Weber, Andrea Maria

Working Paper

An evaluation of single and mixed gender computer science classes

Diskussionsbeitrag, No. 369

Provided in Cooperation with:
School of Economics and Management, University of Hannover

Suggested Citation: Weber, Andrea Maria (2007) : An evaluation of single and mixed gender computer science classes, Diskussionsbeitrag, No. 369, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät, Hannover

This Version is available at:
https://hdl.handle.net/10419/27179

An Evaluation of Single and Mixed Gender Computer Science Classes

Andrea M. Weber

***Darmstadt University of Technology**
and Leibniz University of Hannover

July 2007

Andrea M. Weber
Darmstadt University of Technology
Residenzschloss S313/136, Marktplatz 15
64283 Darmstadt
Germany

Phone: *++49 – 6151 –16-3292*
Fax: *++49 – 6151 – 16-3897*
E-Mail: weber@vwl.tu-darmstadt.de

Abstract: Discussions on the benefits of single-gender education on girls' science outcomes are popular in the German education literature. However, most empirical evidence tends to be qualitative work and the causal effects of single-gender education are hardly identified using appropriate statistical methods. This paper provides insights from a recent single-gender-education school project conducted in computer science classes at a German lower secondary school. About 80 students participated in this intervention study repeatedly answering specifically designed questionnaires and tests. The project fails to identify positive effects from single-gender education but the interpretation is impeded by several confounding factors. When directly asked, most students prefer to be educated in mixed-gender groups, while the participating teachers judge their teaching experience with the project groups in favour of single-gender education.

JEL classification: I21, J16

Keywords: gender, education, identification, coeducation, segregation, experiments

Acknowledgement: Part of this research was supported by the Anglo-German Foundation within the project 'The Economics and Politics of Employment, Migration and Social Justice', which is part of the Foundation's research initiative "Creating Sustainable Growth in Europe". I am obliged to the school principal, teachers, pupils, parents, and those responsible at the local school authority who kindly supported this project. I am grateful to Patrick Puhani for encouraging the study and for valuable comments. I also thank Horst Entorf, Olaf Hübler, Christian Pfeifer and participants of the Ph.D. Seminars in Applied Labor Economics at Darmstadt University of Technology and at the Leibniz University of Hannover for helpful comments as well as Björn Schumacher and Philip Savage for excellent research assistance. All remaining errors in this paper are my own.

Contents

1 Motivation and Stylized Facts

Coeducation was broadly introduced in West Germany in the 1950s and 1960s and in East Germany in 1945 and was taken to be an important measure for the assurance of equal educational and professional opportunities for both genders. However, educational experts soon started to doubt the universal benefits of mixed-gender education. In the 1980s, German universities realised that the overwhelming proportion of girls choosing technical study programmes had graduated from single-gender high schools. In 1989, the German feminist magazine "EMMA" even came up with the slogan 'coeducation makes girls stupid' (*Koedukation macht Mädchen dumm*)[1]. In the same year, the federal states' women's representatives (*Frauenbeauftragte der Länder*) suggested that instruction techniques should eliminate gender stereotypes in the classroom and compensate disadvantages for female students (cf. Volmerg *et al*., 1996, p.11).

Thus, debates on the advantages and disadvantages of coeducation and how to design coeducation made a return to the political agenda in the 1980s. These discussions were accompanied by evolving educational, psychological and sociological research studies on coeducation and educational performance by gender. The classic arguments used in this education literature to explain girls' reluctance to choose technical subjects are (1) little experience related to the contents of these subjects before entering, or outside of, the classroom combined with the feeling that they are unable to catch up with the boys' head start[2], (2) a self-critical assessment of their own abilities, which is typically observed for female students[3], (3) a lack of *female* role models related to the respective subjects and (4) a lack of educational techniques focusing on the needs or interests of female students.

Theoretically, there are two major ways out of this coeducational dilemma: First of all, researchers suggested going back to single-gender education in technical subjects. In fact, nowadays (beside the principle of general coeducation of the genders) the different states' education laws (*Schulgesetze*) explicitly include the possibility of (temporary) single-gender education in some subjects. The second option is to foster instruction techniques *within* co-

[1] Quoted according to Kraus (1998). A comparable phrase that spread in English-speaking countries is 'better dead than coed'.

[2] Compare for example Schuld (1997) and Heidtmann (1998) for gender differences in computer use at home. From this literature it is obvious that teenage boys more often possess a computer than girls and use computers in their leisure time.

[3] If students are asked to judge their own performance, girls generally judge their performance worse compared to boys' self-assessment (given the same state of knowledge). Generally, there is also a consensus in the education literature that boys and girls behave differently in class: Girls are more often cooperative and behave according to teachers' expectations, while boys show more competitiveness. Compare for example Funken *et al.* (1996), Volmerg *et al.* (1996) or Rustmeyer und Jubel (1996) for discussions of these phenomena.

education classes which focus on the specific needs and interests of girls. The present paper follows the first approach examining whether coeducation really provides means to improve girls' performance in typically male-dominated subjects.

Until now, research on the benefits of single-gender education mainly focused on higher level secondary schools (*Gymnasium*). However, it is a fact that the group of graduates from the lower level schools (*Hauptschule*) experiences most problems in the labour market (*e.g.* high unemployment and low labour earnings). Girls (specifically those with an immigrant background) are known to be an especially disadvantaged group among lower level secondary school students: They are often raised according to traditional gender role models and are characterized by low social mobility (cf. Brendel, 1995 and Thierack, 1995). Therefore, it is crucial to develop specific educational measures in order to improve the educational situation of these 'losers' in the education system.[4] Single-gender education might be one promising measure in order to improve the situation of poorly educated girls.

Table 1: Most popular apprenticeship choices among lower secondary school graduates

Male		Female	
Subject of apprenticeship	%	Subject of apprenticeship	%
Vehicle Construction and Maintenance	12.75	Health and Hygiene	32.20
Construction	9.11	Retail	17.71
Painters, Varnishers	7.79	Commerce	13.46
Metal Processing	7.37	Hotel and Catering Industry	8.18
Mining, Manufacturing	6.98	Clerical Employees	8.08
Plumbers, Installers	6.71	Agriculture, Farming, Forestry	4.24
Electronics	6.29	Textile, Clothing, Nutrition (without cooks)	3.03
Commerce	5.87	Cooks	2.45
Textile, Clothing, Nutrition (without cooks)	5.56	Transportation, Stocking, Cleaning	1.42
Transportation, Stocking, Cleaning	5.43	Mining, Manufacturing	1.01

Note: The figures refer to the south-western German state of *Hessen* for which the required data is available. There are missing observations on the subject of apprenticeships for 2.66 % of the male students and 4.06 % of the female students.
Source: Student-Level Data of the Statistics of General Schools of the State of *Hessen* 2005/2006 provided by the State Statistical Office. Sample of 24,383 male and 11,116 female lower secondary school graduates in vocational schools. Own calculations.

Table 1 presents evidence on the ten most popular apprenticeship professions of graduates of the lower secondary schools in one German state by gender.[5] Obviously, girls' apprenticeship choices are much more focused on specific professional fields than boys' choices. Two-thirds of the female lower secondary graduates work in the fields of health and hygiene, retail or

[4] The lack of studies examining lower secondary school students has been discussed before, for example in the papers by Brendel (1995) and Thierack (1995).
[5] The numbers refer to apprentices in the German state of *Hessen* since this is the only state for which the required information on apprenticeship choice and attained secondary degree is available from the official school statistics. Aggregated statistics for all apprentices independent of the secondary degree they hold are provided by the Federal Statistical Office and yield a similar pattern. Examination of industry codes of former lower secondary school graduates in the German-Socio Economic Panel (GSOEP) also yields similar results.

commerce, where they are trained for 'classical' female professions like nurses, saleswomen or secretaries. Male students on the other hand tend to be trained in more technical professions (*e.g.* vehicle construction) which are hardly on the list of female graduates' choices. Additionally, male graduates of the lower secondary schools spread to a larger variety of apprenticeships compared to the female graduates. All in all, the choice of professions among lower secondary school graduates seems to be rather gender-specific and corresponds to classical gender-stereotypes.

In order to improve the labour market perspectives of female lower secondary school graduates it might be a promising step to foster their interest in the neglected study fields of technical subjects. As an example, and because of the growing importance of computer related skills in today's labour markets, the present paper focuses on computer studies (cf. for example Black and Spitz-Oener, 2007 for the change in gender-related job tasks in light of the technological progress). Specifically, the research question to be answered is whether girls' interest and performance in computer education may be improved through single-gender education. Therefore, a new school project was conducted during the school years 2004/2005 and 2005/2006 in an intermediate size lower level secondary school in the German state of **B**aden-**W**ürttemberg (hence the project name 'BW-project' which is used for the remainder of this paper). The intervention refers to computer science lessons of approximately eighty students who were in fifth grade in the first year of the project.

This paper proceeds as follows: First of all it reviews empirical strategies and existing evidence in order to assess the benefits and detriments of coeducation on girls' performance in computer studies (Section 2). The literature review focuses on German studies but also provides a summary of international evidence. Furthermore, the set-up of the BW-project is presented in Section 3 together with the findings of this project. When directly asked, most students prefer to be educated in mixed-gender groups, while the participating teachers judge their teaching experience with the project groups in favour of single-gender education. Considering different performance measures, the project does not allow conclusion in favour of coeducation. However, this interpretation is impeded by several confounding factors. Section 4 summarizes and discusses the current state of knowledge. The need for further empirical evidence is emphasized.

2 Review of Empirical Designs and Literature

The research question of interest is whether pupils, and specifically girls, benefit from single-gender science education. Generally, it is possible to measure the 'benefits' by focusing on different types of performance measures, which might be categorized as 'hard' and 'soft' measures[6]: Hard performance measures directly assess educational performance based, for example, on grades, test scores or students' self-assessment, while soft measures relate to behavioural or social outcomes like self-confidence, interest or motivation.

Given the set of outcome variables of interest, the empirical researcher aims to compare these outcomes of persons participating in specific single-gender education measures (the treated group) and persons not participating (the untreated) in order to identify a 'treatment effect'. For example, the outcomes (*e.g.* test scores, motivational or behavioural indicators) of students in single-gender classes are to be compared to outcomes of students in coeducational classes. Logically, it is not possible to observe identical students under both the single-gender regime and the coeducational regime at the same time. If one simply compares students from single-gender classes to students in mixed-gender classes, it is very likely that these students differ in observed or unobserved characteristics which may in turn influence the outcome variables. In other words: The observed effects will be biased because of the selection of students to the different groups. This is the typical evaluation problem which we have in the social sciences (compare for example the formalization in Heckman and Smith, 1995). Thus, in order to evaluate the impact of single-gender education, the crux is to find an adequate control group which did not receive single-gender treatment and does not distinguish from the treatment group in background characteristics determining educational success.

2.1 General Evaluation Strategies

How can the present evaluation problem thus be solved? Generally, there are three methodological bags of tricks differing in the way in which an adequate control group is created. The three strategies, which will be discussed for the present purpose, are: (1) *social experiments*, (2) *natural experiments,* and (3) *non-experimental approaches:*

Social experiments imply an intervention study where students are randomly assigned to coeducational and single-gender classes. Due to this procedure, it is expected that the

[6] Compare for example Hoffmann *et al.* (1997), chapter 6 for these different measures and their assessment in the German literature.

groups of treated students (single-gender education) and untreated students (mixed-gender education) do not differ in the distribution of their observed and unobserved characteristics.

A general critique with respect to the required randomisation refers to the fact that this procedure might change the pool of persons who are willing to participate in the evaluation study and the behaviour of participants. This effect is known as the *randomization bias* of social experiments. The major practical problem of randomisation is, however, that random assignment of students to classes may be disapproved by school principals or other participating persons: Sometimes random treatment (as compared to free choice of received treatment) is considered to be politically incorrect if there are concerns that the untreated students are arbitrarily deprived of beneficial measures. Furthermore, randomisation of students into groups (*e.g.* in the technical subjects considered) may be difficult for administrative reasons. Specifically, students are usually instructed in the same class context for all subjects in Germany which complicates the random assignment to different groups in a specific subject. Additionally, schools face limited resources in relation to, for example, teachers' work hours and the available science or computer classroom capacities.

As an alternative identification strategy, one stream of the existing literature considers ***natural experiments***. A natural experiment consists of an exogenous policy change affecting only sub-groups of the population where assignment to the affected groups is exogenous to the outcome. Few previous studies drawing on natural experiments consider the effects that arise from a transition of schools from a coeducational to a single-gender regime or *vice versa*, as will be summarized in section 2.3. Besides simply looking at changes over time, natural experiments can make use of pre-post designs, where performance changes of students experiencing a school's transformation (treatment group) are compared to the outcome patterns of students in similar schools or classes which did not experience such a regime transformation (control group). Thus, the chosen approach corresponds to a difference-in-difference approach where the performance change over time (first difference) is compared between a treatment and a comparison group (second difference). The lack of evidence related to natural experiments in the existing literature is symptomatic of the absence of such present regime changes.

Since natural experiments are rare and social experiments are hardly feasible, practical alternatives consist in ***non-experimental methods.*** Such studies draw on a comparison of students educated in single-gender and mixed-gender groups, explicitly taking initial selection into account via appropriate statistical methods. For example, matching techniques might be used in order to compare the outcomes of persons in treatment and control groups which are

similar in (observed) individual background variables.[7] Regression adjustment methods regress the outcome of interest on a treatment dummy and individual background variables. Specifically, one might think of directly modelling initial selection into the treatment group *via* appropriate econometric techniques (*e.g.* using switching regressions, cf. section 2.3).

2.2 General Implementation Issues

The detailed design of the intervention project needs to address different potential sources of biases. Such biases may occur independent of the underlying research strategy (social experiment, natural experiment or non-experimental evaluation). To start with, problems may emerge due to ***treatment substitution***, *i.e.* if the persons in the control group receive a substitutive treatment similar to the treatment under consideration.[8] A substitution bias would occur if, for example, students in single-gender (treatment) groups were taught by standard educational methods while students in mixed-gender (control) groups were taught by instruction methods focusing on girls' interests. In this case, the specific education methods are a substitute for the single-gender treatment and it is not possible to identify the treatment effect by comparing both groups. Generally***, multiple treatments*** may make it impossible to disentangle the effects of these different treatment measures. In order to avoid a substitution or multiple treatment bias, ideally all students have to be instructed in the same (or at least similar) way by the same (similar) teacher under the same (similar) circumstances.

The determination of treatment effects might also be complicated if the participating persons know that they are subjects of an evaluation study. Specifically, if students know that their results will be evaluated and related to their gender, this is likely to change their behaviour where the respective changes might differ by gender. As a consequence, (again) observed effects might be biased. This issue is referred to as an ***observation bias***. If students adjust their behaviour under treatment because they think this is what their *teachers* expect, this is called a ***Pygmalion effect*** in the education literature (cf. for example Ziegler *et al.*, 1998 for the discussion of the importance of this effect). In order to prevent these sources of biases, one possibility is to conceal the subject of observation and expected outcomes from the project participants.

A general problem of intervention studies considering a longer time span is that there might be significant sample attrition. Attrition will generate ***attrition biases*** in estimating the desired effects if the students dropping out of the sample systematically differ from the re-

[7] See, for example, Heckman *et al.* (1998) for a theoretical discussion of matching as an econometric evaluation estimator.
[8] See Heckman and Smith (1995) for a discussion of the substitution bias and other biases in social experiments.

maining students. Attrition may occur due, for example, to non-response to the research questions or if students who have initially been assigned to one of the compared groups are not satisfied with this educational situation and decide to leave the group. In the most extreme case students would leave the school or have to leave the school, respectively. Generally, it might be that school dropouts change the composition of the groups considered of interest.

2.3 Literature Review with a Focus on German Studies

A great many existing studies conducted by researchers from different social science disciplines around the world try to empirically evaluate the effects of single-gender education. Mael *et al.* (2005) provide a comprehensive literature and methods overview for industrialized countries. However, this survey article demonstrates that an overwhelming proportion of the present literature simply compares outcomes of students in single-gender and coeducational classes without taking selection to different educational regimes into account. In many cases the compared groups are even located at different schools. Thus, such simple correlation studies are likely to suffer from severe selection biases (cf. section 2.1) and are not adequate for identifying a causal effect of single-gender education. One consequence of lacking identification strategies is that previous studies provide mixed evidence with some equivocal support of positive impacts of single-gender education related to school performance.

Since the survey by Mael *et al. (*2005) only considers evidence from studies in the English language and since the existing studies for Germany are mostly German-language research reports, hardly any evidence on Germany is summarized. Table A in the 'Appendix to the Literature Review' provides an overview of recent research on single-gender computer education in Germany.[9] Due to the limited number of such studies, related fields of education (*i.e.* mathematics and science) are included in this review. Studies are categorized according to the way in which they solve the evaluation problem. The four categories considered are (1) correlation studies which do not solve the evaluation problem, (2) non-experimental evaluation studies drawing on matching or similar techniques, (3) natural experiments implying exogenous policy changes and (4) social experiments where students are randomly assigned to treatment and control groups.

The overview given in Table A does not claim to be exhaustive: Especially, a large variety of articles (including results only published in newspaper articles) related to the 'correla-

[9] While the focus of the following literature review is on the effects of coeducation on girls' outcomes, fewer studies consider the impacts on male students. Holz-Ebeling *et al.* (2000) is an exceptional study focusing on boys' outcomes. One conclusion of the study is that boys' educational and social outcomes are hardly affected by single-gender education. However, since the paper examines boys in coeducational and single-gender educational schools without appropriately accounting for the selectivity into different school types, the findings are again to be interpreted with reservation.

tion study' category exists. However, because of the inherent problematic identification technique these studies are not of primary interest to the present paper. Additionally, most of the summarized papers do not solely concentrate on the specific topic related to the evaluation of coeducation *vs.* single-gender education but also cover more general topics (*e.g.* discussions of curricula or overviews of the history of coeducation). Since these special topics are not the focus of this paper, the given summary is restricted to relevant evaluation results. Moreover, only *recent* evidence, starting in the mid 1990s, is considered. One older study is presented when discussing the evidence from natural experiments since this is the newest available study for this category.[10] Finally, further recent German language publications refer to intervention studies in Austria or Switzerland (*e.g.* Faulstich-Wieland, 2004a provides recent evidence). Since the interest of this summary is related to studies for Germany, I do not summarize these papers.

The studies summarized in category (1) in Table A are mostly qualitative reports including simple descriptive statistics. Funken *et al.* (1996), Meyer (1996) and Volmerg *et al.* (1996) explicitly focus on computer studies while Nyssen (1996) considers the related context of science classes. The size of the studies in terms of underlying samples varies considerably, ranging from observations for only 29 students in Meyer (1996) to 1,128 students in Funken *et al.* (1996). While Funken *et al.* (1996) covers girls in single-gender and mixed-gender schools, the other studies compare students in mixed and single-gender classes within coeducational schools. The report by Nyssen (1996) refers to comprehensive schools, while the remaining studies consider students in the highest secondary school track (*Gymnasium*). Evidence is based on outcome measures such as the students' self-assessed interest and motivation. The measured effects are mixed with a tendency to interpret in favour of single-gender education.

Besides the selection bias, these qualitative reports are also likely to suffer from substitution biases: Different treatment measures are applied to different groups (cf. Nyssen, 1996; Volmerg *et al.*, 1996) so that a simple comparison of outcomes is hardly appropriate for identifying an effect of interest. One issue related to this critique is that it is not assured that

[10] Concerning older studies, especially in the 1990s, different publicly financed measures were conducted in order to increase girls' participation in technical subjects and science (including computer science). Such and further measures (supported and funded by the federal commission Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, BLK) primarily focusing on *tertiary* education are summarized in BLK (2002). Additionally, compare Kessels (2002), Hoffmann *et al.* (1997) or Volmerg *et al.* (1996) for brief summaries of older studies for Germany. For the most part, older studies simply compare girls in single-gender and mixed schools and exhibit the same problems as discussed for the studies summarized in Table A in the 'Appendix to the Literature Review'. A further stream of literature, which is not the focus of this review, considers gender related subject of degree choice in higher secondary schools (see Roeder and Gruehn, 1997 and Heinrichs and Schulz, 1989 for examples of such studies).

the different groups are taught by similar teachers. Additionally, it is reported in at least one of the four studies (Volmerg *et al.*, 1996) that girls anticipated the expected results so that the measured effects are probably driven by an observation bias or Pygmalion effect (cf. section 2.2).

