* | 1.203 (.477)** | 1.005 (.461)** |
| *Grade* | | .014 (.266) | .041 (.251) |
| *Sex* | | .786 (.445)* | .895 (.440)** |
| *School* | | .918 (.428)** | .669 (.424) |
| *Mathmark* | | | .639 (.272)** |
| *Sudoku* | | | 1.527 (.519)*** |
| *Constant* | 3.408 (.399)*** | 2.576 (.903)*** | −1.287 (1.399) |
| $R^2$ | .0361 | .0842 | .1869 |

* Significance level 0.1.
** Significance level .05.
*** Significance level .01.

**Table 4**
Estimation of cognitive peer effects in the first test.

| | | | |
|---|---|---|---|
| OLS: $N = 96$, dependent variable: *first test*; coefficients (robust standard error) | | | |
| *Sudokupeer* | −1.115 (.559)** | | −.849 (.568) |
| *Mathpeer* | | −.071 (.028)** | −.061 (.028)** |
| *Grade* | .233 (.311) | .219 (.314) | .208 (.313) |
| *Sex* | 1.065 (.512)** | 1.304 (.530)** | .1.229 (.525)** |
| *School* | 1.228 (.510)** | 1.048 (.499)** | 1.161 (.508)** |
| *Constant* | 3.841 (.985)*** | 6.120 (1.540)*** | 6.384 (1.548)*** |
| $R^2$ | .1248 | .1412 | .1577 |

* Significance level 0.1.
** Significance level .05.
*** Significance level .01.

**Table 5**
Estimation of differences between the treatment groups in the second test.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| OLS: $N = 145$, dependent variable: *second test*; coefficients (robust standard error) | | | | | |
| *Treatment* | .946*** (.334) | .441* (.263) | .427 (.277) | .519* (.278) | .473* (.282) |
| *First test* | | .472*** (.051) | .468*** (.052) | | |
| *First test*$^2$ | | | | .055*** (.006) | .051 (.006) |
| *Grade* | | | .155 (.159) | .153 (.154) | .154 (.156) |
| *Sex* | | | −.156 (.266) | −.212 (.269) | −.112 (.283) |
| *School* | | | .266 (.270) | .339 (.261) | .276 (.262) |
| *Mathmark* | | | | | .016 (.019) |
| *Sudoku* | | | | | .651 (.330)* |
| *Constant* | 2.429*** (.265) | .820*** (.222) | .369 (.545) | .892 (.542) | −.182 (1.011) |
| $R^2$ | .0508 | .4384 | .4487 | .4558 | .4783 |

* Significance level 0.1.
** Significance level .05.
*** Significance level .01.

**Table 6**
Estimation of cognitive peer effects in the second test.

| | | | |
|---|---|---|---|
| OLS: $N = 96$, dependent variable: *second test*; coefficients (robust standard error) | | | |
| *Partnerscore* | .085 (.062) | | |
| *Sudokupeer* | | .845 (.445)* | .775 (.466)* |
| *Mathpeer* | | | .019 (.020) |
| *First test* | .477 (.070)*** | .494 (.071)*** | .505 (.074)*** |
| *Grade* | .363 (.208)* | .392 (.205)* | .397 (.204)* |
| *Sex* | −.214 (.329) | −.086 (.339) | −.150 (.347) |
| *School* | −.065 (.355) | −.097 (.357) | −.090 (.360) |
| *Constant* | −.045 (.648) | −.526 (.761) | −.806 (1.151) |
| $R^2$ | .4060 | .4223 | .4273 |

* Significance level 0.1.
** Significance level .05.
*** Significance level .01.

experiment and allowed some insight into the mechanisms of peer interaction. The experimental approach circumvents key econometric problems which greatly restrict the analysis of peer interaction with administrative or survey data. The results show that prospective cooperation has a motivational effect which induces a learning benefit. The benefit from cooperation shows no robust relationships with the characteristics of the partner. Partners with an interest in logical puzzles improve test-performance in the second test. On the other hand subjects with a high math score reduce the motivation of their partner.

