Scandinavian Journal of Educational Research
Vol. 55, No. 2, April 2011, 121–144

Routledge
Taylor & Francis Group

The Effect of Mixed-Age Classes in Sweden

Elly-Ann Lindström and Erica Lindahl
Uppsala University and The Institute for Labour Market Policy Evaluation

Mixed-aged (MA) classes are a common phenomenon around the world. In Sweden, these types of classes increased rapidly during the 1980s and 1990s, despite the fact that existing empirical support for MA classes is weak. In this paper, the effect of attending an MA class during grades 4–6 on students' cognitive skills is estimated. Using a unique survey with information on students, parents, and teachers, it is possible to control for many factors that could otherwise bias the results. A negative effect on short-run cognitive skills, as measured by grade 6 cognitive tests, was found. This effect is relatively large—almost 5 percentile points—and robust to a rigorous sensitivity analysis. On grade 9 credits the effect is still negative but smaller in size and not statistically significant.

Keywords: education, mixed-age classes, multi-grade classes, cognitive skills

Mixed-age (MA) classes are a common phenomenon in schools both in Sweden and in other countries. In MA classes, students from different grades are mixed into one class for two major reasons: either out of demographic and economic necessity, or because it is believed that these classes have pedagogical benefits. For example, it is argued that students of different age and school experience interact and learn from each other. This belief contributed to the rapid increase of MA classes in Swedish schools during the 1980s and 1990s. However, the scientific evidence on the effects of MA grouping is ambiguous. Among the Swedish studies used to motivate the introduction of MA classes there is, to our knowledge, no study using representative samples. Many studies are simply questionnaires collected among teachers in MA classes (Andrae Thelin, 1991; Edlund & Sundell, 1999). International studies are available, but they are of varying quality (with very few studies using representative samples) and yield contradictory results. However, most studies conclude that the effect, if any, is small in magnitude.

From an economic point of view, investigating the effect of MA classes is important since it may be one possible way towards greater cost-efficiency within schools. If it is the

Elly-Ann Lindström, Uppsala University and The Institute for Labour Market Policy Evaluation; Erica Lindahl, Uppsala University and The Institute for Labour Market Policy Evaluation.

Correspondence concerning this article should be addressed to Elly-Ann Lindström, Uppsala University and Institute for Labour Market Policy Evaluation, Box 513, 751 20 Uppsala, Sweden. E-mail: elly-ann.lindstrom@ifau.uu.se

We are grateful to Peter Fredriksson and Per Johansson for valuable guidance. We would also like to thank Mikael Elinder, Patrik Hesselius, Jenny Nykvist, Peter Skogman Thoursie, Andreas Westermark, Nabanita Datta Gupta, Mårten Palme, and seminar participants at the Department of Economics, Uppsala University, and anonymous referees, for valuable suggestions and comments. Åsa Arnell is acknowledged for the research idea. The financial support from the Swedish Council for Working Life and Social Research, FAS (dnr 2004-1222) is also acknowledged.

ISSN 0031-3831 print/ISSN 1470-1170 online

DOI: 10.1080/00313831.2011.554692
http://www.informaworld.com

case that MA classes, as is often claimed, are a less expensive way to organize students than traditional classes, and if the students in these classes perform equally well or better than students in traditional classes, introducing MA classes on a larger scale would be an efficient way towards reduced costs and/or increased student performance. This is particularly interesting in relation to one of the most debated policies in the economic and educational literature during recent years, namely reducing class-size. In contrast to class-size reductions, introducing MA would imply practically no extra costs.[1]

Examining the effect of MA classes also sheds light on the question of how knowledge is produced. Economic research has mainly focused on quantitative aspects of education—if, and how much, resources matter for student achievement. But equally important are more qualitative aspects of the educational production function, and the effect of MA classes is one such aspect.

The purpose of this paper is to estimate the effect of MA classes in Sweden on students' cognitive skills. We focus both on short-term effects on grade 6 cognitive tests and on long-run effects on grade 9 credits. We also allow the effect of attending an MA-class to vary between different groups of students considered potentially important: girls, low performing students, and students with a non-Swedish background.

The analysis is based on a rich and representative data set. In addition to register data on important socio-economic variables, we have access to a unique survey with information on parents and teachers and their attitudes towards school-related issues. These data allow us to control for many potential selection problems and to perform a rigorous sensitivity analysis.

The results show a negative effect of attending an MA class in grades 4–6 on the grade 6 cognitive tests. The effect is large—almost 5 percentile points—and robust to a rigorous sensitivity analysis. Also the effect on grade 9 credits is negative, but smaller in magnitude (around 1 percentile point) and not statistically significant. This pattern is consistent with catch-up effects (in the long run, children from MA classes may catch up with children from traditional classes), but it could also be the case that grades and cognitive tests measure slightly different types of skills.