Concerning non-experimental evaluation studies ('type 2'), the overview article of Mael *et al.* (2005) documents that only few of the existing international, English-language studies control for relevant characteristics driving the selection. Specifically, it seems that the more appropriately selection is controlled for (*i.e.* the more control variables are added) the higher the reduction in the estimated effect of single-gender education.[11] Billger (2006) is an exceptional study using *econometric* techniques in order to take selection to different educational regimes into account. Since it is a recently published paper, it is not included in the 2005 review of Mael *et al.* The paper examines effects of single-gender school attendance on education and individual labor market outcomes in the U.S. Results from regression analyses (including switching regressions for starting salaries) controlling for a variety of students' background characteristics show modest positive effects of single-gender education. However, most regression results are probably biased because of selection in unobserved variables and endogeneity of some regressors. The switching regression results may suffer from the lack of exclusion restrictions in the regime equation, *i.e.* it is not discussed whether variables are included in the regime equation that have no direct potential effect on starting salaries. Consequently, it is not clear whether the study really identifies the causal effect of single-gender education or if the results are still biased due to selection.

The non-experimental evaluation literature ('type 2') for Germany is summarized in the second panel of Table A. Rost and Pruisken (2000) discuss the selection problem and aim to solve it by comparing similar students. For this purpose they compare samples of fifth and sixth graders (649 students) in single-gender and mixed-gender higher level secondary schools where all three schools considered are similar in that they are run by the Catholic Church. The study finds no significant effect of the organizational class type on different psychological and social outcome variables (including students' self-assessment in mathematics and biology). The contribution of Rost and Pruisken (2000) is that they address the selection problem and possibly even identify a (causal) effect of single-gender education for the relatively limited group of Catholic private school students. However (as is mentioned in this very study), this effect may probably not be generalized for the whole of secondary schools.

[11] The same sceptical conclusion is for example drawn in the German papers by Rost and Pruisken (2000), Faulstich-Wieland (1999) and Baumert (1992) which briefly review the international (and German) evidence on single-gender education.

Furthermore, the authors do not present evidence on the distribution of student background variables among the compared schools. It is not clear whether they really solve the selection problem even for the very specific group of students.

A series of recent publications refers to an intervention study in the German state of *Schleswig-Holstein* in the 1990s (Häußler and Hoffmann, 2002, 1998, 1990 and Hoffmann *et al.*, 1997) and focuses on science (physics) classes. A core sample includes 150 girls and 139 boys in treatment classes from six schools which are taught by six different teachers. Furthermore, 103 girls and 64 boys from two schools are in control classes and are taught by six different teachers. All students considered are seventh graders in the higher level secondary school (*Gymnasium*) in 1992/93. Students are assessed by several standardized written tests and questionnaires during the school year. Differences between groups and difference-in-differences (*i.e.* the development over time between differently treated groups) are calculated. Outcome measures refer to the students' interest, self-concept and achievement. The study may be considered to be a 'type 1' (correlation study) or a 'type 2' study (non-experimental evaluation study) since it is not clear whether the different groups are really comparable in the beginning of the intervention study. There are just brief statements (cf. Häußler and Hoffmann, 2002, page 879 and page 882) suggesting that there were no significant initial differences between groups. Yet, treatment and control groups are located at different schools and may not be comparable. Hoffmann *et al.* (1997) presents regression-adjusted results controlling for students' initial performance and 'learning environment' (as determined by class and teacher, cf. page 149). This regression analysis is likely to suffer from endogeneity of the regressors (*e.g.* self-concept which is used as a control-variable may already be affected by the treatment). The studies conclude that there is a positive impact of single-gender education especially on girls' outcomes.[12]

While correlation studies are frequent and evidence on experimental studies is already rather limited, hardly any truly experimental evidence on single-gender education effects is provided by the literature: As mentioned in section 2.2, few studies draw on natural experiments considering the effects that arise from a transition of schools from a coeducational to a single-gender regime or *vice versa*.[13] Little evidence is documented referring to the transition

[12] Hoffmann *et al.* (1997) additionally provides evidence for chemistry classes. However, due to problems related to the realization of the intervention study (small sample size, deviation of teachers from standardized curricula, cf. page 10), this evidence is difficult to interpret. Previous evidence is additionally presented in Häußler and Hoffmann (1990), observing single-gender physics classes in 1988/89. I do not report on this evidence in detail since the study is rather qualitative.

[13] Marsh *et al.* (1988) conducted such a study for Australia, examining two singe gender schools in the same neighbourhood that were restructured to be coeducational. The study does not find significant effects of the regime change on students' educational performance but positive effects on students' self concept. However, Mael

from former single-gender schools to coeducational schools after 1950 in Germany (cf. the third panel of Table A). Baumert (1992) refers to the regime switch concerning higher secondary schools between 1965 and 1975 and draws on the fact that some schools had already adapted to the new (mixed-gender) regime while other schools had not. Using data from a survey of 12,000 seventh graders including standardized German, English and mathematics outcomes, the study thus compares students in the different school types. Baumert (1992) shows that there is selection into the differently organized school types in regions where both types coexist so that students (or actually their parents) may choose to attend single or mixed-gender schools: Generally, more able students seem to prefer single-gender schools. The further study aims to take this selection into account by analysis of variance controlling for the organizational type of the school. However, the study does not make use of the natural experiment through the potential before-after-comparisons or difference-in-difference estimates. All in all, within the setup of the study the author concludes that both genders perform significantly better in mathematics in single-gender schools than in coeducational schools. Similarly, girls' interest in mathematics decreases notably in seventh grade when girls are taught in mixed-gender classes. No effects are found for the English and German performance.

Concerning social experiments ('type 4'), Kessels (2002) reports on a German intervention project which stands out from the other studies because of an effort to randomly assign students to treatment and control groups.[14] Seven coeducational schools in Berlin participated in this project, where coeducational and single-gender education classes were compared in science (physics and chemistry) lessons. The core sample of the study contains 270 eighth-graders from four comprehensive schools in Berlin (87 girls and 62 boys in coeducational classes as well as 56 girls and 65 boys in single-gender classes). Each participating teacher instructs at least one single-gender and one coeducational class. Information on socio-demographic variables and outcome measures such as motivation and self-concept are assessed by standardized questionnaires and evaluated by analysis of variance. Kessels (2002) finds that there is a positive impact of single-gender education on girls' motivation and self-confidence.

One drawback related to the design of the study by Kessels (2002) is that the study gives no information to verify the assumption that the (randomly constructed) treatment and

---

et al. (2005) raise some doubts on the study's identification strategy which are mainly related to Pygmalion effects.

[14] Also, from an international perspective there is hardly any evidence based on truly experimental studies. Marsh and Rowe (1996) summarize and re-analyze evidence from studies relating to an experiment in an Australian school. However, the paper raises severe critiques concerning the implementation of the studies and the underlying 'experiment' (*e.g.* related to non-random assignment in the second year of the project).

control groups do not differ (by hazard) in observed characteristics. Related to this, there is significant sample attrition (non-respondence) which might be systematic and thus bias the results. A further problem might be that students are informed about the subject of the intervention study so that the observed *psychological* outcomes are probably biased because of Pygmalion effects. No evidence on hard performance measures is provided. It is not clear whether an increase in girls' perceived motivation related to the science class is accompanied by a real increase in their science knowledge.

To sum up, even if several studies discuss the implication of single-gender education and refer to intervention projects that were conducted in order to shed light on this topic, most work is rather qualitative and forgoes using appropriate statistical identification strategies.[15] In other words, most of the existing studies imply conclusions (mostly in favor of single-gender education) which are rather equivocal from a methodological point of view. The major problem inherent to previous studies is that they do not solve the selection problem arising from the fact that compared single-gender and coeducational groups probably consist of students with different characteristics.

Similarly, most studies suffer from the impossibility of identifying effects from multiple treatments: If these studies try to identify the effect of single-gender education by comparing students in treatment (single-gender) and control (coeducational) groups, where besides the organizational treatment both groups differ by further treatment measures (*e.g.* different teachers, different curricula and educational methods), the effect of the single-gender treatment can hardly be identified. In these cases it is strictly not possible to disentangle the effects of different measures and the presented conclusions of these studies remain speculative in nature. Additionally, most studies refer to psychological variables or social as opposed to hard performance measures. This is probably partly due to the fact that German data protection laws are rather strict, implying that school representatives are not willing to report on their students' grades.

[15] The same conclusion is drawn by Ludwig (2003) who also reviews the literature with a focus on German studies.

3 Detailed Facts and Findings from the BW-project

This section provides detailed background information on the BW-school project and its evaluation: The general framework and implementation issues of the project are described in sub-section 3.1. Sub-section 3.2 discusses the identification strategy and provides evidence on selection of students to different groups. General gender differences emerging among the students considered are addressed in sub-section 3.3, while sub-section 3.4 presents the evaluation results of different groups. Specifically, overall outcome differences of students (and especially girls taught by the same teacher) in single and mixed-gender groups are shown.

3.1 Facts and Implementation of the Project

The BW-project was conducted during the school years 2004/2005 and 2005/2006 in a lower level secondary school *(Haupt- und Werkrealschule)* in the city of Rastatt in the state of Baden-Württemberg.[16] The specific school is of intermediate size, consisting of about 370 students in grades five to ten. The intervention refers to computer science lessons of approximately eighty students who are in fifth grade (aged 11-13) in the first year of observation. These computer lessons take part once a week for one hour. Primary goals are to familiarize students with the computer in general (*i.e.* the different components of the hardware) and to teach the application of specific standard software, especially related to text processing and the Internet. While the fifth graders are taught in three separate classes in all subjects, there are six computer study groups. Random assignment of pupils to the different groups was not feasible. The school's principal opposed random assignment explaining that necessities related to the students schedule and the available computer classrooms ruled the design of the computer groups.

In the first year of the project, four of the groups were mixed groups (coeducational) and two groups were single-gendered (one all-boy group and one all-girl group). Table 2 illustrates the (gender) composition of the six groups in which fifth graders are taught in computer sciences and indicates which teacher is responsible for which group.

[16] Rastatt is located in the higher plain of the river Rhine (Rhein) between the Rhine and the Black Forrest in the direct vicinity of the French region Alsace and the German city of Karlsruhe. The population size amounted to about 47.000 individuals in 2004 (the year when the project started). The population density corresponded to the intermediate population density in the state of Baden-Württemberg. Compared to the entire German population, the population is representative concerning the gender and age structure. In 2004, the unemployment rate in Rastatt amounted to 6.3 % which was below the West German rate of 9.4 % and slightly lower than the state Baden-Württemberg's unemployment rate (6.9 %). Generally, the proportion of persons with an immigrant background (holding citizenships from foreign states) is higher in the state of Baden-Württemberg (12.1 %) than in most other German states (with an average of 8.9 %). Compare Rastatt (2005) and IW (2005) for these and more statistical details.

Table 2: Division of groups in grade 5

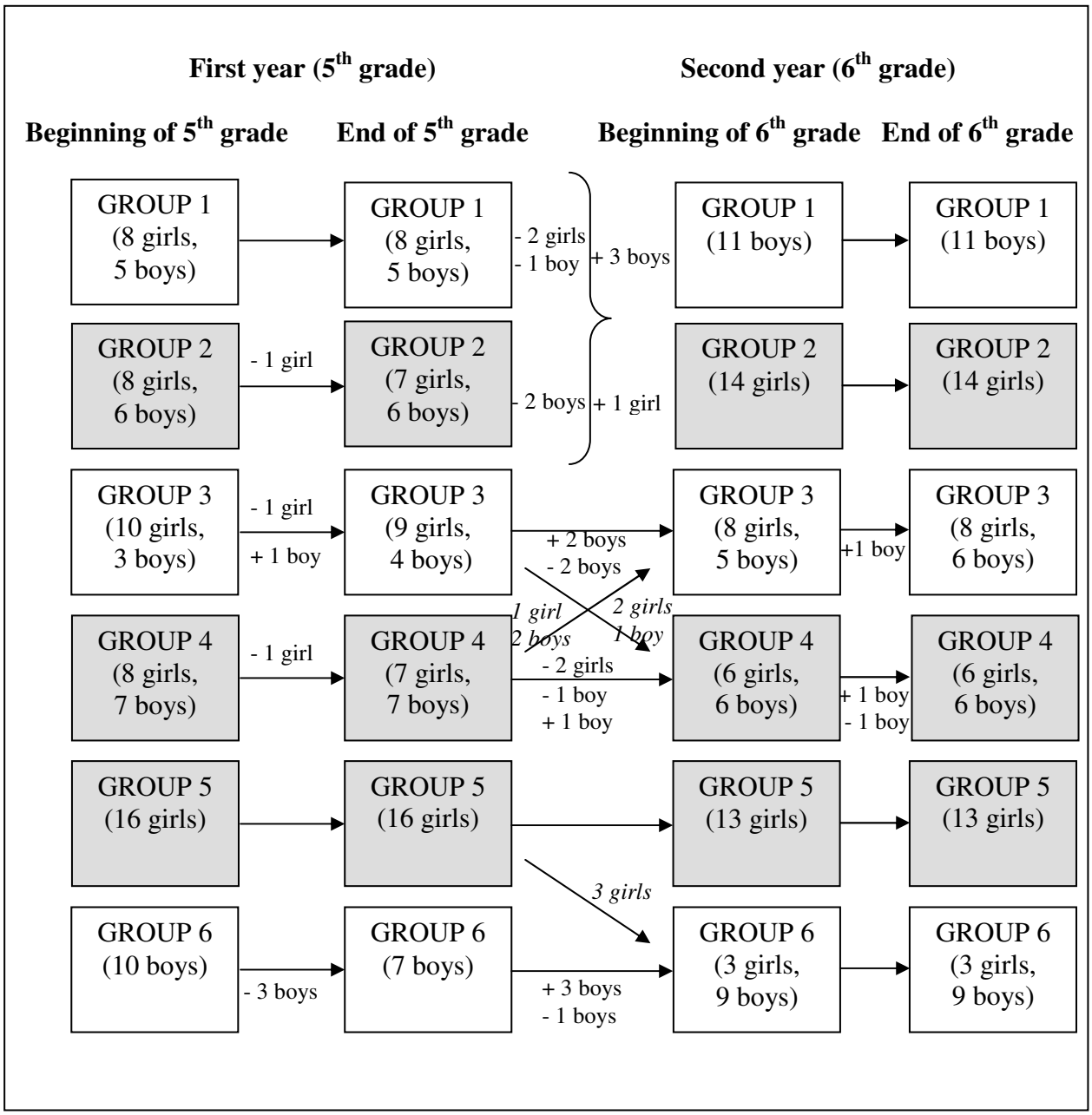
Group	Class	Teacher	Group size*	Number of girls *	Share of females
1	5a	A	13	8	62 %
2	5a	B	13 (+ 1)	6 (+ 1)	50 %
3	5b	A	13 (+-1)	9 (+ 1)	69 %
4	5b	B	14 (+ 1)	7 (+ 1)	50 %
5	5c	B	16	16	100 %
6	5c	C	7 (+ 3)	0	0 %

Note: * (+) indicates the number of students who were in this group at the beginning of the school year but dropped out of class until the end of term. This number of students is not included in the total group size. (-) indicates the number of students who were not observed in the beginning of the school year but joined class during the school year. This number of students is included in the total group size. Grey-shaded lines refer to groups taught by the same teacher (B).

Groups five and six are single-gender groups while the remaining groups are coeducational. The all-girl group (group 5) is taught by the same computer science teacher who instructs groups two and four. Group sizes changed over time due to students moving to other school districts or schools and due to students entering the school from other schools, respectively: Seven students dropped out during the first year while one additional student joined group 3 during the school year.

Figure 1 gives an overview of the group compositions by gender at different points of measurement, including the second year of observation. In addition to a switch of groups 1 and 2 from coeducational to single-gender groups, three girls from the former all-girl group 5 were now coeducated in group 6 (the former all boys' group). The latter measure was allowed by the school principal in order to equalize the group sizes and additionally complicates the comparability of groups. Additionally, the teachers allowed some students to switch between groups three and four regardless of the intervention project. Furthermore, all in all, 18 students drop out by the end of the project. Ten students, nine boys and one girl, join the groups as new students in the second year (sixth grade). Taken together, during the first and in the second year of observation three boys (one in the first year and two in the second year) joined the groups. Considering all four measurement times (mid-term and end of term of both years), only 64 students were observed throughout the two-year period.

Figure 1: Group compositions by gender



Note: + indicates the number of students who were in this group in the previous term but dropped out until the current term. - indicates the number of students who were not observed in the previous term but joined the group until the current term. Grey-shaded groups refer to students taught by the same teacher (B).

If the sample is restricted to these 64 students, the proportion of females amounts to about two thirds of the students in the resulting sample (43 out of 64). Eight girls are not observed throughout the two-year span. Thus, fluctuations are especially high among boys, *i.e.* 22 boys are not observed throughout the entire time span while 21 boys remain in the same school from the beginning of grade five until the end of grade six. Table 3 shows the numbers of continuously observed students and the numbers of students not observed throughout the entire time span by gender and group number.

Table 3: Students observed over time and dropouts

	Continuing Students		Changers	
Group	Male	Female	Male	Female
1	7	0	6	2
2	0	13	2	2
3	4	8	4	1
4	4	6	3	3
5	0	13	0	0
6	6	3	7	0
Total	21	43	22	8

Note: The group number refers to the group indicated in the second year. Continuing students are students observed throughout the two-year time span of the project. Grey-shaded lines refer to students taught by the same teacher (B).

From the above considerations it becomes clear that the time dimension of the project may hardly be used. Due to the unintended switch between groups of some students in the second year and the dropping out of other students, the group of students that might potentially be compared over time becomes too small. Specifically, only few girls are taught by the same teacher in both years: There are seven girls in group 2, six girls in group 4 and thirteen girls in group 5. In other words, too many confounding factors (including teacher differences and group compositions) in combination with the limited and decreasing sample size rule out comparisons over time. Thus, the following analyses are confined to examining the results for the first year of the project.

Student, class and teacher background characteristics are assessed through mainly self-contained questionnaires. Table 4 summarizes the different dates of measurement. At the beginning of the project (in January 2005), parents and student characteristics are assessed. Variables collected in these questionnaires are primarily related to individual age, overall school performance and immigrant background, the age and number of siblings, parental education and employment. The assessment of student characteristics is repeated once every six months until the end of the project. Additionally, the teacher questionnaires were distributed at the end of the first project year. Here, further information concerning the different groups as well as teacher characteristics and their gender views are assessed. All the questionnaires are presented in the 'Appendix of Questionnaires'. Stylized facts on the general student's characteristics which are drawn from the questionnaires are also summarized in the Appendix ('Appendix on the Students' Background').

Table 4: Dates of measurement and questionnaires

Wave	Date	Questionnaires
1	January 2005 (mid-term)	Parent and student questionnaires
2	June 2005 (end of term)	Student and teacher questionnaires
3	January 2007 (mid-term)	Student questionnaire
4	June 2007 (end of term)	Student questionnaire

Note: Questionnaires for the first year are given in the 'Appendix of Questionnaires'.

Besides the control variables, outcome variables of interest need to be collected. As indicated in section 2, there are different kinds of performance measures which may be categorized as 'hard' and 'soft' measures. Hard performance measures directly relate to the educational performance and may be assessed by grades, students' self-assessment of their own performance or by test scores. Direct assessment of performance via students' grades might be problematic because grades are likely to be relative instead of absolute performance measures, *i.e.* they indicate the students' performance *within* a given group. Students' self-assessment might be problematic as well (especially in the gender context) because girls tend to understate their own performance (compare section 1).

Thus, it is more reliable to base performance statements on standardized students' tests.[17] In the BW-project computer tests are conducted at two points in time in order to assess students' performance over time: about two weeks before mid-term and two weeks before the distribution of end of term school reports. The test questions are shown in the 'Appendix of Tests'. These tests are not a part of the students' term grade, yet the students are not aware of this which is a promising strategy in order to assure that the students put the required efforts and seriously answer the questions. The contents of the examinations were jointly developed by all of the three participating computer science teachers with the objective of not giving an advantage to one of the tested groups. All teachers graded the tests according to a linear scale. Even if the assessments are rather short, according to the teachers they yield an overall measure which is generally representative for the students' real performance.