The results have some clear implications. Peers matter in learning processes. In particular, they induce higher motivation. Further research is required to identify how peers induce motivation. The paper also contributes to the understanding of peer effects in general, in that peers have an impact on performance not just during interaction but also before and afterwards.

The econometric literature has investigated peer effects in schools in various ways and with different conclusions.[11] However, motivational aspects have largely been ignored, partly because administrative and survey data do not provide the relevant data. The results of this paper suggest that motivation shapes the interaction among students and, in consequence, a students' performance. Future research may show that differences in student motivation explain the heterogeneity of results regarding peer effects in schools.

The results do not imply that cognitive or other aspects of peer interaction are irrelevant. A simple variation of the experiment – and an increased subject pool – could provide more robust outcomes in that area. A brief discussion should suffice to sketch possible variations of the experiment to achieve this. The actual treatment starts 25 min after the experiment has begun. Reducing this time decreases the performance in the first test and enhances the relative importance of the treatment period for the success in the second test. Increasing the complexity of the task (e.g. by changing the puzzles in the first test to a $5 \times 5$ matrix) can have a similar effect because it takes longer to "digest" the logic behind the puzzle. The reported differences between the schools support this consideration. Variations in the timing may also provide some evidence about the optimal composition of learning groups.

The results suggest some recommendations for the organization of education although these have to be confirmed with more representative data. There is some discussion about optimal group composition and selection in schools. The selection criteria are usually based just on characteristics of a student (e.g. gender, test-performance or socio-economic background) but not on expectations about student's behavior. There are good reasons to do so, as these characteristics are easy to measure and difficult to manipulate. In contrast, Lazear's (2001) theoretical approach to selection is based on expected interruption in schools. The results of our experiment suggest that behavioral aspects play a key role for the optimal peer group composition. Non-experimental psychological research provides similar evidence (Ryan, 2001). Future research should indicate how observable characteristics of a peer group relate to the expected behavior of a student. This research in turn may lead to more sensitive and differentiated criteria for efficient selection in schools.

The paper provides another applied insight, too. Moore (2002) argues that one reason for Finland's exceptional performance in the PISA tests is her system of full-day schooling, where students spend most of the day at the school and do homework and exam preparations there. As a response to their own relatively poor performance in PISA, the governments in Austria and Germany provide strong support for similar full-day schools. The results from this paper suggest that this is a sensible policy if it induces higher cooperation between the students.

## Appendix A. Introductory letter (Translated from German)

The participants in the different treatment groups received different introductory letters. The appendix merges both letters. The participants did not know anything about the treatment in the other group. The letter is translated from German.

Dear participant,

You take part in an research project about the learning of logical puzzles. The experiment will go on for about one hour. Thank you very much for your participation.

We will test you twice in the experiment, at the beginning and at the end. In both tests we investigate, how many puzzles you can solve within a specific time.

### *A.1. Instruction for single treatment group*

> Please put your solution for each puzzle in the solution sheet only.

Between both tests you have 20 min to learn more about the logic of the puzzle. You get a sheet with more puzzles to do so.

> Please work alone during the entire experiment.

### *A.2. Instruction for pair treatment group*

> Please solve these tests alone.
> Please put your solution for each puzzle in the solution sheet only.

Between both tests you have 20 min to learn more about the logic of the puzzle. You get a sheet with more puzzles to do so. In this period you can cooperate with another participant. The seat number of your partner is on the sheet with your own number.

[11] Take for example, Table 1 in Ammermüller and Pischke (2006).

*A.3. Instruction for both groups*

We ask you to answer two questionnaires, one at the beginning and one at the end of the experiment. We ensure the anonymity of the results. You will get 20 Francs once you hand in the last questionnaire at the end.