Background

In this section we discuss the concept of MA and the prevalence of these types of classes across time and countries. We also present arguments used for and against these classes and review empirical evidence of the effects of MA classes on cognitive and non-cognitive abilities.

Mixed Age Classes: Development and Definition

MA classes arise for two main reasons: either through economic necessity (too few children in an area to form a class or too few teachers to cover all grades) or through choice (a belief that MA classes are pedagogically superior). Generally, the knowledge is poor about the prevalence of MA classes across time and countries (Little, 2001), but some general patterns can be described. Historically, MA classes were the only possible way of

[1] For reviews of the class-size literature, see Hanushek (1999) and Krueger (2003).

organizing schools due to low population density. Still today, MA classes are mainly found in rural areas where they have been formed out of necessity. However, from the 1960s onwards, a belief in the pedagogical benefits of MA classes started to spread in many countries and, today, MA classes within urban areas is usually the result of an active pedagogical choice (Little, 2001).

In Sweden, the number of MA classes increased rapidly from 1980 onwards (Vinterek, 2001, 2003). In 2000, approximately one third of all Swedish students in the first three years of school attended MA classes and about one quarter of the students in grades 4 and 5. That is nearly twice as many as only five years earlier. The share of students attending an MA class during the last three years of compulsory schooling in Sweden is still rather small, about 2% of all students in these grades were in MA groups between 1996 and 1998.

We do not know whether this rapid increase in the number of MA classes in Sweden is due to pedagogical reasons or economic reasons (Vinterek, 2003). There is some evidence that pedagogical motives dominated in the lower grades (1–3) whereas economic motives dominated in the higher grades (4–6) (Sandquist, 1994). In grades 7–9, MA classes are scarce, and, if they do exist, they tend to be motivated by demographic necessities (Sandquist, 1994).

The initiative to start an MA class has usually come from groups of teachers within a school, often supported by the school management (Vinterek, 2003). However, since the beginning of 1990 it seems to be the case that MA classes have been introduced by politicians against the will of teachers and parents (Edlund & Sundell, 1999; Sandquist, 1994; Sundell, 2002; Vinterek, 2003). There is also evidence that MA classes are more prevalent in schools with many low performing students (Vinterek, 2003).

The basic definition of an MA class is a class consisting of students of different age and from different grades (as compared to a conventional class where all students are from the same grade and all or most students are of the same age). In practice, the term MA education often also implies a different type of teaching method, although there is little consensus about what characterizes this teaching method.

There are a number of different teaching strategies available: teaching the whole class simultaneously, ability grouping within the class irrespective of grades, and grouping by grades in some subjects while teaching the whole class simultaneously in others. Which strategy is most prevalent is unknown and, according to Little (2001), MA teaching is "invisible" in textbooks, syllabi, and teachers' education.

For the Swedish setting, there is some evidence suggesting that students in MA classes work more individually (Sandqvist, 1994; Vinterek, 2003).[2] There are also tendencies to grade-specific teaching. However, there are large differences depending on subject. Social sciences are often taught to all grades simultaneously, and leave large possibilities for group work and a thematic organization of the subject. In subjects like mathematics and athletics, teaching is more often done separately for each grade. This can be achieved in different ways. Sometimes one grade within the class works individually with one subject while the

[2] One reason could be the large heterogeneity within the class, making cooperation between students and group activities more difficult since they are at different knowledge levels. This implies that learning takes place through quiet reading and writing more than through listening and speaking. This is somewhat contradictory, since one common argument for MA classes is that the larger heterogeneity within the class enhances learning through group activities.

other grade(s) listens to the teacher lecturing. In other cases, the highest graders stay in school later in the afternoon and have time to learn more advanced mathematics when their younger classmates have left for the day (Sandqvist, 1994; Vinterek, 2003).

Arguments For and Against MA Classes

Mixed-age classes may differ from conventional classes in two ways: the composition of students, and the teaching methods. Most arguments in favor of MA classes focus on the former, that is, the effects of greater student heterogeneity. In the following we give an exposition of the most commonly used arguments for MA classes. Since the literature is mainly concerned with the supposed benefits of MA classes, this exposition will be one-sided; this does not mean that the pro-arguments have more empirical support.

Veenman (1995) discusses the following benefits of MA classes: MA classes are claimed to enhance the children's security and confidence as they form relationships with a wider variety of children. Mixed-age classes also invite cooperation, and children benefit from learning from and teaching each other. Furthermore, MA classes are considered to have a more relaxed atmosphere, and to be more stimulating as children from different ability levels meet. It is also claimed that the self-concepts of slower, older students are specially enhanced when they are asked to tutor younger students.