In addition to standardized test outcomes, the BW-project draws on 'comparative' hard performance measures which relate to the students rating of their own performance relative to the other students. The underlying scale ranges from 1 ("I perform much better than my classmates.") to 5 ("Other students perform much better than me."). A value of 3 indicates that the student thinks that she performs equally to the average student. This performance measure is provided for the overall performance in all subjects taken together, for computer science, for math and for the German class. Note that the underlying scale is not an absolute performance measure but a relative self-ranking of students towards their classmates. At the

[17] For example Hoffmann *et al.* (1997) contains detailed recommendations for the construction of such tests.

same time it is likely that the indicated value is influenced by the grade the student achieves in the respective subject.

Furthermore, students are asked about the grades they expect to achieve in the end of term school report. Generally, the German school grades follow a scale from 1 ("outstanding") to 6 ("failed"). According to the school principal's information, in this school all grades are usually based on a linear scale where the distances between two grades are proportional to the performance difference (for examples related to the test scores in underlying examinations).[18]

As mentioned above, besides these different 'hard performance' measures, soft performance measures, as they are often used in sociological studies, are observed as well.[19] These measures relate to the students' gender perception and motivation. Motivation is assessed by the question whether the student likes working with the computer. Gender perceptions are deduced from the answers to the question whether the student thinks that boys or girls (or both) know more about computers relative to the other gender. In addition, students are directly asked whether they prefer to be taught in mixed or single-gender groups.

3.2 Identification Strategy and Selection Issues

Given the non-experimental set-up of the school project (*i.e.* no random assignment to treatment and control groups), the evaluation calls for statistical methods in order to take possible selection effects into account. However, popular evaluation techniques like matching or regression analysis crucially depend on the given sample size: Degrees of freedom drop if more control variables are included (*i.e.* the more appropriately selection is corrected) and it might be hard to identify significant effects if few students are observed in the treatment and control groups. Specifically, in the present case sample sizes are too small in order to employ such common methods. Too few comparable girls (*i.e.* those taught by the same teacher) are observed in mixed and single-gender classes respectively. Additionally, it is hard to impose parametrical assumptions on the data, which would be required for simple standard statistical tests (*e.g.* the t-Test).

Therefore, the analysis opts for a feasible solution and proceeds as follows: First of all, it is demonstrated that students in different groups do not - in fact - differ in their observed characteristics in the given case. Based on this insight, students in single and mixed-gender

[18] Thus, in this regime, grades can be considered to be metrically scaled, which is important when discussing the results from the project.

[19] For the interested reader, the development of detailed soft performance scales is extensively discussed in Kessels (2002, chapter 5.5) or Hoffmann *et al.* (1997, chapter 6.2), for example.

groups taught by the same teacher are compared using non-parametrical χ^2-homogeneity-tests or Kruskal-Wallis tests. The advantage of these statistics is that they are not observation-intensive but may be applied to the small samples.

Specifically, for the nominal and ordinal variables standard χ^2-independency tests are used to assess whether the characteristics are similarly distributed within groups. The conducted independency test may be interpreted as a homogeneity test with the null hypothesis that the different groups are drawn from the same population.

For metrical variables Kruskal-Wallis rank tests are feasible in order to examine whether the samples (*e.g.* the six different groups) are drawn from the same population:[20] The null hypothesis assumes that the population medians are equal (H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_6$, H_1: $\mu_i \neq \mu_j$ for at least one set of i and j, where $i \neq j$ *and* $i,j \in [1,2,3,\ldots,6]$.). The test statistic is based on ranking the combined sample of all observations. Then, the sum of the ranks is computed for each of the groups (R_1 to R_6). The test statistic is:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1),$$

where n_i *(i = 1, 2, ..., k)* represents the sample size for each of the k groups. The intuition behind this test statistic is that if the groups really have the same median, the sum of ranks for each group should not differ too much. If the null hypothesis of equal populations is true, this statistic is approximately χ^2-square distributed with *k-1* degrees of freedom. The distributional assumption is valid if each of the n_i is at least five.

Distribution of Group Background Variables

Detailed contingency tables of group characteristics and means of the background variables by group are presented in the 'Appendix of Tables of Group Characteristics'. Furthermore, Table 5 presents the results from χ^2-independency tests for the nominal and ordinal variables. Since the tests require at least five observations per cell (otherwise the distribution of the test statistic cannot be approximated by the χ^2-distribution), groups are aggregated into coeducational and mixed-gender groups. Table 5 shows that the null hypothesis (for each of the observed variables) cannot be rejected.

Table 6 reports the Kruskal-Wallis test results for observed metrical variables. According to these results, one cannot reject (at any conventional level of significance) the assumption that the samples are drawn from the same population.

[20] The test is developed in Kruskal and Wallis (1952).

Table 5: χ^2-Tests of group homogeneity (dummy variables)

Variable	$\chi^2(1)$	Probability
German-born	0.03	0.87
German-born parents	0.21	0.65
German language spoken at home	0.23	0.63
Some immigrant background	0.35	0.55
Grandparents live nearby	0.05	0.83
Higher secondary education of parents	0.01	0.92
Parental interest in school affairs	0.01	0.90

Note: The χ^2 –test compares distributions of coeducational and single-gender groups. A comparison on a less aggregated level (*i.e.* for the six study groups) is not possible due to the limited sample size.
Source: Student questionnaires. Own calculations.

Table 6: Kruskal-Wallis tests by group for observed metrical variables:

Variables	$\chi^2(5)$	Probability
Age at first measurement date	0.75	0.98
Number of siblings at home	4.17	0.52
… siblings more than 3 years younger	1.28	0.94
… siblings (less than) 3 years younger	2.25	0.81
… siblings (less than) 3 years older	1.84	0.87
… siblings more than 3 years older	0.49	0.99

Note: The test statistic is explained in section 3.2.
Source: Student questionnaires. Own calculations.

Thus, all the presented findings provide confirmation that the groups do not differ considerably in their observed characteristics. However, for the following analyses one needs to bear in mind that only three of the groups considered are taught by the same teacher. Comparisons among these groups are expected to be especially meaningful.

Teachers' statements

In addition to the above considerations, information from the teacher questionnaire is used in order to determine (1) whether the groups are comparable according to the teachers' opinion and (2) whether there are teacher differences concerning gender views. The teacher questionnaire is given in the 'Appendix of Questionnaires'.

Table 7 shows that the teachers' estimations on the proportion of students with an immigrant background in each group differ from the true proportions (as indicated by the individual students). This probably stems from the fact that teachers are only vaguely informed about the students' family background. According to information from the school principal the parents' participation in school events like parent-teacher conferences is rather low (or practically non-existent) which might make it more difficult for teachers to judge students' family background. The share of immigrant children is underestimated in groups 1, 3 and 6 and overestimated in the other half of the groups (all of which are instructed by teacher B). Generally, (as discussed in section 2.4.1) the share of immigrant children is high in each

group (at least 50 %). Table 7 includes children speaking a foreign language at home in the definition of immigrant children. Based on this definition drawn from the student-questionnaires, the shares of immigrants are between 69% and 92% depending on the group.

According to the teachers, in each group there are at most two students with serious language problems and up to one student is classified as showing behavioural or learning problems. There seems to be no clustering of disadvantaged or problem-children into one group or to one teacher. In addition, the questions related to the teachers' educational treatment of the class show that there is some homogeneity: All teachers but one (teacher C of group 6) assign homework less than once a month. Teacher A (groups 2 and 4) also instructs the religion and social study classes of the children in his computer group. It is to be expected that the homogeneity of the educational treatment is the largest between the groups taught by the same teacher B, where this teacher does not meet one of his groups in any other class.

Table 7: Teachers' information on group background

Variable / Group	1	2	3	4	5	6
Share of foreigners (born abroad or parents born abroad)	38 %	100 %	67 %	86 %	100 %	57 %
Real share of foreigners (student information)	*69 %*	*75 %*	*92 %*	*50 %*	*69 %*	*86 %*
Share (number) of children showing serious language problems	15 % (2)	0 %	0 %	14 % (2)	0 %	0 %
Share (number)of children showing learning / behavioural problems	8 % (1)	0 %	0 %	7 % (1)	6 % (1)	4 % (1)
Frequency of homework in a month*	<1	<1	<1	<1	<1	1-2
Number of other subjects in which teacher instructs these students	2	0	1	0	0	0
Kind of other subjects in which teacher instructs these students	social studies		social studies			
Teacher ID	A	B	A	B	B	C

Note: <1 'sometimes but less than once a month', 1-2 'once or twice a month, but less than once in two weeks'.
Source: Teacher questionnaires. Own calculations.

Background information on the teachers is given in Table 8. All the teachers are male (which might be important concerning their gender views). Teachers A and B are more similar in their age, work experience and instructed subjects: They are relatively young (36 and 37 years respectively), teach for three and eight years respectively and generally instruct children in all kinds of subjects, while teacher C is a 56 year old science teacher with 33 years of work experience. From this and the previous tables it becomes obvious that group six (instructed by teacher C) is hardly comparable to the other groups if one expects that differences in teacher characteristics influence the educational treatment.

Table 8: Teacher characteristics

Variable / Teacher	A	B	C
Gender	male	male	male
Age	36	37	56
How many years have you been a teacher?	3	8	33
Number of subjects taught	6	9	4
Kind of subjects taught	*all fields**	*all fields**	science

Note: **all fields* indicates that the teacher instructs science/math as well as languages and social studies.
Source: Teacher questionnaires. Own calculations.

Teachers' attitudes towards gender views and stereotypes have been inquired into through a series of questions which are repeated in Table 9. Teachers B and C, who have gained experience with single-gender classes in the first project year, are consistent with the view that it is reasonable to segregate pupils into single-gender classes. According to teacher B who teaches two coeducational and one all-girl computer class, girls benefit from single-gender education because they interact more in class compared to the situation in a mixed-gender classroom where they ask fewer questions. Teacher C who teaches the all-boy class thinks that boys generally believe to be more talented in science than girls and therefore debar girls from actively participating in science classes. Teacher A states that he cannot judge the issue since he does not have any experience with single-gender classes. The indetermination of teacher A might also be attributable to his relatively few years of work experience.

Generally, all three teachers do not think that one gender is more talented in computer studies compared to the other gender. They are uniformly of the opinion that girls are more talented in learning languages than boys, while boys are not more talented in technical subjects than girls (teacher B is unsure concerning the second point). None of the teachers expresses objections to girls being talented enough to work in technical professions and only teacher A thinks that mainly boys rather than girls should seek technical professions. The last statement might be considered to be consistent with the statement that girls are more talented in learning languages.

The teachers think that different strengths and weaknesses of girls and boys in different school subjects are *both* instilled by society and innate, while the first source is considered to be of greater importance. Comparing boys' and girls' overall performance in lower secondary school, teachers A and C state that boys have more problems keeping up with the educational contents than girls. Teacher B thinks that none of the genders has more or less problems compared to the other.[21]

[21] In addition to the facts presented, the questionnaires provide the information that all teachers agree that the school is well equipped with computers and software.

Table 9: Teachers' gender views and stereotypes

Question / Response of teacher	A	B	C
What do you think …			
… Is it wise to teach boys and girls separately in the computer class?	do not know	yes	yes
… Are boys or girls more talented with respect to computer studies?	neutral	neutral	neutral
According to your professional experience, which of the following statements are correct?			
… Girls are more talented than boys when it comes to learning languages.	yes	yes	yes
… Boys are more talented than girls when it comes to learning maths or science.	no	yes/unsure	no
… Girls have no talent for technical professions.	no	no	no
… Boys rather than girls should seek technical professions.	yes	no	no
… Different strengths and weaknesses are instilled rather than innate.	yes	yes	yes
… Different strengths and weaknesses are both instilled and innate.	yes	yes	yes
… On average, girls have more problems keeping up with the secondary school level's curriculum than boys.	no	no	no
… On average, boys have more problems keeping up with the secondary school level's curriculum than girls.	yes	no	yes

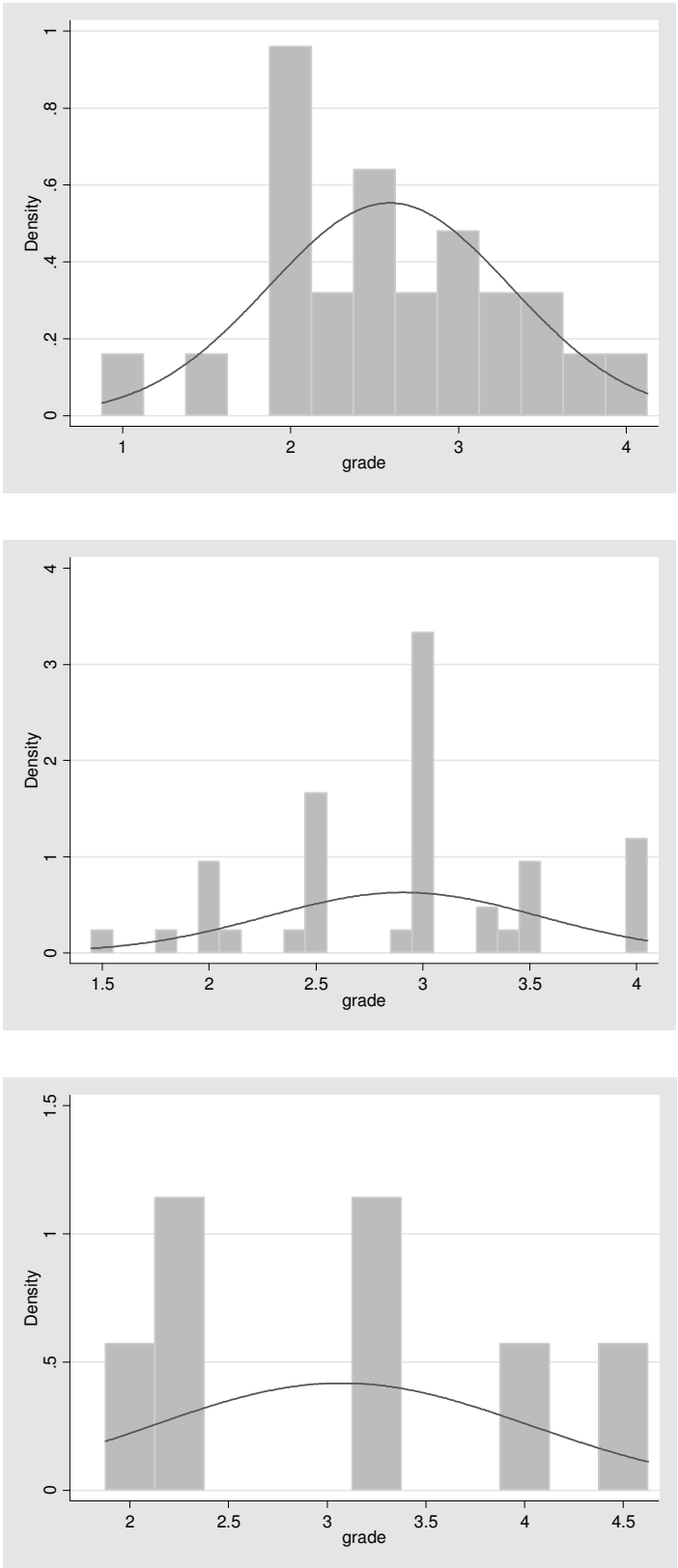
Source: Teacher questionnaires. Own calculations.

All in all, the answers do not indicate that one of the teachers is affected by serious 'traditional' gender stereotypes. However, one has to bear in mind that the teachers were informed about the contents of the school project. Thus, it cannot be ruled out that their answers are biased in the direction of the answers they expected to be politically correct.

Performance by teacher

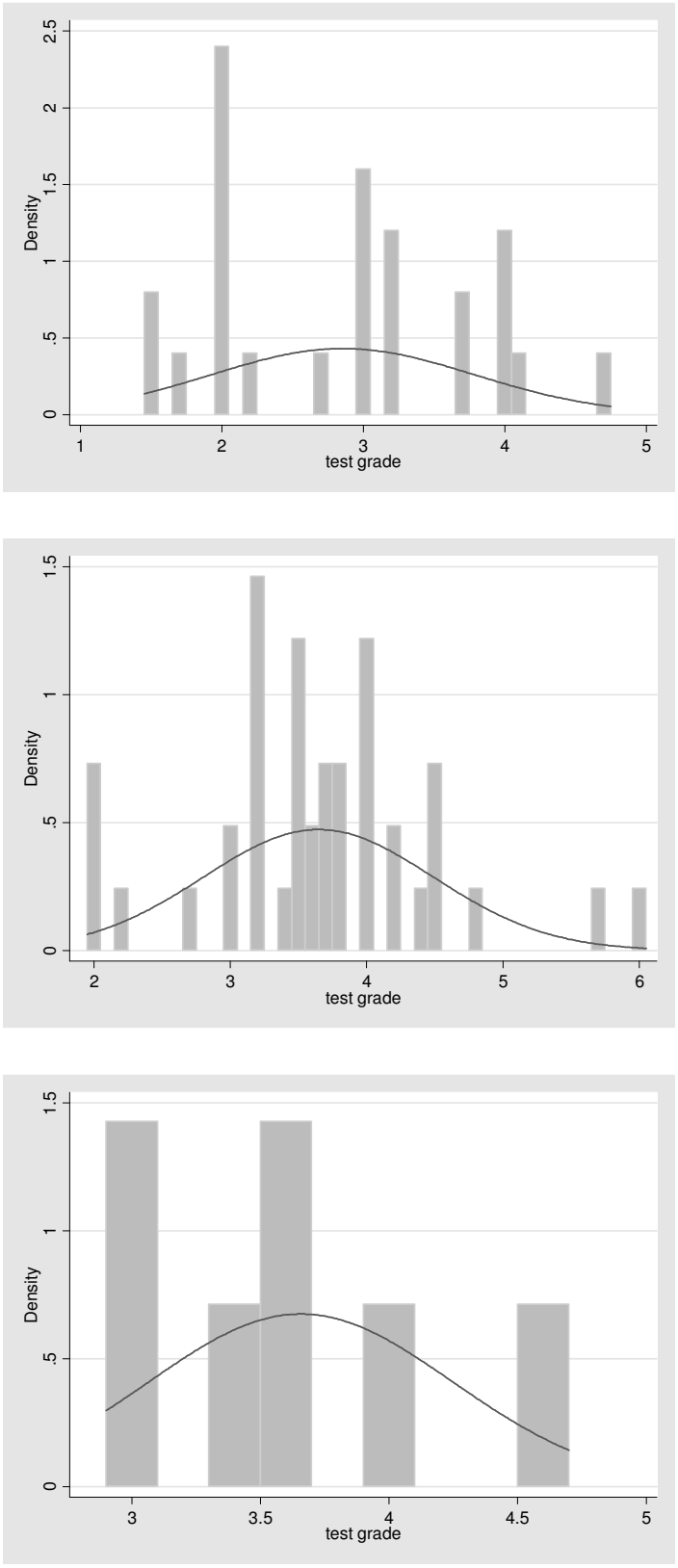
Even if the different groups are similarly composed, the students' outcomes will hardly be comparable between groups if teacher quality varies between groups. Figure 2 shows the distributions of grades by teacher in the computer test at mid-term (together with a sample normal distribution). Figure 3 refers to the computer tests at the end of term. Further statistics related to the distribution of grades by teacher are given in Table 10.

Figure 2: Distribution of mid-term computer test grades by teacher



Note: The distributions refer to teachers 1 to 3 respectively (from top to bottom).
Source: Test results. Own calculations.

Figure 3: Distribution of end of term computer test grades by teacher



Note: The distributions refer to teachers 1 to 3 respectively (from top to bottom).
Source: Test results. Own calculations.

It is obvious that the grade distributions differ substantially by teacher. While, for example the mode of the distribution related to teacher A at mid-term is grade 2, it is grade 3 for teacher B. The corresponding Kruskal-Wallis test related to the end of term grade indicates that one can reject the hypothesis of equality of populations at any conventional level of significance.[22] Since the teachers use the same scale in order to score the tests[23] and given that the socio-economic background of the students is very similar among groups, it is likely that these differences are not only driven by differences due to the single-gender and coeducational treatment but also by teaching differences.