You will find the rules for the puzzle on the next sheet. Please read them carefully. We cannot answer any questions in order to keep the conditions identical for all participants. Please do not worry if you do not understand the rules immediately. It is neither uncommon nor embarrassing.

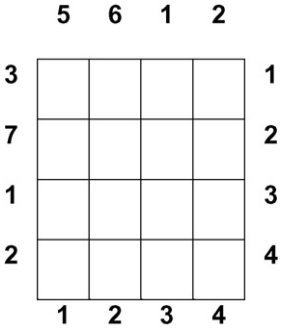
Yours sincerely.

## Appendix B. Rules (Translated from German)

Your task in the test: Solve as many puzzles as possible by marking the correct boxes in the puzzle.

The composition of a puzzle:

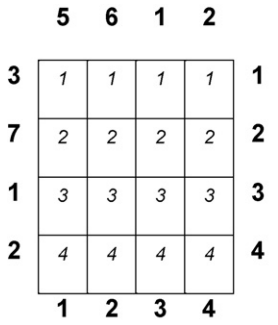
We can explain it with the help of the following example:




The puzzle is surrounded by numbers. These are very important. The numbers at right boundary and at lower boundary assign values to the different boxes.

Please bear in mind:
Each box has two different values!

1. At the right boundary you find the values for each box in the respective row.



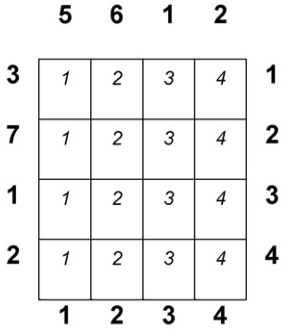

All boxes in row one have value 1.

All boxes in row two have value 2.

All boxes in row three have value 3.

All boxes in row four have value 4.

2. At the lower boundary you find the values for each box in the respective column.



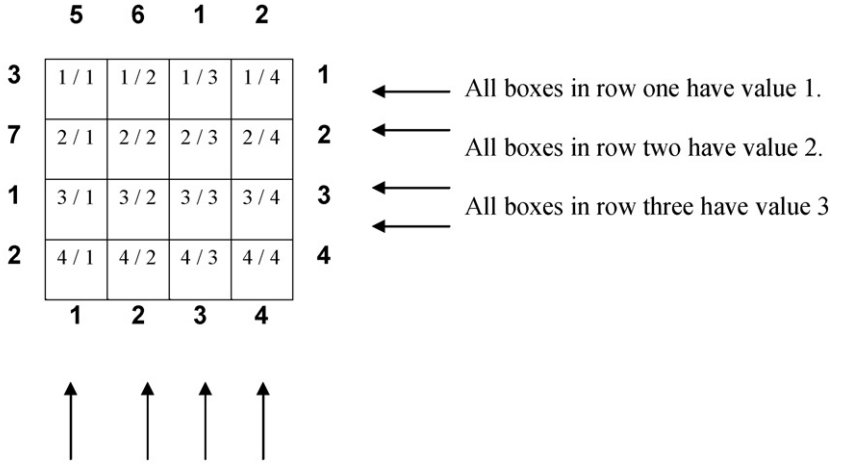

All boxes in the first column have value 1

All boxes in the second column have value 2

All boxes in the third column have value 3

All boxes in the fourth column have value 4

Here you find the values for each box again. (row value/column value).



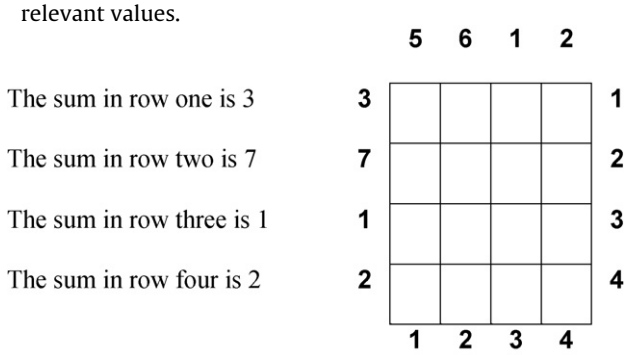

All boxes in the first column have value 1.