Table 10: Distribution of grades by teacher

Teacher	A	B	C
Test grades at mid-term			
Mean	2.59	2.91	3.07
(s.d.)	0.72	0.64	0.95
Median	2.50	3.00	3.25
Mode	2.00	3.00	2.25, 3.25
Skewness	0.00	-0.07	0.29
Kurtosis	2.59	2.52	1.67
Kruskal-Wallis $\chi^2_{(2)}$		3.42	
Test grades at end of term			
Mean grade	2.86	3.65	3.66
(s.d.)	0.93	0.84	0.59
Median grade	3.00	3.60	3.70
Mode grade	2.00	3.20	3, 3.7
Skewness	0.18	0.33	0.53
Kurtosis	1.89	4.06	2.50
Kruskal-Wallis $\chi^2_{(2)}$		10.48	
Number of observations	25	41	7

Note: The Kruskal-Wallis test refers to the comparisons of the respective distributions for students instructed by different teachers.
Source: Test results. Own calculations.

To sum up, the descriptive evidence presented in this sub-section indicates that the groups are similar in their socio-economic background while they are principally educated in a similar way by the different teachers. However, there seem to be notable teacher differences in grading. Therefore, the following analysis needs to focus on comparisons of students taught by the same teacher to guarantee that similar students (under similar conditions) are considered.

[22] The Kruskal-Wallis test related to the mid-term grade can only be rejected at the 18%-level of significance.
[23] There are minor differences in the valuation of the first test, where teacher 2 uses a more exact scale based on intervals of first digits, *i.e.* [1.0; 1.1; 1.2; … ; 5.9; 6.0] while teacher 1 and 2 refer to intervals of quarter s of grades, *i.e.* [1.00; 1.25; 1.50; … ; 5.75; 6.00]. For the end of term test all teachers use the first digit scale.

3.3 Gender Related Findings

Before addressing the issue of coeducation, this section sheds light on the question whether there are overall performance differences by gender among the students observed in the intervention study. First of all hard performance measures related to different subjects are considered. Later on, soft performance measures are compared by gender.

Table 11 presents mean grade and performance measures separately by gender together with Kruskal-Wallis tests on the equality of the gender specific distributions. Kruskal-Wallis tests are appropriate if it might be assumed that the scales are metrical. Again, this assumption is reasonable for the grades in this case because of an underlying linear scale. The remaining scales can be taken to be metrical as well if one assumes that students' judge their own relative performance based on their grades.

Table 11: Performance in general subjects by gender

Group		All students		Girls		Boys	
Variable	Kruskal-Wallis $\chi^2_{(1)}$	Obs	Mean (s.d.)	Obs	Mean (s.d.)	Obs	Mean (s.d.)
Performance measured at mid-term:							
Overall comparative performance	0.34	74	2.70 (0.77)	46	2.76 (0.57)	28	2.61 (1.03)
Comparative performance in math	3.96**	74	2.80 (0.89)	46	2.96 (0.79)	28	2.54 (1.00)
Comparative performance in German	0.16	74	2.95 (0.90)	46	2.89 (0.88)	28	3.04 (0.96)
Performance measured at end of term:							
Overall comparative performance	0.46	73	2.84 (0.60)	45	2.89 (0.61)	28	2.75 (0.59)
Expected average grade (all subjects)	0.59	67	2.77 (0.79)	40	2.69 (0.81)	27	2.89 (0.76)
Comparative performance in math	3.83*	74	2.84 (0.79)	46	2.98 (0.68)	28	2.63 (0.91)
Expected math grade	2.47	73	3.05 (0.90)	46	3.19 (0.87)	27	2.82 (0.92)
Comparative performance in German	0.16	74	3.07 (0.75)	46	3.07 (0.68)	28	3.09 (0.86)
Expected German grade	1.41	74	3.28 (0.78)	46	3.19 (0.72)	28	3.44 (0.86)
Performance change[+]							
Change in overall comparative performance	0.00	73	0.12 (0.71)	45	0.11 (0.61)	28	0.14 (0.85)
Change in comparative math performance	0.05	74	0.05 (1.06)	46	0.02 (0.95)	28	0.09 (1.23)
Change in comparative German performance	0.79	74	0.13 (0.92)	46	0.17 (0.68)	28	0.05 (1.23)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for boys and girls. ** Significant at the five percent level. * Significant at the 10 percent level.
[+] The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

According to the comparative performance measures in Table 11, the average girl attributes a worse relative position to herself as compared to her classmates than the average boy does. This is especially true for math. However, for the German performance the reverse is true. These findings are true for both measurement dates, while the average person of the group indicating a better self-ranking (*i.e.* boys for their general performance and math, and girls for German) experiences a larger drop in his or her self-ranked position. Concerning expected end of term grades, the average girl generally (and especially in German) performs better than the average boy, while in math the reverse is true. The latter finding may explain the higher self-ranking of the average boy concerning his math performance. According to the Kruskal-Wallis tests, the hypothesis of equality of populations cannot be rejected for all available measures, but the comparative math performance measure can. Thus, one might assume that boys in the observed sample generally rate their relative math performance better than girls.

Table 12: Computer science performance by gender

Group		All students		Girls		Boys	
Variable	Kruskal-Wallis $\chi^2_{(1)}$	Obs	Mean (s.d.)	Obs	Mean (s.d.)	Obs	Mean (s.d.)
Performance measured at mid-term:							
Computer test grade	1.71	74	2.82 (0.71)	46	2.72 (0.68)	28	2.97 (0.73)
Comparative performance in computer science	1.27	73	2.41 (0.85)	45	2.49 (0.76)	28	2.29 (0.98)
Performance measured at end of term:							
Computer test grade	2.95	73	3.38 (0.92)	45	3.22 (0.92)	28	3.63 (0.89)
Comparative performance in computer science	0.52	74	2.74 (0.66)	46	2.80 (0.58)	28	2.64 (0.78)
Expected computer science grade	0.00	70	2.26 (0.67)	44	2.21 (0.52)	26	2.35 (0.87)
Performance change:[+]							
Change in computer test grade	0.62	73	0.54 (0.96)	45	0.47 (0.93)	28	0.65 (1.00)
Change in comparative performance	0.23	73	0.34 (0.77)	45	0.33 (0.64)	28	0.36 (0.95)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for boys and girls. ** Significant at the five percent level. * Significant at the 10 percent level.
[+] The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Table 12 additionally shows performance measures specifically related to the computer class. Concerning the mid-term and end of term computer tests, the average girl performs slightly better than the average boy. The same is true for the expected end of term grade. Even though, the average boy ranks himself on a relatively higher position compared to his classmate than the average girl does. This may be a hint for a higher self-esteem of boys. However, all the

differences are not substantial. The Kruskal-Wallis tests indicate that it is not possible to reject the hypothesis of equal populations for any of the available variables.

Table 13 shows the results related to the soft performance measures together with χ^2-homogeneity tests.[24] Most students state that they like working with computers when they are asked around mid-term. Only 9 % of the responding female students (4 out of 42) and 18 % of the male students (5 out of 23) do not like computer work. Based on the χ^2-test, it is not possible to reject the equality of the distribution of outcomes for boys and girls. At the end of term the proportion of students disliking computer work is larger and especially high among male students (46 % *vs.* 20 % of responding female students). The χ^2-test now rejects equality of the male and female distributions on the one-percent level of significance.

Table 13: Soft-performance measures by gender

Group		All students		Girls		Boys	
Variable	$p[\chi^2_{(1)}]$	Yes	No	Yes	No	Yes	No
Variables measures at mid-term:							
I like working with computers.	0.24	65	9	42	4	23	5
Boys know more about computers than girls.	0.00	16	56	0	45	16	11
Girls know more about computers than boys.	0.08	9	63	8	37	1	26
Both genders know about computers.	0.00	47	25	37	8	10	17
I prefer to be in a single-gender computer group.	0.29	32	41	18	28	14	13
Variables measures at end of term:							
I like working with computers.	0.01	52	22	37	9	15	13
Boys know more about computers than girls.	0.00	11	61	1	44	10	17
Girls know more about computers than boys.	0.05	6	66	6	39	0	27
Both genders know about computers.	0.04	55	17	38	7	17	10
I prefer to be in a single-gender computer group.	0.00	17	55	17	29	0	26

Note: The χ^2-Tests tests refer to the comparisons of the respective distributions for boys and girls. $p[\chi^2_{(1)}]$ indicates the level of significance.
Source: Student questionnaires. Own calculations.

The majority of students think that both genders are equal or similar as concerns their computer knowledge. However, girls' and boys' statements differ with respect to this question: At mid-term none of the girls think that boys are superior concerning their computer skills while the majority of male students (59 %) state that boys know more about computers. These notions converge somewhat around the end of term when one girl (2 %) indicates that boys know more about computers but only 37 % of the boys take this position. The proportion of students thinking that girls know more about computers is rather low: 17 % (13 %) of girls take this position at mid-term (end of term) and only one boy agrees at mid-term and none of the boys at the end of the term. While at mid-term boys most often state that they know more than what girls know about computers, most girls (82 %) assume that both genders know about the same as girls. The proportion of boys sharing the latter opinion rises at the end of term (from 37 % to 63 %).

[24] For the purpose of the χ^2-statistics, some outcome measures are recoded to be binary in order to assure that there are enough observations per cell so that the approximation of the test statistic is valid.

Around mid-term 44 % of all students, 39 % of the girls and even 52 % of the boys, would prefer to be taught in a single gender class. At the end of term, the proportion remains similar for girls (37 %) while, surprisingly none of the boys wants to be segregated. It is hard to interpret the reasons for the change in the boys' opinion concerning the institution of an all boys' class. Potential reasons might be related to the specific treatment of the teacher who instructed the all-boy class.

3.4 Main Results: Group Related Findings

This section examines whether there are feasible group differences related to the outcome measures. First of all, results related to hard performance measures are discussed in detail. Beside evaluation based on Kruskal-Wallis tests some evidence from regression analysis is presented. However, due to the impossibility of controlling for a variety of background characteristics because of the limited sample size, both techniques yield (by definition) the same results. Later on in this section, evidence related to the soft performance measures is presented.

Table 14: Computer science performance by group type

Group Type		Mixed Gender Groups						Single Gender Groups					
Sample	all	all		girls		boys		all		girls		boys	
Variable	K-W. $\chi^2_{(1)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Test grade	2.31	51	2.72 (0.70)	30	2.57 (0.69)	21	2.94 (0.67)	23	3.02 (0.71)	16	3.00 (0.61)	7	3.07 (0.95)
Comperat. performance	2.52	51	2.27 (0.78)	30	2.37 (0.72)	21	2.14 (0.85)	22	2.73 (0.94)	15	2.73 (0.80)	7	2.71 (1.25)
Performance measured at end of term:													
Test grade	0.79	50	3.32 (0.99)	29	3.11 (0.95)	21	3.61 (0.99)	23	3.50 (0.77)	16	3.43 (0.84)	7	3.66 (0.59)
Comperat. performance	0.24	51	2.75 (0.63)	30	2.80 (0.41)	21	2.67 (0.86)	23	2.74 (0.75)	16	2.81 (0.83)	7	2.57 (0.53)
Expected grade	0.56	47	2.33 (0.71)	28	2.26 (0.57)	19	2.42 (0.89)	23	2.13 (0.57)	16	2.13 (0.43)	7	2.14 (0.85)
Performance change:[+]													
Change in grade	0.08	50	0.57 (0.99)	29	0.50 (0.93)	21	0.67 (1.07)	23	0.47 (0.90)	16	0.43 (0.96)	7	0.59 (0.78)
Change in comp. perf.	4.1**	51	0.47 (0.76)	30	0.43 (0.63)	21	0.52 (0.93)	22	0.05 (0.72)	15	0.13 (0.64)	7	-0.14 (0.90)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in single-gender and mixed groups. ** Significant at the five percent level. * Significant at the 10 percent level. [+] The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

A simple comparison of outcomes by group types may give a first hint at whether single-gender education is effective. Table 14 shows means of hard performance measures separately

for students in single-gender and mixed groups. The numbers are provided separately by gender. Overall, there are no substantial performance differences between students in single-gender and mixed groups. On average, test grades of students in mixed-gender groups are better than those of students in single-gender groups and this is especially true for girls: At mid-term (end of term) the average girl in a mixed-gender group performs about 0.4 (0.3) grade points better than the average girl in a single-gender group. Additionally, the average students' comparative performance ranking is better for the average student educated in a mixed-gender group than for an average student from a single-gender group when measured at mid-term. However, the average end of term comparative performance measure takes about the same value for both groups and the average of the expected end of term grade is 0.2 grade points better for the single-gender group.

Table 15: Computer science performance by group

Group		Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
Variable	K-W. $\chi^2_{(5)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Computer test grade	4.88	13	2.58 (0.70)	12	2.98 (0.62)	12	2.60 (0.78)	14	2.75 (0.70)	16	3.00 (0.61)	7	3.07 (0.95)
Comparative performance	6.51	13	2.08 (0.95)	12	2.42 (0.67)	12	2.00 (0.60)	14	2.57 (0.76)	15	2.73 (0.80)	7	2.71 (1.25)
Performance measured at end of term:													
Computer test grade	17.08**	13	3.28 (0.80)	12	4.11 (0.97)	12	2.40 (0.85)	13	3.49 (0.56)	16	3.43 (0.84)	7	3.66 (0.59)
Comparative performance	2.28	13	2.62 (0.65)	12	2.92 (0.29)	12	2.67 (0.78)	14	2.79 (0.70)	16	2.81 (0.83)	7	2.57 (0.53)
Expected computer grade	6.38	12	2.39 (0.64)	11	2.59 (0.63)	12	2.38 (0.96)	12	1.98 (0.47)	16	2.13 (0.43)	7	2.14 (0.85)
Performance change:[+]													
Change in test grade	11.15**	13	0.70 (0.94)	12	1.13 (1.00)	12	-0.20 (0.88)	13	0.65 (0.73)	16	0.43 (0.96)	7	0.59 (0.78)
Change in comparative performance	5.86	13	0.54 (0.52)	12	0.50 (0.80)	12	0.67 (0.89)	14	0.21 (0.80)	15	0.13 (0.64)	7	-0.14 (0.90)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in different groups. ** Significant at the five percent level. * Significant at the 10 percent level. [+] The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Table D in the 'Appendix of Tables of Group Characteristics' additionally shows means of performance measures related to other subjects. There are no substantial performance differences between students in single gender and mixed gender groups in the first term. It is only in the second term that some of the performance measures indicate that students in the mixed groups perform worse than students in the single gender groups. It is not possible to identify whether this difference is an outcome related to the treatment in the mixed gender groups.

Table 15 provides means of performance variables related to the computer science class for all six groups. Average test grades around mid-term are similar for the different groups as it is confirmed by the Kruskal-Wallis test. However, as a matter of fact and according to the Kruskal-Wallis test the average end-of-term grades differ substantially between groups with the best average result for group 3 (2.40) and the worst for group 2 (4.11). The mean test grade of the all-girl group takes a value of 3.43 which is quite close to the overall average (3.38).

As discussed above, the huge differences in grades are driven by the different teachers instructing different groups. Because of the teacher differences, it is reasonable to specifically compare the all-girl group to the two other groups instructed by the *same* teacher. The results related to the corresponding groups are shaded grey in Table 15. In fact, there might be a hint that single gender education is effective: The average end of term performance is much worse in group 2 and slightly worse in group 4 (3.49) compared to the single gender group 5. However, conducting a Kruskal-Wallis test for the relevant groups reveals that one cannot reject the hypothesis of equal populations up to the 19 %-level of significance.

Table 16: Girls' computer science performance by group type

Group		Group 1		Group 2		Group 3		Group 4		Group 5	
Variable	K-W. $\chi^2_{(4)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:											
Computer test grade	5.67	8	2.38 (0.57)	6	2.87 (0.73)	9	2.61 (0.86)	7	2.50 (0.58)	16	3.00 (0.61)
Comparative performance	4.40	8	2.38 (0.92)	6	2.67 (0.52)	9	2.00 (0.71)	7	2.57 (0.53)	15	2.73 (0.80)
Performance measured at end of term:											
Computer test grade	10.44**	8	2.90 (0.69)	6	4.07 (1.01)	9	2.57 (0.93)	6	3.25 (0.56)	16	3.43 (0.84)
Comparative performance	0.32	8	2.75 (0.46)	6	2.83 (0.41)	9	2.78 (0.44)	7	2.86 (0.38)	16	2.81 (0.83)
Expected computer grade	2.38	7	2.24 (0.48)	6	2.58 (0.66)	9	2.22 (0.51)	6	2.03 (0.65)	16	2.13 (0.43)
Performance change: +											
Change in test grade	6.28	8	0.53 (0.98)	6	1.20 (0.95)	9	-0.04 (0.89)	6	0.58 (0.49)	16	0.43 (0.96)
Change in comparative performance	4.32	8	0.38 (0.52)	6	0.17 (0.75)	9	0.78 (0.67)	7	0.29 (0.49)	15	0.13 (0.64)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in different groups. ** Significant at the five percent level. * Significant at the 10 percent level. $^+$ The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students. Results are not shown for group 6 because there are no girls in this group.
Source: Student questionnaires, test results. Own calculations.

Concerning the comparative performance measure, girls in the single-gender group 5 tend to perform at the lower end of the group averages at mid-term and close to the overall average

when measured at the end of term. For the expected end of term computer grade, girls in the single-gender group also perform close to the overall average.

The above analysis refers to both boys and girls. However, in order to measure the effect of education in the all-girls group, it might be more appropriate to compare their performance to the performance of other *girls* in the remaining groups. Therefore, Table 16 shows the averages of performance measures of girls only.

The averages and Kruskal-Wallis test presented in Table 16 confirm the findings from Table 15. Again, there are substantial differences concerning the distribution of the end-of-term computer test grade. However, the effect of single gender education is not clear because at the end of term girls educated in the single gender group (group 5) perform only better than girls educated in *one* of the two coeducational groups educated by the same teacher (group 2) but worse than girls in the other group (group 4) if test grades and the expected computer grade are considered.

Table 17 shows averages of the performance measures for boys in each group. However, the group sizes drop to very small numbers so that it is hard to pin down any effect. Kruskal-Wallis tests are not feasible because of the small group sizes.

Table 17: Boys' computer science performance by group type

Group	Group 1		Group 2		Group 3		Group 4		Group 6	
Variable	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:										
Computer test grade	5	2.90 (0.82)	6	3.08 (0.52)	3	2.58 (0.63)	7	3.00 (0.76)	7	3.07 (0.95)
Comparative performance	5	1.60 (0.89)	6	2.17 (0.75)	3	2.00 (0.00)	7	2.57 (0.98)	7	2.71 (1.25)
Performance measured at end of term:										
Computer test grade	5	3.88 (0.61)	6	4.15 (1.03)	3	1.90 (0.17)	7	3.70 (0.52)	7	3.66 (0.59)
Comparative performance	5	2.40 (0.89)	6	3.00 (0.00)	3	2.33 (1.53)	7	2.71 (0.95)	7	2.57 (0.53)
Expected computer grade	5	2.60 (0.82)	5	2.60 (0.65)	3	2.83 (1.89)	6	1.92 (0.20)	7	2.14 (0.85)
Performance change: +										
Change in test grade	5	0.98 (0.91)	6	1.07 (1.13)	3	-0.68 (0.79)	7	0.70 (0.93)	7	0.59 (0.78)
Change in comparative performance	5	0.80 (0.45)	6	0.83 (0.75)	3	0.33 (1.53)	7	0.14 (1.07)	7	-0.14 (0.90)

Note: Kruskal-Wallis tests are not feasible because of the small sample sizes. Numbers of observations vary due to missing information for some students. [+] The change variables indicates the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students. Results are not shown for group 5 because there are no boys in this group.
Source: Student questionnaires, test results. Own calculations.

In addition to the assessment by group types *via* Kruskal-Wallis tests, the following paragraphs report results from ordinary least squares regressions. The OLS results may be consid-

ered to be complementary to the above results. They are presented for the sake of completeness since regression analysis is the tool most often used by empirically oriented economists. Generally, regression analysis provides a simple tool for measuring the effect of interest controlling for different background variables. However, it needs to be verified that the underlying assumptions of the classical linear regression model are met and (in order to use common test statistics and in light of the small sample size) that the errors are normally distributed.

Table 18: Simple OLS regressions of mid-term and end of term grades

Outcome variable	Mid-term grade			End of term grade		
Sample	All students	Girls	Girls, Teacher B^{+}	All students	Girls	Girls, Teacher B^{+}
Single-gender effect	0.2972*	0.4267**	0.3308	0.1737	0.3147	-0.2333
(s.e.)	(0.1758)	(0.2046)	(0.2343)	(0.2336)	(0.2856)	(0.3287)
Sample Size	74	46	29	73	45	28

Note: $^{+}$ This sample covers only girls instructed by the same teacher B (who teaches two coeducational and one single-gender groups).
* Significant at the ten percent level. ** Significant at the five percent level.
Source: Student questionnaires, test results. Own calculations.

Table 18 presents regression results for different samples of interest for the mid-term and end of term grades. Regressing the outcome variables on a dummy variable for single-gender treatment for all students yields a significant effect for the mid-term grade and an insignificant effect for the end of term grade: It seems that students instructed in single-gender classes perform *worse* than students in coeducational classes around mid-term. The effect is especially pronounced if the sample is restricted to girls.[25] However, if the restriction refers to girls taught by the same teacher, the single-gender effect disappears. Therefore, it seems that the effect observed when not considering students taught by the same teachers is an artificial effect due to the teacher and not due to the organizational type.

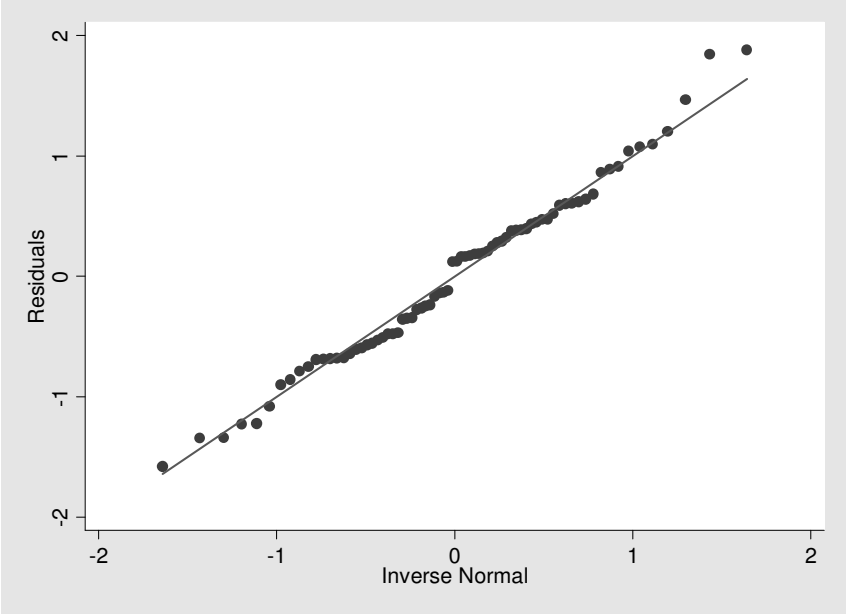
Further regressions which have been conducted additionally control for students' background variables. However, as one would probably expect (given the limited sample size) the coefficient of the single-gender group indicator is always insignificant. This is why these results are not documented in detail.

The meaning of the regression results relies (among other assumptions) on the validity of the assumption that the error terms are normally distributed.[26] Figure 4 presents the normal probability plot according to Chambers *et al.* (1983), which is a simple check of the assump-

[25] By definition, the effects presented here are identical to the ones implied in Table 14.
[26] Several tests have been conducted in order to verify assumptions of the classical linear regression model under which the OLS-estimates have the standard desired properties. Specifically, given the fact that students are taught in different groups, one might assume that errors are heteroskedastic. However, Breusch-Pagan and White-tests for heteroskedasticity based on different sets of variables could not reject the null of homoskedastic errors.

tion that the error terms are normally distributed: Fractiles of the error distribution are plotted versus the fractiles of a normal distribution having the same mean and variance.[27] Eyeballing suggests a rather linear pattern of the normal probability plot, *i.e.* the plot falls close to the diagonal line. Additionally, a Shapiro-Wilk test for normality yields the same result in favour of the normal distribution of the errors. Therefore, the assumption may be taken to be valid.[28]

Figure 4: Normal probability plot



Source: Student questionnaires, test results. Own calculations.

Besides the hard performance measures which have been discussed so far, evidence on soft performance measures has been assessed as well. However, in light of the limited sample size it is hard to deduce anything from categorical variables. Therefore, the results are not discussed in detail. Table 19 presents the results for girls in the different groups, where grey-shaded columns refer to groups taught by the same teacher.

Table 19: Motivation and gender perception of girls by group

Group	Group 1		Group 2		Group 3		Group 4		Group 5	
Variable	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Variables measures at mid-term:										
I like working with computers.	8	0	5	1	7	2	7	0	15	1
Boys know more about computers.	0	7	0	6	0	9	0	7	0	16
Girls know more about computers.	0	7	2	4	0	9	0	7	6	10
Both genders know about computers	7	0	4	2	9	0	7	0	10	6
I prefer single-gender group.	7	1	3	3	0	9	1	6	7	9
Variables measures at end of term:										
I like working with computers.	4	4	6	0	8	1	5	2	14	2
Boys know more about computers.	1	6	0	6	0	9	0	7	0	16
Girls know more about computers.	1	6	0	6	0	9	2	5	3	13
Both genders know about computers	5	2	6	0	9	0	5	2	13	3
I prefer single-gender group.	4	4	3	3	8	1	2	5	7	9

Source: Student questionnaires. Own calculations.

[27] Exemplarily, the presented plot refers to the most general regression discussed above.
[28] The additional test whether the errors are log normally distributed yielded rejection of the null.

From the numbers presented in Table 19 it is not possible to infer that girls taught in the single-gender group (group 5) differ from girls in the other groups taught by the same teacher (group 2 and 4) concerning the soft performance measures.

4 Summary and Discussion

The present paper documents that there is still plenty of scope for conducting research on the effects of single-gender education. Existing studies are problematic by design since the effects they identify are either only valid for a very specific group or biased because of selection or further implementation problems. Also the presented BW-project does not identify an effect of single-gender education. Performance differences between the single-gender and mixed-gender groups seem to be caused by teacher differences rather than by the organizational treatment. Causal effects are hard to identify in light of a small sample size, confounding factors and sample attrition. Additionally, the small sample based on pupils from one single school's fifth graders may hardly be taken to be representative for the entire German school population. A further drawback of this and many existing studies is that the restriction to a very limited time-span (one observation year) does not allow examining long-run effects.

From all this evidence it is possible to deduce general elements for a potential 'ideal' intervention study. Briefly, in the absence of feasible experiments, features of such a study relate to: (1) large sample sizes together with a standardized assessment of pupils' characteristics in order to be able to control for pupils' selection into different groups using appropriate statistical techniques, (2) a longer project period of several schooling years in order to observe potential long-run effects (and probably also to draw on difference-in-difference techniques), (3) the reduction of confounding factors related to the (educational) treatment of pupils in different groups, *e.g.* through assuring that the same teachers instruct coeducational as well as single-gender groups, (4) the prevention of *Pygmalion* effects, *e.g.* through concealing the gender-related purpose of the intervention project, (5) the reduction of sample attrition (due to missing observations) or group changes through a careful supervision of the project. Especially points (3) to (5) require the support of the participating teachers. Written guidelines and regular meetings of all project partners (teachers and researchers) are certainly important in order to assure the appropriate implementation of the project.

Even if these guidelines are followed, there might still be some problems related to single gender education which have not been addressed so far. Critics of such an intervention study may ask the (legitimate) question why female students should be specifically supported

in science studies while male students are not supported in languages and other subjects where they are outperformed by girls. The PISA studies have shown that male students especially, and especially those from the lower secondary schools, show poor reading abilities (cf. for example Röhner, 2003). A stream of the recent education literature emphasizes that most 'problem children' (*e.g.* children displaying aggressive behaviour or other behavioural problems as well as extremely poor school outcomes) are boys and call for new educational measures which focus on boys' development and performance (cf. for example Böhmann, 2003 and Kraus, 1998). In fact, a coherent project could extend the topic of analysis and focus on both genders and different fields of studies.

A further point raised by critics of single gender studies is that it is not clear whether students *want* to be educated separately by gender in the first place (cf. this critique in Biermann, 2000). In the presented BW-project for example, most pupils stated that they prefer to be educated in mixed classes. Related to this issue is the argument that it might be important for girls to learn to succeed in the 'real', competitive and mixed gendered world (cf. Meyer, 1996). On the one hand, single gender education might provide an artificial learning environment and it is not guaranteed that pupils succeeding in this environment will succeed when entering, for example, the mixed gendered labour markets. On the other hand, single gender education might be an appropriate tool for enhancing girl's interest in technical subjects in the first place, while it is possible that these girls are well capable of competing with their male classmates later on (once they have started to put an effort in acquiring the specific skills). Therefore, these arguments tend to support the requirement of conducting *long-term* studies on the effects of single gender education than ruling out the meaning of such studies in the first place.

Researchers opposing single gender experiments often argue that girls' interest and success in computer studies and occupations could be enhanced *within* coeducational classes. Existing studies along this line usually emphasize that the present situation in schools is characterized by co-instruction (instead of coeducation), *i.e.* girls and boys are taught in the same class but a 'secret curriculum' (*heimlicher Lehrplan*) focuses on the abilities and interests of boys. One example is that teachers do not consider different problem-solving strategies of boys and girls but favour the 'male techniques' (cf. Funken *et al.*, 1996, p.128). Concerning different interests related to computer studies, traditional curricula do not consider that girls are less interested in pure programming but more in the application of computer skills, for example using computers for creative tasks (cf. Rentmeister, 1992 and Schultz-Zander, 1992). Further strategies that have been suggested in order to foster girls interest in computers within

coeducational classes are for example (1) to discuss the impact of computers on modern societies in the computer class (since girls are usually more interested in social topics) (2) to provide 'good examples', *i.e.* to inform on successful female careers in computer sciences or to (3) increase the number of female computer science teachers.[29] Generally, such approaches which consider the gender perspectives within the coeducational classroom have become known as "reflexive coeducation".

The present (empirical) literature is not conclusive regarding which one of the two ways – single gender education or reflexive coeducation – is more suited to reducing subject-related gender differences. Again, there is plenty of scope for further research on the benefits of single gender education under this perspective.

In summing up, it can be maintained that different streams of discussion call for a more reliable empirical foundation of single gender research. A larger-scale intervention study following the suggested guidelines could crucially improve the insights into this topic given that the presently available evidence is merely speculative.

[29] Cf. Rentmeister, 1992 and, for example, Nossek, 2006 for more specific suggestions of 'reflexive coeducation' in modern computer science classes The study by Hoffmann *et al.* (1997) mentioned above additionally reviews existing evidence and provides evidence on the effects of changed science curricula on girls' and boys' educational outcomes. Lechner (2002) is a follow-up study to the study by Kessels (2002) and emphasizes the role of different educational strategies within science (physics) classes. An extensive discussion and summary of the evidence on reflexive coeducation is given in Faulstich-Wieland (2004b).

References

Baumert, J. (1992): Koedukation oder Geschlechtertrennung, *Zeitschrift für Pädagogik* 38: 83-110.

Biermann, C. (2000): Lernziel Gleichberechtigung – überflüssiger oder aktueller denn je? Der Versuch einer Zwischenbilanz der Koedukationsdebatte, *Der Deutschunterricht* 52: 90-93.

Billger, Sh.M. (2006): *Reconstructing School Segregation: On the Efficacy and Equity of Single-Sex Schooling*, IZA Discussion Paper No. 2037.

Black, S.E. and A. Spitz-Oener (2007): *Explaining Women's Success: Technological Change and the Skill Content of Women's Work*, NBER Working Paper No. 13116.

BLK (2002): *Frauen in den ingenieur- und naturwissenschaftlichen Studiengängen*, Bericht der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 2.Mai 2002, Bonn.

Böhmann, M. (2003): Jungen in der Schule – ein Problemfall? Die neue Diskussion um die Jungen, *Pädagogik* 55: 32-35.

Brendel, S. (1995): Mädchen in der Hauptschule, in Mack, W. (ed.): *Hauptschule als Jugendschule, Beiträge zur pädagogischen Reform der Hauptschulen in sozialen Brennpunkten*, Süddeutscher Pädagogischer Verlag, Ludwigsburg.

Chambers, J., W. Cleveland, B. Kleiner and P. Tukey (1983): *Graphical Methods for Data Analysis*, Wadsworth International, Duxbury Press, Boston.

Faulstich-Wieland, H. (2004a): Wem nützt die Einrichtung einer Mädchenklasse?, *Zeitschrift für Frauenforschung und Geschlechterstudien* 22: 39-57.

Faulstich-Wieland, H. (2004b): *Mädchen und Naturwissenschaften in der Schule*, Expertise für das Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg, Hamburg.

Faulstich-Wieland, H. (1999): Koedukation heute – Bilanz und Chance, in Horstkemperer, M. and M. Kraul (eds.): *Koedukation. Erbe und Chancen*, Deutscher Studien Verlag, Weinheim.

Funken, C., K. Hammerich and B. Schinzel (1996): *Geschlecht, Informatik und Schule*, Academica Verlag, Sankt Augustin.

Häußler, P. and L. Hoffmann (2002): An Intervention Study to Enhance Girls' Interest, Self-Concept, and Achievement in Physics Classes, *Journal of Research in Science Teaching* 39: 870-888.

Häußler, P. and L. Hoffmann (1998): Chancengleichheit für Mädchen im Physikunterricht – Ergebnisse eines erweiterten BLK-Modellversuchs, *Zeitschrift für Didaktik der Naturwissenschaften* 4: 51-67.

Häußler, P. and L. Hoffmann (1990): Physikunterricht für Mädchen und Jungen getrennt? Zwei Interviews aus der Schulpraxis, *Naturwissenschaften im Unterricht. Physik* 1: 32-33.

Heckman, J.J., H. Ichimura and P. Todd (1998): Matching as an Econometric Evaluation Estimator, *Review of Economic Studies* 65: 261-294.

Heckman, J.J. and J.A. Smith (1995): Assessing the Case for Social Experiments, *Journal of Economic Perspectives* 9: 85-110.

Heidtmann, H. (1998): Multimedia und Mädchen, *Praxis Schule 5-10* 9: 25-27.

Heinrichs, U. and T. Schulz (1989): Mädchen und Naturwissenschaften. Wahlverhalten in der gymnasialen Oberstufe, *Pädagogik* 41: 36-39.

Hoffmann, L., P. Häußler and S. Peters-Haft (1997): *An den Interessen von Jungen und Mädchen orientierter Physikunterricht, Ergebnisse eines BLK-Modellversuchs*, Institut für die Pädagogik der Naturwissenschaften, Kiel.

Holz-Ebeling, F., J. Grätz-Tümmers and C. Schwarz (2000): Jungen als „Nutznießer" der Koedukation? Eine empirische Studie zur Bedeutung der Koedukation für Jungen, *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 32: 94-107.

IW (2005): *Deutschland in Zahlen*, Institut der Deutschen Wirtschaft Köln, Deutscher Instituts-Verlag, Köln.

Kessels, U. (2002): *Undoing Gender in der Schule, Eine empirische Studie über Koedukation und Geschlechtsidentität im Physikunterricht*, Juventa Verlag, Weinheim/München.

Kraus, J. (1998): Koedukation oder Desintegration - Schule auf dem Weg zur Quote? *Profil* 4: 14-15.

Kruskal, W.H. and W.A. Wallis (1952): Use of Ranks in One-Criterion Variance Analysis, *Journal of the American Statistical Association* 47: 583–621.

Lechner, H. (2002): Reflexive Koedukation – eine notwendige Bedingung für den Anfangsunterricht Physik?, in R. Brechel (ed.): *Zur Didaktik der Physik und der Chemie, Tagung 2001*, 86-88.

Ludwig, P.H. (2003): Partielle Geschlechtertrennung – enttäuschte Hoffnung? Monoedukative Lernumgebungen zum Chancenausgleich im Unterricht auf dem Prüfstand, *Zeitschrift für Pädagogik* 49: 640-656.

Mael, F., A. Alonso, D. Gibson, K. Rogers and M. Smith (2005): *Single-Sex Versus Coeducational Schooling: A Systematic Review*, U.S. Department of Education, Policy and Program Study Service Doc. 2005-01, Washington.

Marsh, H.W. and K.J. Rowe (1996): The Effects of Single-Sex and Mixed-Sex Mathematics Classes Within a Coeducational School: A Reanalysis and Comment, *Australian Journal of Education* 40: 147-162.

Marsh, H.W., I.D. Smith, M. Marsh and L. Owens (1988): The Transition From Single-Sex to Coeducational High Schools: Effects on Multiple Dimensions of Self-Concept and on Academic Achievement, *American Educational Research Journal* 25: 237-269.

Meyer, M. (1996): Wenn die Mädchen dabei gewesen wären… Schülerinnen und Schüler äußern sich über ITG-Unterricht in geschlechtshomogenen Gruppen, *Computer und Unterricht* 6: 50-52.

Nossek, G. (2006): Reflexive Koedukation, *Computer und Unterricht* 16: 36-37.

Nyssen, E. (1996): *Mädchenförderung in der Schule, Ergebnisse und Erfahrungen aus einem Modellversuch*, Juventa Verlag, Weinheim/München.

Rastatt (2005): *Rastatt in Zahlen*, Stadt Rastatt.

Rentmeister, C. (1992): Computer und Kreativität, in Glumper, E. (eds.): *Mädchenbildung – Frauenbildung*, Verlag Julius Klinkhard, Bad Heilbrunn, 221-227.

Roeder, P. and S. Gruehn (1997): Geschlecht und Kurswahlverhalten, *Zeitschrift für Pädagogik* 43: 877-894.

Röhner, C. (2003): Bildungsverlierer: Jungen? Zur Koedukationsdebatte nach PISA, *Praxis Schule 5-10* 14: 11-15.

Rost, D.H. and C. Pruisken (2000): Vereint schwach? Getrennt stark? Mädchen und Koedukation, *Zeitschrift für pädagogische Psychologie* 14: 177-193.

Rustmeyer, R. and A. Jubel (1996): Geschlechtsspezifische Unterschiede im Unterrichtsfach Mathematik hinsichtlich der Fehleinschätzung, Leistungserwartung, Attribution sowie im Lernaufwand und im Interesse, *Zeitschrift für Pädagogische Psychologie* 10: 13-25.

Schuld, E. (1997): Das Internet: Benachteiligung oder Förderung von Mädchen am Computer?, *Didaktik der Berufs- und Arbeitswelt* 4: 21-30.

Schultz-Zander, R. (1992): Für die Gleichstellung von Mädchen und jungen Frauen in der Informationstechnischen Bildung, in E. Glumper (ed.): *Mädchenbildung – Frauenbildung*, Verlag Julius Klinkhard, Bad Heilbrunn, 210-220.

Thierack, A. (1995): In der Schule benachteiligt, in der Koedukationsdebatte vergessen: Lebenswelten von HauptschülerInnen, in M. Hempel (ed.): *Verschieden und doch gleich. Schule und Geschlechterverhältnisse in Ost und West*, Klinkhardt, Bad Heilbrunn.

Volmerg, B., A. Creutz, M. Reinhardt and T. Eiselen (1996): *Ohne Jungs ganz anders? Geschlechterdifferenz und Lehrerrolle am Beispiel eines Schulversuchs*, Kleine Verlag, Bielefeld.

Ziegler, A., P. Broome and K.A. Heller (1998): Pygmalion im Mädchenkopf, Erwartungs- und Erfahrungseffekte koedukativen vs. geschlechtshomogenen Physikanfangsunterrichts, *Psychologie in Erziehung und Unterricht*, 45: 2-18.

Appendix of Questionnaires

The following pages display the original pupil, parent and teacher questionnaires (in German language) distributed in the first year of the BW-project. The same questions are repeated in the second-year questionnaires. The ‘Appendix on the Students’ Background’ provides descriptive statistics for the variables drawn from these questionnaires.