All boxes in the second column have value 2.

All boxes in the third column have value 3

3. The numbers at the *left* boundary show the sum of the values of all marked boxes in the respective *row*.

In this case, the values at the *lower* boundary are the relevant values.




4. The numbers at the *top* boundary show the sum of the values of all marked boxes in the respective *column*.

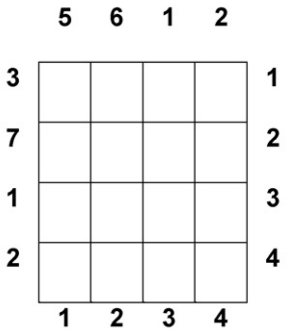
In this case, the values at the *right* boundary are the relevant values.

The sum in column one is 5

The sum in column two is 5

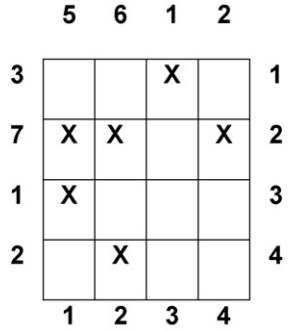
The sum in column three is 1

The sum in column four is 2



Now you have to mark the boxes, ensuring that both the values on the top and at the left boundary are correct.

Our example has the following solution




Is the solution correct?

In *row one*, the sum of all marked boxes has to be 3. Only the third box is marked, its value is 3 (see lower boundary). Hence, the solution in this row is correct.

In *column one*, the sum of all marked boxes has to be 5. The second and the third box are marked: 2+3 = 5 (see right boundary). The solution in this column is correct.

Indeed, the marked boxes ensure a correct solution for the values on the top and the left.

Row one: 3 = 3

Row two: 7 = 1+2+4

Row three: 1 = 1

Row four: 2 = 2

Column one: 5 = 2+3

Column two: 6 = 2+4

Column three: 1 = 1

Column four: 2 = 2

For the sums in each **row**, take the marked boxes in this row with the corresponding values at the **bottom**.

For the sums in each **column**, take the marked boxes in this column with the corresponding values at the **right**.

If you do not understand the solution at the moment, it is no problem. Go through the solution once. Or maybe the following test can help you. Good luck.

## Appendix C. The Questionnaires (translations from German)

Questions in the first questionnaire:

- Sex (Boxes: male and female)
- How old are you?
- What are your preferred subjects in school? Note up to three
- What are your three best subjects in school?
- What was your math mark in last year's final certificate?
- What was your overall average mark in last year's final certificate?
- Do you like to go to school? (Boxes: yes and no)
- Do you often prepare for exams with other students? (Boxes: yes and no)
- Can you explain difficult topics successfully to others (e.g. solution for difficult mathematical tasks)? (Boxes: yes and no)
- Do you get along with most of your fellow students? (Boxes: yes and no)
- Do you like to solve logical puzzles, e.g. Sudoku? (Boxes: yes and no)
- Are you member in a club? (Boxes: yes and no)
- If yes, what type of club is it?

Questions ahead of each test

- Do you understand the rules of the puzzle? (Boxes: yes and no)
- What do you expect? Will you solve the puzzles easily or with difficulty? (5 Boxes, very easy to very difficult)
- Will your performance in this test be above or below the average performance of the other participants? (5 Boxes, much below average to much above average)

Questions in the final questionnaire:

- Compared with your expectations, did you find the puzzles easier or more difficult to solve. (5 Boxes, much easier to much more difficult)
- Do you expect to perform better than the average? (5 Boxes, much worse to much better)
- How was the quality of cooperation with the partner? (5 Boxes, very good to very bad)
- Was your partner helpful during the preparation? (5 Boxes, very helpful to not helpful at all)
- Did you know your partner before the experiment? (Boxes: yes and no)
- Are you in the same grade? (Boxes: yes and no)
- Are you in the same class? (Boxes: yes and no)
- Does your partner live in your vicinity? (Boxes: yes and no)
- Do you prepare occasionally with him for classes or exams? (Boxes: yes and no)
- How much do you like your partner? (5 Boxes, very much to not at all)
- Did you like the experiment? (Boxes: yes and no)
- Are you interested in the results of the experiment? (Boxes: yes and no)

- Do you want to participate in similar experiments in the future? (Boxes: yes and no)

## References


Ammermüller, A., & Pischke, J.-S. (2006). *Peer effects in European primary schools: Evidence from PIRLS*. Mannheim: ZEW Discussion Paper.

Arnott, R., & Rowse, J. (1987). Peer group effects and educational attainment. *Journal of Public Economics*, *32*, 287–305.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, *9*(3), 245–250.

De Fraja, G., & Landeras, P. (2006). Could do better: The effectiveness of incentives and competition in schools. *Journal of Public Economics*, *90*, 189–213.

Deci, E. L., & Ryan, R. M. (1985). *Intrinisc motivation and self-determination in human behavior*. New York: Plenum Press.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, *24*(1), 39–57.

Feinberg, J. M., & Aiello, J. R. (2006). Social facilitation: A test of competing theories. *Journal of Applied Social Psychology*, *36*, 5.

Fleming, V. M., & Alexander, J. M. (2001). The benefits of peer collaboration: A replication with a delayed posttest. *Contemporary Educational Psychology*, *26*(4), 588–601.

Foster, G. (2006). It's not your peers and it's not your friends. Some progress towards understanding the educational peer effect mechanism. *Journal of Public Economics*, *90*, 1455–1475.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*(3), 791–810.

Gould, E., Lavy, V., & Paserman, D. (2005). *Does immigration affect the longterm educational outcomes of natives? Quasi-experimental evidence*. Hebrew University.

Hanushek, E. A., Kain, J. F, Markman, J. M., & Rivkin, R. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*(5), 527–544.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, *90*(5), 1209–1238.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, *116*(3), 777–803.

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, *36*(3), 465–475.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*(3), 449.

Lui, K. M., Chan, K. C. C., & Nosek, J. T. (2008). The effect of pairs in program design tasks. *IEEE Transactions of Software Engineering*, 197–211.

Manion, V., & Alexander, J. M. (1997). The benefits of peer collaboration on strategy use, metacognitive causal attribution, and recall. *Journal of Experimental Child Psychology*, *67*(2), 268–289.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*(3), 531–542.

Mas, A., & Moretti, E. (2006). *Peers at work*. NBER working paper.

McEwan, P. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, *22*(2), 131–141.

Moore, A. (2002). Learning from Pisa: Reasons and remedies for student under-performance in reading, Maths and Science. *EMBO Reports*, *3*(4), 296.

Platania, J., & Moran, G. P. (2001). Social facilitation as a function of the mere presence of others. *The Journal of Social Psychology*, *141*(2), 190–197.

Quarter, J., & Marcus, A. (1971). Drive level and the audience effect: A test of Zajonc's theory. *The Journal of Social Psychology*, *83*(1), 99–105.

Rothschild, M., & White, L. J. (1995). The analytics of the pricing of higher education and other services in which the customers are inputs. *Journal of Political Economy*, *103*(3), 573–586.

Ryan, A. M. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development*, *72*(4), 1135–1150.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, *116*(2), 681–704.

Stinebrickner, R., & Stinebrickner, T. R. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, *90*, 1435–1454.

Zajonc, R. B. (1965). Social facilitation: A solution is suggested for an old unresolved social psychological problem. *Science*, *149*(3681), 269–274.

Zimmerman, D. J. (2003). Peer effects in higher education: Evidence from a natural experiment. *Review of Economics and Statistics*, *85*(1), 9–23.