Schulprojekt: Computer in der Schule

Schülerfragebogen (wird in der Schule ausgefüllt) zum Halbjahr, Klasse 5a Gruppe 1

Bitte kreuze nur die Antworten an, die auf dich zutreffen: ☒

1. Bitte gib die Nummer an, die dir dein Lehrer für diese Befragung gegeben hat:

   ☐

2. Wie alt bist du?

   ☐ Jahre

3. Hat deine Familie zu Hause einen Computer?

   Ja ☐ Nein ☐

4. Wie viel Spaß macht dir der Unterricht am Computer? Kreuze bitte das an, was zutrifft.

   Sehr viel Spaß ☐
   Eher viel Spaß ☐
   Weder viel noch wenig Spaß ☐
   Eher wenig Spaß ☐
   Keinen Spaß ☐

5. Wie sind bisher in der Hauptschule deine Leistungen allgemein im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere. ☐
   Ich erbringe bessere Leistungen als andere. ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere. ☐
   Andere erbringen bessere Leistungen. ☐
   Andere sind viel besser. ☐

6. Wie sind bisher in der Hauptschule deine Leistungen *im Computerunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

   Ich erbringe viel bessere Leistungen als andere. ☐
   Ich erbringe bessere Leistungen als andere. ☐
   Ich erbringe weder bessere noch schlechtere Leistungen als andere. ☐
   Andere erbringen bessere Leistungen. ☐
   Andere sind viel besser. ☐

7. Wie sind bisher in der Hauptschule deine Leistungen *im Matheunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere.	☐
Ich erbringe bessere Leistungen als andere.	☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere.	☐
Andere erbringen bessere Leistungen.	☐
Andere sind viel besser.	☐

8. Wie sind bisher in der Hauptschule deine Leistungen *im Deutschunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere.	☐
Ich erbringe bessere Leistungen als andere.	☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere.	☐
Andere erbringen bessere Leistungen.	☐
Andere sind viel besser.	☐

9. Was würdest du am Unterricht am Computer verbessern?

10. Beantworte diese Frage nur, wenn du ein Mädchen bist: Findest du es angenehmer, in einer Klasse nur mit Mädchen zu sein?
Kreuze bitte an, was du gut findest!

Nur mit Mädchen ☐ Mit Jungen und Mädchen ☐

11. Beantworte diese Frage nur, wenn du ein Junge bist: Findest du es angenehmer, in einer Klasse nur mit Jungen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Jungen ☐ Mit Jungen und Mädchen ☐

12. Wer meinst du, kennt sich besser mit Computern aus: eher die Jungs oder eher die Mädchen?

Eher die Jungs ☐ Eher die Mädchen ☐ Beide ☐

Schulprojekt: Computer in der Schule

Schülerfragebogen (wird in der Schule ausgefüllt)
zum Schuljahresende, Klasse 5a, Gruppe 1

Bitte kreuze nur die Antworten an, die auf dich zutreffen: ☒

1. Bitte gib die Nummer an, die dir dein Lehrer für diese Befragung gegeben hat:

2. Hat deine Familie zu Hause einen Computer? Ja ☐ Nein ☐

3. Wie viel Spaß macht dir der Unterricht am Computer? Kreuze bitte das an, was zutrifft.

Sehr viel Spaß ☐
Eher viel Spaß ☐
Weder viel noch wenig Spaß ☐
Eher wenig Spaß ☐
Keinen Spaß ☐

4. Wie sind bisher in der Hauptschule deine Leistungen allgemein im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere ☐
Ich erbringe bessere Leistungen als andere ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere ☐
Andere erbringen bessere Leistungen ☐
Andere sind viel besser ☐

5. Wie sind bisher in der Hauptschule deine Leistungen *im Computerunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere ☐
Ich erbringe bessere Leistungen als andere ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere ☐
Andere erbringen bessere Leistungen ☐
Andere sind viel besser ☐

6. Wie sind bisher in der Hauptschule deine Leistungen *im Matheunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere ☐
Ich erbringe bessere Leistungen als andere ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere ☐
Andere erbringen bessere Leistungen ☐
Andere sind viel besser ☐

7. Wie sind bisher in der Hauptschule deine Leistungen *im Deutschunterricht* im Vergleich zu denen deiner Mitschüler ausgefallen? Kreuze bitte das an, was zutrifft.

Ich erbringe viel bessere Leistungen als andere ☐
Ich erbringe bessere Leistungen als andere ☐
Ich erbringe weder bessere noch schlechtere Leistungen als andere ☐
Andere erbringen bessere Leistungen ☐
Andere sind viel besser ☐

8. Gib eine Gesamtnote zwischen 1 und 6 an, die Deine Leistungen beschreibt, die Du durchschnittlich im Zeugnis erwartest: ☐

9. Gib eine Gesamtnote zwischen 1 und 6 an, die Deine Leistungen beschreibt, die Du im Computerunterricht im Zeugnis erwartest: ☐

10. Welche Note bekommst Du wahrscheinlich in Deutsch? ☐

11. Welche Note bekommst Du wahrscheinlich in Mathe? ☐

12. Wie viele Brüder hast Du?

☐ Brüder

13. Wie viele Schwestern hast Du?

☐ Schwestern

14. Wie viele Kinder (ohne Dich) leben bei Dir zu Hause?

☐ Kinder

15. Wie alt sind diese Kinder, die bei Dir zu Hause wohnen?

Alter des 1. Kindes (falls vorhanden):	☐	Jahre
Alter des 2. Kindes (falls vorhanden):	☐	Jahre
Alter des 3. Kindes (falls vorhanden):	☐	Jahre
Alter des 4. Kindes (falls vorhanden):	☐	Jahre
Alter des 5. Kindes (falls vorhanden):	☐	Jahre
Alter des 6. Kindes (falls vorhanden):	☐	Jahre

16. Wurdest Du in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

17. Wurde Dein Vater in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

18. Wurde Deine Mutter in Deutschland geboren?

Ja.. ☐ Nein.. ☐ weiß ich nicht.. ☐

19. Sprecht Ihr zu Hause eine Sprache, die nicht die deutsche Sprache ist?

Ja.. ☐ Nein.. ☐

20. Wenn Ihr zu Hause eine Sprache sprecht, die nicht die deutsche Sprache ist, welche Sprache ist das? Schreibe die Sprache in das Kästchen.

21. Wohnen Deine Großeltern (z.B. Deine Oma oder Dein Opa) bei Dir zu Hause oder in der Nähe?

Ja.. ☐ Nein.. ☐

22. Welchen Schulabschluss hat Deine Mutter? Kreuze das Zutreffende an!

Keinen..... ☐
Hauptschule (oder vergleichbar)..... ☐
Realschule (oder vergleichbar)..... ☐
Gymnasium (oder vergleichbar)..... ☐
Weiß ich nicht..... ☐

23. Welchen Schulabschluss hat Dein Vater? Kreuze das Zutreffende an!

Keinen..... ☐
Hauptschule (oder vergleichbar)..... ☐
Realschule (oder vergleichbar)..... ☐
Gymnasium (oder vergleichbar)..... ☐
Weiß ich nicht..... ☐

24. Hast Du einen Vorschlag, wie der Unterricht am Computer besser werden könnte?

---

25. Beantworte diese Frage nur, wenn du ein Mädchen bist: Findest du es angenehmer, im Computerunterricht in einer Klasse nur mit Mädchen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Mädchen.... ☐ Mit Jungen und Mädchen.... ☐

26. Beantworte diese Frage nur, wenn du ein Junge bist: Findest du es angenehmer, im Computerunterricht in einer Klasse nur mit Jungen zu sein? Kreuze bitte an, was du gut findest!

Nur mit Mädchen.... ☐ Mit Jungen und Mädchen.... ☐

27. Wer meinst du, kennt sich besser mit Computern aus: eher die Jungs oder eher die Mädchen?

Eher die Jungs... ☐ Eher die Mädchen... ☐ Beide... ☐

Schüler- Nr.:

Schulprojekt: Computer in der Schule

Fragebogen: Eltern

Bitte kreuzen Sie nur die Antworten an, die auf Sie zutreffen: ☒

1. Sind Sie die Mutter (bzw. weibliche Bezugsperson) oder der Vater (bzw. die männliche Bezugsperson) des Kindes in der 5. Klasse?

   Mutter (bzw. weibliche Bezugsperson): ☐
   Vater (bzw. männliche Bezugsperson): ☐

2. Sind Sie allein erziehend?

   Ja ☐ Nein ☐

3. Bitte geben Sie die Gesamtzahl Ihrer Kinder an: ☐

4. Wie viele Kinder leben in Ihrem Haushalt? ☐

5. Wie alt sind diese Kinder? Bitte geben Sie das Alter aller Kinder, die in Ihrem Haushalt leben, in Jahren an (z.B. schreiben Sie die Zahl 8 in ein Kästchen, wenn Ihr Kind 8 Jahre alt ist).

   Alter des ersten Kindes: ☐ Jahre
   Alter des zweiten Kindes (falls vorhanden): ☐ Jahre
   Alter des dritten Kindes (falls vorhanden): ☐ Jahre
   Alter des vierten Kindes (falls vorhanden): ☐ Jahre
   Alter des fünften Kindes (falls vorhanden): ☐ Jahre

6. Wohnt ein Teil der Großeltern (ein Opa oder eine Oma) des Kindes in Ihrem Haushalt oder in der Nähe?

   Ja ☐ Nein ☐

7. Wurde Ihr Kind, welches die 5. Klasse besucht, in Deutschland geboren?

   Ja ☐ Nein ☐

8. Wurden Sie in Deutschland geboren?

   Ja ☐ Nein ☐

9. Wurde der andere Elternteil des Kindes beziehungsweise die andere erwachsene Bezugsperson des Kindes in Deutschland geboren?

   Ja ☐ Nein ☐

10. Welche Staatsbürgerschaft besitzen Sie (z.B. Deutsch)? ☐

11. Welche Staatsbürgerschaft besitzt die andere erwachsene Bezugsperson des Kindes?

12. Sprechen Sie mit Ihrem Kind eine Sprache, die nicht die deutsche Sprache ist?

Ja ☐ Nein ☐

13. Wenn Sie Frage 12 mit Ja beantwortet haben: Welche Sprache ist das?

14. Haben Sie in Ihrem Beruf mit Computern zu tun?

Ja ☐ Nein ☐

15. Wenn Sie Frage 14 mit Ja beantwortet haben: Beschreiben Sie, auf welche Art und Weise Sie in Ihrem Beruf Computer nutzen:

_____

_____

_____

_____

16. In welchem Umfang sind Sie berufstätig?

Nicht berufstätig ☐ Geringfügig beschäftigt ☐
Teilzeit beschäftigt ☐ Vollzeit beschäftigt ☐

17. In welchem Umfang ist der andere Elternteil des Kindes (beziehungsweise die andere erwachsene Bezugsperson des Kindes) berufstätig?

Nicht berufstätig ☐ Geringfügig beschäftigt ☐
Teilzeit beschäftigt ☐ Vollzeit beschäftigt ☐

18. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) beruflich mit Computern zu tun?

Ja ☐ Nein ☐

19. Wenn Sie Frage 18 mit Ja beantwortet haben: Beschreiben Sie, auf welche Art und Weise der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) beruflich einen Computer nutzt:

_____

_____

_____

_____

20. Welchen allgemeinen Schulabschluss (Hauptschule, Realschule oder Gymnasium) besitzen Sie?

Keinen allgemeinen Schulabschluss ☐

Hauptschulabschluss
oder vergleichbaren Abschluss, z.B. im Ausland erworben ☐

Mittlere Reife / Realschulabschluss oder vergleichbaren Abschluss ☐

Abitur, Fachhochschulreife od. vergleichbaren Abschluss ☐

21. Welchen allgemeinen Schulabschluss (Hauptschule, Realschule oder Gymnasium) besitzt der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes)?

Keinen allgemeinen Schulabschluss ☐

Hauptschulabschluss
oder vergleichbaren Abschluss, z.B. im Ausland erworben ☐

Mittlere Reife / Realschulabschluss oder vergleichbaren Abschluss ☐

Abitur, Fachhochschulreife od. vergleichbaren Abschluss ☐

22. Haben Sie einen beruflichen Abschluss (z.B. Lehre)?

Ja ☐ Nein ☐

23. Wenn Sie Frage 22 mit Ja beantwortet haben: Welchen beruflichen Abschluss besitzen Sie? ☐

24. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) einen beruflichen Abschluss (z.B. Lehre)?

Ja ☐ Nein ☐

25. Wenn Sie Frage 24 mit Ja beantwortet haben: Welcher Abschluss ist das?

☐

26. Haben Sie studiert?

Ja ☐ Nein ☐

27. Wenn Sie Frage 26 mit Ja beantwortet haben: Welchen (Fach-) Hochschulabschluss besitzen Sie? ☐

28. Hat der andere Elternteil (beziehungsweise die andere erwachsene Bezugsperson des Kindes) studiert?

Ja ☐ Nein ☐

29. Wenn Sie Frage 28 mit Ja beantwortet haben: Welchen Hochschulabschluss besitzt die Person? ☐

30. Welches Bruttoeinkommen (einschließlich Einkommen des im Haushalt lebenden Partners, Sozialhilfe etc.) steht Ihrem Haushalt monatlich zur Verfügung? Kreuzen Sie die relevante Einkommensklasse an!

weniger als 800 €	☐	800-1200 €	☐
1200-1500 €	☐	1500-1800 €	☐
1800-2200 €	☐	2200-2500 €	☐
2500-2800 €	☐	2800-3200 €	☐
3200-3700 €	☐	mehr als 3700 €	☐

31. Wurde im Grundschulunterricht Ihres Kindes, welches in der 5. Klasse ist, ab und zu ein Computer eingesetzt?

Ja ☐ Nein ☐

32 Wie wichtig finden Sie, dass Ihr Kind ab Klasse 5 Computerunterricht erhält?

sehr wichtig ☐, wichtig ☐, weniger wichtig ☐, nicht wichtig ☐

33. Haben Sie selbst einen Computer bei der Hausaufgabenvorbereitung Ihres Kindes, welches in der 5. Klasse ist, eingesetzt?

Ja ☐ Nein ☐

34. Wie viele Stunden verbringt Ihr Kind an einem durchschnittlichen Tag in seiner Freizeit am Computer?

☐ Stunden

35. Welche Note hatte Ihr Kind im letzten Zeugnis in Mathematik? ☐

36. Welche Note hatte Ihr Kind im letzten Zeugnis in Deutsch? ☐

37. Vergeben Sie eine Gesamtnote zwischen 1 und 6, die die Leistung Ihres Kindes in der Grundschule beschreibt: ☐

Vielen Dank für Ihre Teilnahme!

Schulprojekt: Computer in der Schule

Lehrerfragebogen
Klasse 5a, Gruppe 1

A Fragen zur Gruppenstruktur

1. Wie viele Kinder ausländischer Abstammung (d.h. im Ausland geboren oder Kinder, deren Eltern im Ausland geboren sind) sind etwa in dieser Gruppe?

☐ Schüler

2. Wie viele Schüler haben **sprachliche** Probleme, dem Unterricht zu folgen?

☐ Schüler

3. Gibt es in der Gruppe Probleme bezüglich des Verhaltens / des Zusammenhalts der Kinder untereinander oder besteht ein gutes Verhältnis unter den Kindern?

Sehr gutes Verhältnis der Kinder in der Gruppe zueinander..... ☐
Eher gutes Verhältnis der Kinder in der Gruppe zueinander..... ☐
Weder besonders gutes noch besonders schlechtes Verhältnis der Kinder..... ☐
Eher schlechtes Verhältnis der Kinder in der Gruppe zueinander..... ☐
Problematisches Verhältnis der Kinder in der Gruppe zueinander..... ☐

4. Gab es in diesem Schuljahr Fluktuation in der Gruppe, d.h. sind während des Schuljahres Kinder abgegangen oder neu in die Schule gekommen?

Ja ☐ Nein ☐

B Fragen zur Gruppenmotivation

1. Wie motiviert schätzen Sie die Schüler dieser Gruppe durchschnittlich ein? Bitte beziehen Sie sich bei dieser Aussage auf den Computerunterricht.

Sehr motiviert..... ☐
Eher motiviert..... ☐
Weder motiviert noch unmotiviert..... ☐
Eher wenig motiviert..... ☐
Unmotiviert..... ☐

2. Gibt es in dieser Gruppe Schüler, die Sie als Problemschüler bezeichnen würden?

Ja ☐ Nein ☐

3. Wenn Sie Frage 2 mit „Ja" beantwortet haben, auf wie viele Schüler trifft das zu?

☐ Schüler

4. Bitte beschreiben Sie ggf., welche Probleme mit den Schülern bestehen.

5. Gibt es häufige Störungen des Unterrichts?

Ja ☐ Nein ☐

6. Wenn Sie Frage 5 mit „Ja“ beantwortet haben, welche Störungen sind das?

7. Wie viele Schüler würden Sie als **inhaltlich** mit dem Unterrichtsstoff als überfordert bezeichnen?

☐ Schüler

8. Welchen Notendurchschnitt erwarten Sie ungefähr für diese Gruppe in diesem Unterricht am Ende des Schuljahres?

☐

9. Wie häufig müssen die Schüler für den Computerunterricht Hausaufgaben machen?

Jede Woche..... ☐
Alle zwei Wochen..... ☐
Etwa ein- bis zweimal im Monat..... ☐
Weniger als einmal im Monat..... ☐
Nie..... ☐

Zum Schluss habe ich noch einige Fragen zu Ihrer Person:

1. Unterrichten Sie diese Klasse auch in anderen Fächern?

Ja ☐ Nein ☐

2. Wenn Sie Frage mit „Ja“ beantwortet haben, welche Fächer sind das?

3. Welche Fächer unterrichten Sie allgemein?

4. Wie finden Sie die Computerausstattung dieser Schule: Ist diese für die Bedürfnisse des Unterrichts ausreichend?

Ja ☐ Nein ☐

5. Wie gut klappt die Zusammenarbeit mit den Schülerinnen und Schülern in dieser Gruppe allgemein?

Sehr gut..... ☐
Gut..... ☐
Weder besonders gut noch schlecht..... ☐
Eher problematisch..... ☐
Sehr problematisch..... ☐

6. Was ist Ihre Meinung: Welche Kinder sind begabter im Computerunterricht?

Mädchen...☐ Jungen...☐ kann man so nicht sagen...☐

7. Was ist Ihre Meinung: Ist es sinnvoll, Jungen und Mädchen im Computerunterricht getrennt zu unterrichten?

Ja...☐ Nein...☐ weiß ich nicht...☐

8. Warum? (bezieht sich auf die Antwort zu Frage 7)

---

9. Aus Ihrer Berufserfahrung heraus, welchen der folgenden Aussagen würden Sie zustimmen?

Mädchen sind eher sprachlich begabt als Jungen:
stimmt...☐ stimmt nicht...☐

Jungen sind eher begabt für Mathematik oder Naturwissenschaften als Mädchen:
stimmt...☐ stimmt nicht...☐

Unterschiedliche Stärken und Schwächen von Jungen und Mädchen für verschiedene Fächer sind wahrscheinlich eher durch die Erziehung bedingt als natürlich gegeben:
stimmt...☐ stimmt nicht...☐

Unterschiedliche Stärken und Schwächen von Jungen und Mädchen für verschiedene Fächer sind wahrscheinlich sowohl durch die Erziehung als auch durch natürliche Veranlagung bestimmt:
stimmt...☐ stimmt nicht...☐

Jungen sollten eher technische Berufe anstreben als Mädchen:
stimmt...☐ stimmt nicht...☐

Mädchen sind nicht begabt für technische Berufe:

stimmt... ☐ stimmt nicht... ☐

Mädchen haben im Durchschnitt stärkere Probleme, dem Unterrichtsstoff der Hauptschule zu folgen als Jungen:

stimmt... ☐ stimmt nicht... ☐

Jungen haben im Durchschnitt stärkere Probleme, dem Unterrichtsstoff der Hauptschule zu folgen als Mädchen:

stimmt... ☐ stimmt nicht... ☐

10. Wie lange üben Sie diesen Beruf aus?

☐ Jahre

11. Wie alt sind Sie?

☐ Jahre

12. Haben Sie sonstige Anmerkungen zum Unterricht mit dieser Gruppe oder zu dieser Umfrage?

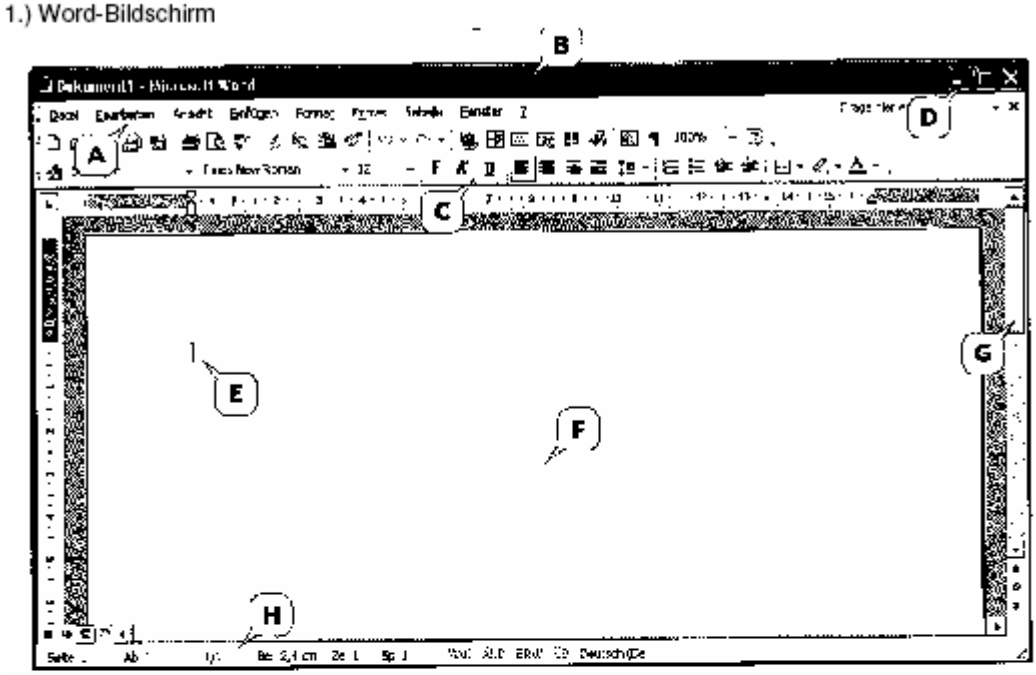
Vielen Dank für Ihre Teilnahme!

Appendix of Tests

The following pages display the original tests (in German language) that were conducted in the first year of the BW-project.

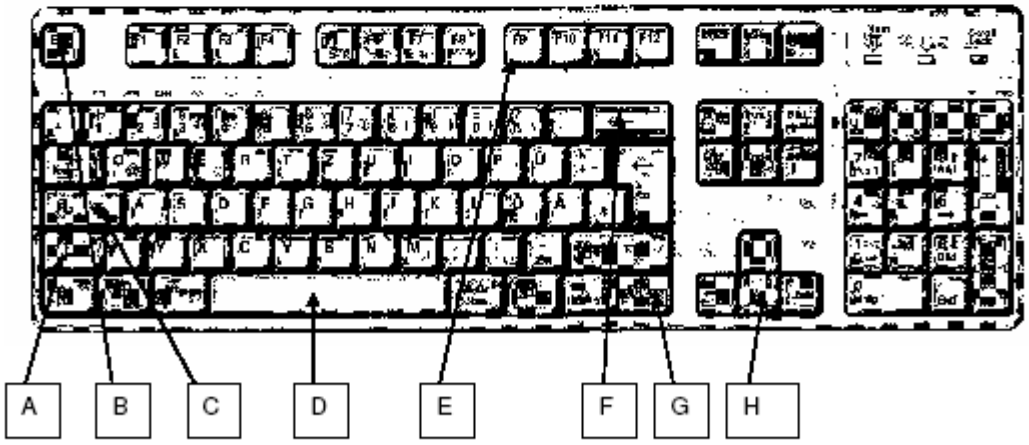
Arbeit Informatik Kl. 5

1.) Word-Bildschirm



Ordne folgende Begriff den Buchstaben zu!
Menüleiste Cursor Titelleiste Statuszeile Symbolleiste Bildlaufleiste Arbeitsfeld
Schaltfläche in der Titelleiste 4P

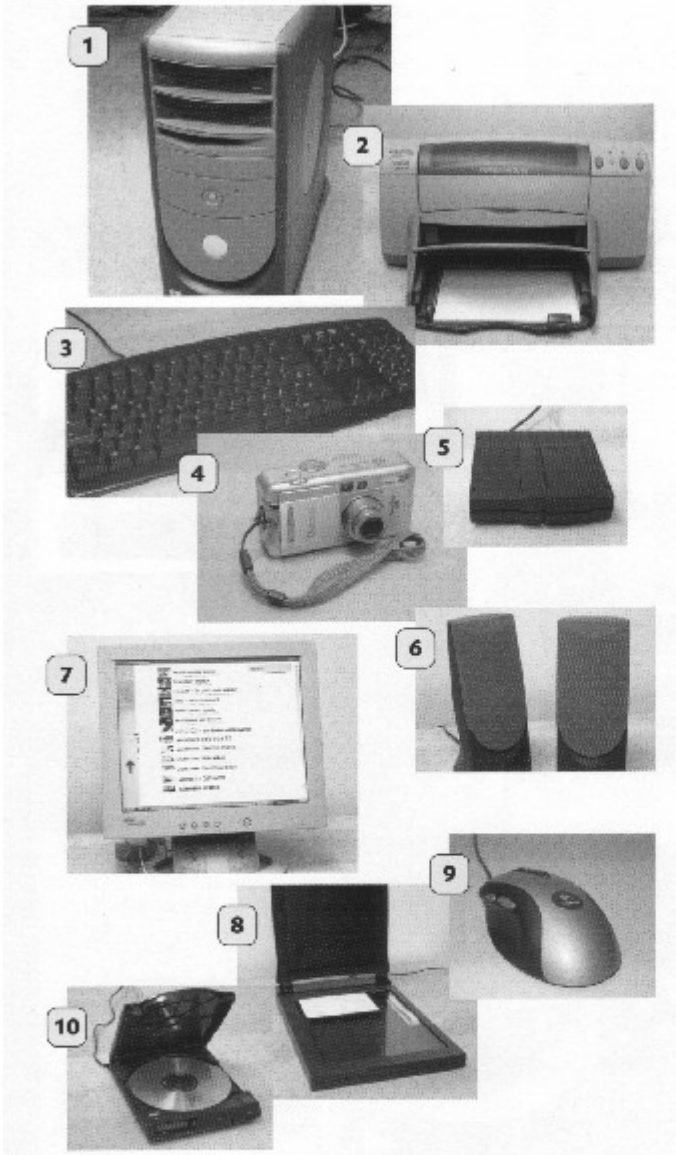
2.) Tastatur



Benenne die Tasten!

A= B=
C= D=
E= F=
G= H= 4P

3.) Teile einer Computer-Anlage



1= _____

2= _____

3= _____

4= _____

5= _____

6= _____

7= _____

8= _____

9= _____

10= _____

5P

gesamt 13P

Informatik-Test Kl.5 Datum:_____

Name:_____ Klasse:_____

1. Beschreibe wie du deine fertigen Dateien speicherst!

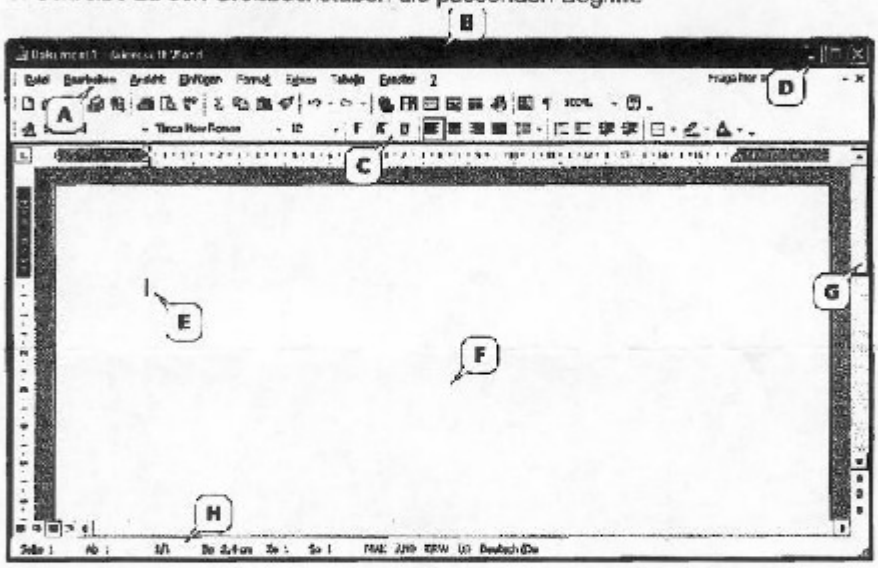
_____

_____

2. Welchen Arbeitsschritt musst du beachten bevor du einen Text ausdruckst?

_____

3. Schreibe zu den Großbuchstaben die passenden Begriffe



A:_____ B:_____

C:_____, D:_____

E:_____, F:_____

G:_____, H:_____

4. Beschreibe in Stichworten, wie das Internet funktioniert!

_____

_____

5. Nenne die Teile, die zu einer Computeranlage gehören!

_____

_____

_____

Appendix on the Students' Background

Table A in this Appendix shows descriptive statistics for major variables describing the general background of the students considered (fifth-graders in 2004/2005). The information is drawn from the students' questionnaires which were distributed in January and June 2005.[30] Generally, girls are a majority in this cohort of lower secondary school students (62 %). The variation in the students' age is rather high, with the youngest students being ten and a half years old and the oldest being more than thirteen years old in January 2005. The average student is eleven and a half years old.

Even if most students were born in Germany (77 %), the majority of them have some immigrant background (in the sense that at least one parent was born abroad or a foreign language is spoken at home). The languages spoken at home, which are reported in the student questionnaires, indicate that most immigrants originate from Russia (42 %) and Turkey (24 %). Fewer are immigrants from the former Yugoslavia (9 %) and Italy (9 %) or Romania (7 %). Three remaining students seem to have an immigrant background from Hungary, Thailand and some Arabian country, respectively.[31]

The average student has one or two siblings. The majority of siblings (36 % of all siblings) are more than three years younger than the considered student. Concerning parental background, parental education is unfortunately missing in many cases: 19 % of the students do not know about their mother's secondary education and 27 % are not able to indicate their father's secondary education. Among those who reported parental education, 39% (18 %) state that their mother's highest secondary degree corresponds to a low (an intermediate) degree, while fewer mothers seem to hold no secondary degree (12 %) or a higher level secondary degree (12 %). Similarly, among the fathers, most reported degrees refer to the lower secondary schools (35 %); only 20 % and 12 % hold intermediate and higher degrees respectively and 5 % seem to hold no secondary degree at all. These numbers indicate that the parents of lower-level high school students generally tend to hold lower (or, at most, intermediate) degrees as well.

Table A includes two variables which can be considered as proxies for 'family care': First of all there is a dummy variable referring to a grand-parent living in the same house or close. Nearly half of the students report that a grand-parent is living nearby. Secondly, paren-

[30] In few cases this information is combined with data from the parents' questionnaires because the reported answers were not clear or did not seem reliable.
[31] This information is confirmed by information from the parents' questionnaires.

tal interest is proxied assuming that parents are at least to some extent interested in the child's school performance if they opted to complete the voluntary parents' questionnaire: 66 % of the children seem to have a mother or father who cares about school matters and completed the questionnaire.

Table A: Descriptive statistics for major background characteristics

Variable	Means	(s.d.)	Min	Max
Gender indicator (male = 1)	0.38	(0.49)	0.00	1.00
Age of students (in January 2005)	11.53	(0.61)	10.52	13.13
Immigrant background (Reference: no such background):				
Student: born in Germany	0.77	(0.42)	0.00	1.00
Both parents born in Germany	0.30	(0.46)	0.00	1.00
Only German language used at home	0.35	(0.48)	0.00	1.00
Some immigrant background	0.78	(0.41)	0.00	1.00
Number of siblings …				
…absolute	1.54	(1.06)	0.00	5.00
…more than 3 years younger	0.55	(0.83)	0.00	4.00
…3 or less than 3 years younger	0.34	(0.56)	0.00	2.00
…older, but at most 3 years	0.19	(0.39)	0.00	1.00
…more than 3 years older	0.36	(0.61)	0.00	2.00
Parental education dummies				
Mother's education: missing value	0.19	(0.39)	0.00	1.00
Mother's education: no secondary degree	0.12	(0.33)	0.00	1.00
Mother's education: lower secondary degree	0.39	(0.49)	0.00	1.00
Mother's education: intermediate degree	0.18	(0.38)	0.00	1.00
Mother's education: higher secondary degree	0.12	(0.33)	0.00	1.00
Father's education: missing value	0.27	(0.45)	0.00	1.00
Father's education: no secondary degree	0.05	(0.23)	0.00	1.00
Father's education: lower secondary degree	0.35	(0.48)	0.00	1.00
Father's education: intermediate degree	0.20	(0.40)	0.00	1.00
Father's education: higher secondary degree	0.12	(0.33)	0.00	1.00
Family Care:				
Grand-parents live nearby (dummy variable)	0.46	(0.50)	0.00	1.00
Parental interest (participation in project dummy)	0.66	(0.48)	0.00	1.00
Number of observations		74		

Note: The calculations only include students observed throughout the year since most background characteristics are only available for these students.
Source: Student questionnaires. Own calculations.

Table B of this Appendix reports means of the considered background variables separately by gender. Most variables take a similar value for boys and girls. However, it is obvious that girls more often fail to report their fathers' secondary degree (30 %) than boys (21 %), while boys more often state that they do not know about their mothers' secondary education (25 %) than girls (15 %). In addition, the further parental education categories seem to differ according to the gender of the student (for those who reported on parental education). For example, 20 % of the girls declare that their father holds a higher secondary degree, while none of the boys' fathers seems to hold such a degree. It is hard to tell if this is due to an incorrect assignment of degrees by the children or to real differences in the parents' education.

Table B: Descriptive statistics for major background characteristics by gender

Variable / Group	Girls		Boys	
	Means	(s.d.)	Means	(s.d.)
Age of students (in January 2005)	11.44	0.58	11.67	0.64
Immigrant background (Reference: no such background):				
Student: born in Germany	0.74	0.44	0.82	0.39
Both parents born in Germany	0.33	0.47	0.25	0.44
Only German language used at home	0.37	0.49	0.32	0.48
Some immigrant background*	0.76	0.43	0.82	0.39
Number of siblings …				
…absolute	1.65	1.10	1.36	0.99
…more than 3 years younger	0.57	0.83	0.54	0.84
…3 or less than 3 years younger	0.35	0.57	0.32	0.55
…older, but at most 3 years	0.22	0.42	0.14	0.36
…more than 3 years older	0.43	0.69	0.25	0.44
Parental education dummies				
Mother's education: missing value	0.15	0.36	0.25	0.44
Mother's education: no secondary degree	0.17	0.38	0.04	0.19
Mother's education: lower secondary degree	0.39	0.49	0.39	0.50
Mother's education: intermediate degree	0.13	0.34	0.25	0.44
Mother's education: higher secondary degree	0.15	0.36	0.07	0.26
Father's education: missing value	0.30	0.47	0.21	0.42
Father's education: no secondary degree	0.04	0.21	0.07	0.26
Father's education: lower secondary degree	0.30	0.47	0.43	0.50
Father's education: intermediate degree	0.15	0.36	0.29	0.46
Father's education: higher secondary degree	0.20	0.40	0.00	0.00
Family Care:				
Grandparents live nearby (dummy variable)	0.48	0.51	0.43	0.50
Parental interest (participation in project dummy)	0.65	0.48	0.68	0.48
Number of observations	46		28	

Note: The calculations only include students observed throughout the year since most background characteristics are only available for these students.
Source: Student questionnaires. Own calculations.

Furthermore, information given in the parents' questionnaire is used in order to learn more about the students' socio-economic background. However, since only two thirds of the parents completed the questionnaires, it may be that the impression given by the parents' information is not representative for all students' parents.

Table C includes information related to the parents' socio-economic background. Most of the questionnaires (two-thirds) were filled in by the students' mothers. In addition, girls' parents more often participate in the survey (61 %) than boys' parents, while the participation rate of fathers is higher for girls (40 %) than for boys (21 %).

Since the sample of responding parents is a selected group one can not directly compare the given information to the children's general information. For example, compared to the children's statements there are fewer persons with an immigrant background in this group.

Table C: Parental information on socio-economic background

	All parents			Girls' parents			Boys' parents		
	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)
Respondent of questionnaire…									
…female parent	49	0.65	(0.48)	30	0.60	(0.50)	19	0.74	(0.45)
…male parent	49	0.33	(0.47)	30	0.40	(0.50)	19	0.21	(0.42)
…both parents	49	0.02	(0.14)	30	0.00	(0.00)	19	0.05	(0.23)
Family situation:									
Single parent	49	0.31	(0.47)	30	0.33	(0.48)	19	0.26	(0.45)
Number of children*	49	2.24	(1.11)	30	2.40	(1.25)	19	2.00	(0.82)
Grandparents live nearby	49	0.39	(0.49)	30	0.40	(0.50)	19	0.37	(0.50)
Immigration background:									
Female parent native born	48	0.44	(0.50)	30	0.43	(0.50)	18	0.44	(0.51)
Male parent native born	48	0.38	(0.49)	30	0.43	(0.50)	18	0.28	(0.46)
Foreign language at home	49	0.53	(0.50)	30	0.50	(0.51)	19	0.58	(0.51)
Mothers' secondary education:									
…no secondary degree	45	0.13	(0.34)	16	0.06	(0.25)	29	0.17	(0.38)
…lower secondary degree	45	0.51	(0.51)	16	0.56	(0.51)	29	0.48	(0.51)
…intermediate degree	45	0.24	(0.43)	16	0.31	(0.48)	29	0.21	(0.41)
…higher secondary	45	0.11	(0.32)	16	0.06	(0.25)	29	0.14	(0.35)
Fathers' secondary education:									
…no secondary degree	40	0.08	(0.27)	16	0.13	(0.34)	24	0.04	(0.20)
…lower secondary degree	40	0.53	(0.51)	16	0.44	(0.51)	24	0.58	(0.50)
…intermediate degree	40	0.30	(0.46)	16	0.44	(0.51)	24	0.21	(0.41)
…higher secondary	40	0.10	(0.30)	16	0.00	(0.00)	24	0.17	(0.38)
Tertiary degree dummies (reference: no tertiary degree):									
Mother: vocational degree	47	0.57	(0.50)	30	0.53	(0.51)	17	0.65	(0.49)
Father: vocational degree	42	0.69	(0.47)	25	0.68	(0.48)	17	0.71	(0.47)
Mother: higher tertiary	46	0.07	(0.25)	29	0.07	(0.26)	17	0.06	(0.24)
Father: higher tertiary	43	0.05	(0.21)	26	0.08	(0.27)	17	0.00	(0.00)
Mothers' employment status:									
…not employed	47	0.34	(0.48)	29	0.41	(0.50)	18	0.22	(0.43)
…minor employment	47	0.15	(0.36)	29	0.14	(0.35)	18	0.17	(0.38)
…part-time	47	0.23	(0.43)	29	0.14	(0.35)	18	0.39	(0.50)
…full-time	47	0.28	(0.45)	29	0.31	(0.47)	18	0.22	(0.43)
Fathers' employment status:									
…not employed	40	0.15	(0.36)	25	0.16	(0.37)	15	0.13	(0.35)
…minor employment	40	0.03	(0.16)	25	0.04	(0.20)	15	0.00	(0.00)
…part-time	40	0.10	(0.30)	25	0.12	(0.33)	15	0.07	(0.26)
…full-time	40	0.73	(0.45)	25	0.68	(0.48)	15	0.80	(0.41)
*Computer use at work dummies (reference: person does not use computer)***									
Computer use of mother	46	0.24	(0.43)	28	0.18	(0.39)	18	0.33	(0.49)
Computer use of father	44	0.34	(0.48)	26	0.27	(0.45)	18	0.44	(0.51)
*Household income (classified information)****									
Household income***	42	4.76	(2.36)	28	4.82	(2.60)	14	4.64	(1.86)

Note: The calculations only include parents of students observed throughout the year since most background characteristics are only available for these students. *Number of children living in the household. **Among the parents using computers for work 37.5 % indicate to use it for standard software or the internet and 62.5% use special application software. ***Gross household income in categories. If I interpolate between categories, the mean of 4.76 corresponds to EUR 1,730.
Source: Parents' questionnaires. Own calculations.

However, the information confirms the impression of intergenerational immobility: Most of the lower secondary students' responding parents have obtained a lower secondary degree as

well. While 57 % of the responding mothers and 69 % of the fathers hold a vocational degree (mostly on the apprenticeship level), there are only few exceptions in which a parent holds a higher tertiary degree (three out of 46 mothers and two out of 43 fathers). The relatively low education of parents goes hand-in-hand with low full-time employment rates (only 28 % of mothers and 73 % of fathers in the available sample are employed) and a low median household income (EUR 1,730)[32].

Table D: Children's computer use and performance

	All parents			Girls' parents			Boys' parents		
	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)	Obs	Mean	(s.d.)
Do you think it is important that your child attends computer lessons in fifth grade?*	48	1.21	0.46	30	1.30	0.53	18	1.06	0.24
Did you use a computer for the preparation of the child's homework at home (dummy)?	49	0.55	0.54	30	0.50	0.57	19	0.63	0.50
How many hours a day does your child spend on a computer during leisure time?	40	1.19	0.94	24	0.89	0.72	16	1.63	1.08
Did your child's elementary school teacher use a computer to teach (dummy variable)?	49	0.47	0.50	30	0.43	0.50	19	0.53	0.51
What grade did your child have in math last year?*	48	3.42	0.74	30	3.57	0.68	18	3.17	0.79
What grade did your child have in German last year?*	48	3.49	0.72	30	3.52	0.72	18	3.44	0.73
Which overall grade describes your child's performance in elementary school?*	74	2.70	0.77	46	2.76	0.57	28	2.61	1.03

Note: The calculations only include parents of students observed throughout the year since most background characteristics are only available for these students. * Grades refer to the German grading system, where the best grade is 1 (= excellent) and the worst grade is 6 (= failed).
Source: Parents' questionnaires. Own calculations.

Table D shows additional parental information on the children's skills and computer use. Nearly half of the responding parents (48 %) support the view that it is important to teach elementary computer skills to fifth graders. However, it might be that the survey respondence rate is higher for those who also support computer studies and this number might thus be upwardly biased. The same is true for the number of parents using a computer to help the student prepare his homework (55 %). Additionally, the average respondent's child spend about one hour of her daily leisure time on the computer. Further information from the parents' questionnaires relates to the computer use at elementary school and elementary school grades. Nearly half of the students' have experienced some computer based elementary school lessons. The overall elementary school performance of the students compared is rather low, which is consistent with the fact that these students attend the lower-level secondary schools.

[32] Household income is an interpolated value from reported income categories (cf. questionnaires). The low value suggests that parents mostly include labour income only.

Appendix of Tables of Group Characteristics

Tables A: Contingency tables

Birth country / Group	Co-educational	Mono-educational	Total
Foreign born	12	5	17
German-born	39	18	57
Total	51	23	74

Parents' birth countries / Group	Co-educational	Mono-educational	Total
At least one parent foreign born	35	17	52
Both parents German-born	16	6	22
Total	51	23	74

Language at home / Group	Co-educational	Mono-educational	Total
Foreign language	34	14	48
Only German	17	9	26
Total	51	23	74

Immigrant background / Group|	Co-educational	Mono-educational	Total
No immigrant background	12	4	16
Some immigrant background	39	19	58
Total	51	23	74

Grand-parent(s) / Group	Co-educational	Mono-educational	Total
Do(es) not live nearby	28	12	40
Live(s) nearby	23	11	34
Total	51	23	74

Parents' education / Group	Co-educational	Mono-educational	Total
Both less than Realschule	22	10	32
At least one parent Realschule or Abitur	21	9	30
Total	43	19	62

Parents' interest / Group	Co-educational	Mono-educational	Total
Show no interest in school	17	8	25
Show interest in school	34	15	49
Total	51	23	74

Source: Student and parents questionnaires. Own calculations.

Table B: Means of observed variables by group

Groups	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Variables	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
Age of student (in January 2005)	11.42 (0.43)	11.50 (0.60)	11.53 (0.71)	11.64 (0.61)	11.55 (0.74)	11.50 (0.57)
Immigrant background (Reference: no such immigrant background):						
Student: Born in Germany	0.77 (0.44)	0.58 (0.51)	0.75 (0.45)	0.93 (0.27)	0.81 (0.40)	0.71 (0.49)
Both parents born in Germany	0.31 (0.48)	0.33 (0.49)	0.08 (0.29)	0.50 (0.52)	0.31 (0.48)	0.14 (0.38)
Only German language used at home	0.31 (0.48)	0.33 (0.49)	0.17 (0.39)	0.50 (0.52)	0.38 (0.50)	0.43 (0.53)
Some immigrant background	0.69 (0.48)	0.83 (0.39)	0.92 (0.29)	0.64 (0.50)	0.81 (0.40)	0.86 (0.38)
Number of siblings …						
…absolute	1.54 (1.33)	1.42 (1.00)	2.08 (0.90)	1.43 (1.28)	1.38 (0.81)	1.43 (0.98)
…more than 3 years younger	0.54 (0.88)	0.58 (0.90)	0.83 (1.11)	0.50 (0.76)	0.44 (0.63)	0.43 (0.79)
… 3 or less than 3 years younger	0.23 (0.60)	0.17 (0.39)	0.42 (0.51)	0.50 (0.76)	0.31 (0.48)	0.43 (0.53)
… older, but at most 3 years	0.31 (0.48)	0.08 (0.29)	0.25 (0.45)	0.21 (0.43)	0.19 (0.40)	0.00 (0.00)
… more than 3 years older	0.31 (0.48)	0.50 (0.80)	0.33 (0.65)	0.29 (0.47)	0.38 (0.72)	0.43 (0.53)
Parental education (Reference: higher secondary degree of mother / father):						
Mother's education: missing value	0.23 (0.44)	0.50 (0.52)	0.00 (0.00)	0.07 (0.27)	0.25 (0.45)	0.00 (0.00)
Mother's education: no secondary degree	0.08 (0.28)	0.08 (0.29)	0.17 (0.39)	0.14 (0.36)	0.19 (0.40)	0.00 (0.00)
Mother's education: lower secondary degree	0.23 (0.44)	0.25 (0.45)	0.50 (0.52)	0.50 (0.52)	0.38 (0.50)	0.57 (0.53)
Mother's education: intermediate degree	0.31 (0.48)	0.08 (0.29)	0.17 (0.39)	0.07 (0.27)	0.13 (0.34)	0.43 (0.53)
Father's education: missing value	0.23 (0.44)	0.58 (0.51)	0.17 (0.39)	0.07 (0.27)	0.44 (0.51)	0.00 (0.00)
Father's education: no secondary degree	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.14 (0.36)	0.13 (0.34)	0.00 (0.00)
Father's education: lower secondary degree	0.23 (0.44)	0.25 (0.45)	0.50 (0.52)	0.50 (0.52)	0.19 (0.40)	0.57 (0.53)
Father's education: intermediate degree	0.38 (0.51)	0.17 (0.39)	0.17 (0.39)	0.07 (0.27)	0.13 (0.34)	0.43 (0.53)
Family Care:						
Grand-parents live nearby (dummy)	0.38 (0.51)	0.42 (0.51)	0.50 (0.52)	0.50 (0.52)	0.63 (0.50)	0.14 (0.38)
Parental interest (participation in project dummy)	0.85 (0.38)	0.58 (0.51)	0.83 (0.39)	0.43 (0.51)	0.56 (0.51)	0.86 (0.38)
Performance (in other subjects):*						
Child's overall performance (self reported)	2.92 (0.76)	2.50 (0.90)	2.75 (0.45)	2.71 (0.83)	2.69 (0.60)	2.57 (1.27)
Child's math performance (self reported)	2.69 (0.85)	2.75 (0.97)	2.67 (0.65)	2.64 (0.93)	3.19 (0.83)	2.71 (1.25)
Child's German performance (self reported)	3.08 (0.64)	2.67 (0.78)	3.42 (0.79)	3.07 (1.14)	2.50 (0.89)	3.14 (0.90)
Number of observations	13	12	12	14	16	7

Note: The calculations only include students observed throughout the year. Grey-shaded lines refer to the groups taught by the same teachers. * Self reported performance scales range from 1 (= much better than average) to 5 (much worse than average).

Source: Student and parents questionnaires. Own calculations.

Table C: Means of observed variables by group for females

Groups	Group 1	Group 2	Group 3	Group 4	Group 5
Variables	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
Age of student (in January 2005)	11.35 (0.34)	11.64 (0.65)	11.38 (0.48)	11.20 (0.46)	11.55 (0.74)
Immigrant background (Reference: no such immigrant background):					
Student: Born in Germany	0.88 (0.35)	0.33 (0.52)	0.67 (0.50)	0.86 (0.38)	0.81 (0.40)
Both parents born in Germany	0.50 (0.53)	0.33 (0.52)	0.11 (0.33)	0.43 (0.53)	0.31 (0.48)
Only German language used at home	0.50 (0.53)	0.33 (0.52)	0.11 (0.33)	0.57 (0.53)	0.38 (0.50)
Some immigrant background	0.50 (0.53)	0.83 (0.41)	0.89 (0.33)	0.71 (0.49)	0.81 (0.40)
Number of siblings …					
…absolute	1.75 (1.67)	1.67 (0.82)	2.22 (0.97)	1.43 (1.27)	1.38 (0.81)
…more than three years younger	0.50 (1.07)	0.50 (0.55)	1.00 (1.22)	0.43 (0.53)	0.44 (0.63)
… younger, but less than three years	0.25 (0.71)	0.17 (0.41)	0.44 (0.53)	0.57 (0.79)	0.31 (0.48)
… older, but less than three years	0.38 (0.52)	0.00 (0.00)	0.22 (0.44)	0.29 (0.49)	0.19 (0.40)
… more than three years older	0.38 (0.52)	0.83 (0.98)	0.33 (0.71)	0.43 (0.53)	0.38 (0.72)
Parental education (Reference: higher secondary degree of mother / father):					
Mother's education: missing value	0.13 (0.35)	0.33 (0.52)	0.00 (0.00)	0.00 (0.00)	0.25 (0.45)
Mother's education: no secondary degree	0.13 (0.35)	0.17 (0.41)	0.22 (0.44)	0.14 (0.38)	0.19 (0.40)
Mother's education: lower secondary degree	0.25 (0.46)	0.33 (0.52)	0.56 (0.53)	0.43 (0.53)	0.38 (0.50)
Mother's education: intermediate degree	0.38 (0.52)	0.00 (0.00)	0.11 (0.33)	0.00 (0.00)	0.13 (0.34)
Father's education: missing value	0.25 (0.46)	0.50 (0.55)	0.11 (0.33)	0.14 (0.38)	0.44 (0.51)
Father's education: no secondary degree	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.13 (0.34)
Father's education: lower secondary degree	0.25 (0.46)	0.33 (0.52)	0.56 (0.53)	0.29 (0.49)	0.19 (0.40)
Father's education: intermediate degree	0.25 (0.46)	0.17 (0.41)	0.11 (0.33)	0.14 (0.38)	0.13 (0.34)
Family Care:					
Grand-parents live nearby (dummy)	0.88 (0.35)	0.50 (0.55)	0.89 (0.33)	0.43 (0.53)	0.56 (0.51)
Parental interest (participation in project dummy)	0.50 (0.53)	0.33 (0.52)	0.33 (0.50)	0.43 (0.53)	0.63 (0.50)
*Performance (in other subjects):**					
Child's overall performance (self reported)	3.00 (0.53)	3.00 (0.63)	2.78 (0.44)	2.43 (0.53)	2.69 (0.60)
Child's math performance (self reported)	3.00 (0.53)	3.33 (0.82)	2.78 (0.67)	2.29 (0.76)	3.19 (0.83)
Child's German performance (self reported)	3.13 (0.64)	3.00 (0.63)	3.11 (0.60)	3.14 (1.35)	2.50 (0.89)
Number of observations	8	6	9	7	16

Note: These calculations only include students observed throughout the year. Grey-shaded lines refer to the groups taught by the same teachers. * Self reported performance scales range from 1 (= much better than average) to 5 (much worse than average).
Source: Student and parents questionnaires. Own calculations.

Table D: Performance in general subjects by group type

Group Type		Mixed-gender groups						Single-gender groups					
Sample	All	All		Girls		Boys		All		Girls		Boys	
Variable	K-W. $\chi^2_{(1)}$	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)	N	Mean (s.d.)
Performance measured at mid-term:													
Comparative performance	0.09	51	2.73 (0.75)	30	2.80 (0.55)	21	2.62 (0.97)	23	2.65 (0.83)	16	2.69 (0.60)	7	2.57 (1.27)
Com. math performance	1.73	51	2.69 (0.84)	30	2.83 (0.75)	21	2.48 (0.93)	23	3.04 (0.98)	16	3.19 (0.83)	7	2.71 (1.25)
Com. German performance	1.65	51	3.06 (0.88)	30	3.10 (0.80)	21	3.00 (1.00)	23	2.70 (0.93)	16	2.50 (0.89)	7	3.14 (0.90)
Expected average grade	0.80	45	2.85 (0.73)	25	2.77 (0.68)	20	2.95 (0.79)	22	2.61 (0.91)	15	2.56 (1.01)	7	2.71 (0.70)
Performance measured at end of term:													
Comparative performance	1.18	51	2.88 (0.52)	30	2.87 (0.57)	21	2.90 (0.44)	22	2.73 (0.77)	15	2.93 (0.70)	7	2.29 (0.76)
Com. math performance	0.15	51	2.89 (0.73)	30	2.97 (0.61)	21	2.79 (0.87)	23	2.74 (0.92)	16	3.00 (0.82)	7	2.14 (0.90)
Expected math grade	3.17*	50	3.16 (0.83)	30	3.25 (0.74)	20	3.01 (0.94)	23	2.83 (1.03)	16	3.07 (1.09)	7	2.29 (0.64)
Com. German performance	4.57**	51	3.21 (0.78)	30	3.23 (0.68)	21	3.17 (0.91)	23	2.78 (0.60)	16	2.75 (0.58)	7	2.86 (0.69)
Expected German grade	4.80**	51	3.42 (0.75)	30	3.35 (0.71)	21	3.51 (0.82)	23	2.98 (0.78)	16	2.88 (0.67)	7	3.21 (0.99)
Change com. performance	0.36	51	0.16 (0.67)	30	0.07 (0.52)	21	0.29 (0.85)	22	0.05 (0.79)	15	0.20 (0.77)	7	-0.29 (0.76)
Performance change:[+]													
Change math performance	1.76	51	0.21 (0.99)	30	0.13 (0.82)	21	0.31 (1.21)	23	-0.30 (1.15)	16	-0.19 (1.17)	7	-0.57 (1.13)
Change German performance	0.00	51	0.15 (0.98)	30	0.13 (0.68)	21	0.17 (1.32)	23	0.09 (0.79)	16	0.25 (0.68)	7	-0.29 (0.95)

Note: The Kruskal-Wallis tests refer to the comparisons of the respective distributions for students in single-gender and mixed groups. ** Significant at the five percent level. * Significant at the 10 percent level. [+] The change variables indicate the difference in the performance measure between the end of term and mid-term. Numbers of observations vary due to missing information for some students.
Source: Student questionnaires, test results. Own calculations.

Appendix to the Literature Review

Table A: Previous research on single-gender computer or science education in Germany

Study	Major Research Question	Data and Method	Main Conclusions	Main Problem
		Type 1: correlation studies*		
Funken *et al.* (1996)	Is coeducation less favourable for girls' **computer**-related performance (computer use and interest)?	Written questionnaires: 1,128 *Gymnasium* students (11th graders) in the school year 1989/90 in Nordrhein-Westfalen (55 % female; 19 % in all girls' schools, no evidence on boys in single gender schools), analysed by descriptive statistics (mainly Chi-squared tests).	More favourable performance-related outcomes for girls from single gender schools (compared to girls from mixed gender schools); similar outcomes of girls in single gender schools and boys in mixed schools.	Selection to single gender schools.
Meyer (1996)	Do students prefer to be taught in single gender **computer** classes?	Project in the city of *Bremen* in the mid-1990s: 14 boys and 15 girls taught in two single gender computer science classes for half a year in ninth grade of the higher secondary school. Assessment by oral interviews. Qualitative presentation of results.	Boys generally prefer coeducation; mixed statements among girls.	Small size of study. Effects are not assessed in a systematic way (*e.g.* using standardized questionnaires).
Nyssen (1996)	Which factors improve girls' confidence, (professional) interest and performance in selected subjects (with a focus on practical **science classes**)? What difference does single gender education make?	Comprehensive school classes in the state of Nordrhein-Westphalen. Longitudinal observation (questionnaires, interviews, video assessment) of 109 students (50 girls, 59 boys) in six groups (2 all girls' groups taught by female teacher, 2 all boys' groups and 2 co-educational groups taught by male teacher, observed in 5th grade (1991/92) and in 7th grade (1993/94). Qualitative presentation of results.	Results are mainly interpreted in favour for single gender education (especially as concerns girls' performance).	Different treatment measures (gender composition of groups, different teachers and instruction methods) are used simultaneously. Especially, instruction of different groups by different teachers reduces comparability of group outcomes. Selection to groups.

Volmerg *et al.* (1996)	Which factors improve girls' interest and participation in **computer** science? What is the role of coeducation?	Observation (oral interviews) of five computer classes in two high schools (*Gymnasium*) in Bremen during 1989/90-1991/92: 3 female single-gender classes (first year: 13 girls in 11th grade followed until 13th grade; second year: 11 girls in 11th grade, observed until 12th grade, third year 15 girls in 11th grade, observed until 12th grade) and 2 mixed classes (10 and 14 students). Qualitative presentation of results.	Mixed findings; introduction of all girls' computer classes increased (initial) participation of girls in this subject	Comparability of different groups is not assured (taught by different teachers; different contents of lessons). Girls anticipate *expected* outcomes of the intervention study and adjust behaviour.
Type 2: non–experimental evaluation studies*				
Rost and Pruisken (2000)	What is the impact of single gender education on psychological and social outcomes (including students' self-assessment in **mathematics** and biology) when similar students in single and mixed gender classes are compared?	Questionnaire-based assessment of performance in Catholic private higher secondary schools in 1997: 649 fifth and sixth graders in three single and mixed gender schools (161 girls in an all girls' school, 243 girls in single gender classes in two mixed gender schools, 154 girls in mixed gender classes in the two mixed gender schools, 91 boys in mixed gender classes in the two mixed gender schools). Evaluated by comparison of mean outcomes.	No significant impact of the regime.	Effect is only identified for the very specific (selective) group of Catholic private school students. Selection into different school types (single *vs.* mixed gender).
Häußler and Hoffmann (2002, 1998, 1990) and Hoffmann *et al.* (1997)	Which factors (curricular changes, teacher behaviour, class size and single gender *vs.* coeducation) improve girls' interest, self-concept and achievement in **science** (physics) classes?	Sample of 150 girls and 139 boys in 12 experimental classes from six schools and taught by six different teachers, 103 girls and 64 boys in 7 control classes from two schools and taught by six different teachers. All students are seventh graders in the higher level secondary school (*Gymnasium*) in the state of Schleswig-Holstein in 1992/93. Assessment by several standardized written tests and questionnaires. (Regression adjusted) group differences (and differences-in-differences) are calculated.	Positive impact of single gender education especially on girls' outcomes.	Treatment and control groups are located at different schools. No extensive information on selection into different groups. Regression analysis in Hoffmann *et al.* (1997) may suffer from endogeneity of some of the regressors (*e.g.* self-concept).

		Type 3: natural experiments *		
Baumert (1992)	How does girls' performance and interest in **mathematics** (and other subjects) differ between coexisting mixed and single gender education schools?	Based on a representative survey of German seventh graders in 1968/1969 including standardized German, English and mathematics outcomes: Girls and boys in mixed and single gender schools (each of the four groups includes 2,100-2,900 observations). Compares seventh graders in single gender and mixed higher secondary schools in the time of a regime switch in the organizational school types through analysis of variance.	Evidence for selection into the differently organized school types in regions where both types coexist. Both genders perform significantly better in mathematics in single gender schools than in coeducational schools. Girls' interest in mathematics decreases notably in seventh grade in mixed gender classes.	Natural experiment is not really used to evaluate the causal effect of coeducation (no difference-in-difference approach; no before-after-comparisons).
		Type 4: social experiments *		
Kessels (2002)	Are girls more motivated and self-confident if they are taught in single gender **science** (physics) classes?	Core Sample: 270 compehensive school students (eighth graders) from four schools in *Berlin* (87 girls and 62 boys in coeducational classes; 56 girls and 65 boys in single gender classes). Random assignment into single gender and mixed groups. Each participating teacher instructs at least one single gender and one coeducational class. Assessment of sociodemographic variables and outcomes by standardized questionnaires, evaluated by analysis of variance.	Positive impact of single gender education on girls' motivation and self-confidence.	No information is given on background characteristics of the different groups. There seems to be significant sample attrition.

Note: * The four categorized types of studies are: 1) correlation studies which do not solve the evaluation problem, 2) non-experimental evaluation studies controlling for background characteristics, 3) natural experiments implying exogenous policy changes and 4) social experiments where students are randomly assigned to treatment and control groups.