In order to motivate the introduction of MA classes in Sweden, the following arguments have been used by many local politicians in local school directives (Sandqvist, 1994). Mixed-age classes enable greater adaptation to individual maturity in different subjects and generate greater social training, since the group is more heterogeneous with respect to age. In addition, MA grouping is claimed to give rise to more acceptance for deviating behavior among classmates.

Some local politicians also refer to the pedagogical idea that students are assumed to be naturally curious and hungry for knowledge and that children spontaneously learn from each other and willingly teach each other.[3] Given this view of schooling and children, a more heterogeneous group is desirable. Another argument, connected to the former, is that the new post-modern information-intensive society requires knowledge about how to *search* for information. To work in project teams and to cooperate among students in order to search for information are new features in the school directives that is claimed to fit well with MA teaching.

Sundell (1995) also describes the arguments used in directives from the former Swedish National Agency for Education. Among the arguments in favor of MA classes, the supposed positive impact on students' cognitive development is claimed to stem from the teaching adapted to the individual that is connected with MA classes, as well as the idea that younger students learn from their older peers. The reason for the former argument is that in an MA class, working groups are formed in accordance with the individual child's mental maturity rather than its actual age.

Further, it is often claimed that the individually adapted teaching connected with MA classes specially benefits low performing students. The reasons are several. First, it is argued that the individually adapted teaching results in more teaching time to those in special need. Second, teaching in an MA class is to a higher degree organized in small

[3] The Montessori pedagogy is mentioned in some local school directives (Sandqvist, 1994).

groups, which benefits low performing students. Finally, as stated above, in an MA-class, low performing students have the possibility to compare themselves with younger children and in this way they do not need to perform worst.

Arguments are sometimes contradictory. For example, student heterogeneity is viewed as beneficial either because heterogeneity in itself is positive or because this heterogeneity allows more ability grouping, that is, less heterogeneity within the classroom. As another example, while some claim that MA classes give the teachers a better working environment as only a share of the class is new every year (Sundell, 1995), others instead argue that MA classes impose a greater workload on the teachers and that most teachers are not adequately prepared to deal with MA groups (Veenman, 1995).

In sum, there is no theoretical consensus about the mechanisms behind MA classes.

Empirical Studies

The empirical evidence on the effects of MA classes is ambiguous and many studies are of poor quality and it is questionable if they really capture causal effects. For example, Veenman (1995) summarizes evidence from 56 international studies. No study is based on exogenous variation. Experimental studies with random assignments into treatment and control groups have for natural reasons not been conducted. However, there is no study using any kind of "natural experiment," such as policy interventions creating exogenous variation in the availability of MA classes across time and area. Further, there are virtually no studies based on representative samples of the student population with well-defined treatment and comparison groups. Many studies do not even make any attempts to condition on initial differences between students in MA classes versus traditional classes. The overall evidence from this literature is contradictory, and when summarizing the results from the studies of best quality, the average effect of attending an MA class becomes zero.

The reason for this zero effect is discussed by Mason and Burns (1996). They argue that selection of better students and/or teachers into MA classes are counteracted by less effective instruction in these classes. In another review (Mason & Burns, 1997) they investigate research on "combination classes"—MA classes formed out of economic necessity—and find negative instructional effects, but positive selection effects of students and/or teacher effort, yielding on average zero effects. In yet another review, Lloyd (1999) finds positive effects of MA classes on high-ability students.

Using Swedish data, Sundell (2002) estimates the effect of MA class attendance in grade 2 on a number of abilities. Important to note is that the 752 students included in his study were not randomly sampled. When controlling for social and pedagogical background as well as initial achievements, the results show that students in MA classes performed worse than other students in 12 out of 13 dimensions. The MA students had, for example, lower mathematical ability, a less developed vocabulary and were perceived as more shy and troublesome by their teachers. However, they did perform better in reading comprehension.

Data

In this section, we describe the data used and show the differences between MA and traditional classes in terms of some important aspects.

Data Sources

Our main data source is a stratified panel data set: Student Panel 4, provided by Statistics Sweden.[4] In this panel, one cohort of students is followed through grades 3 to 9. In the first stage, 35 municipalities were selected. In the second stage, a random sample of grade 3 classes within these municipalities were selected.[5] Within the selected classes, information from all students in grade 3 was collected. This means that for students in traditional classes, we have information on the whole class, while for students in MA classes we only have information on the part of the class that spent their third year in school in 1992. That is usually one half or one third of the class, depending on how the MA-class is constructed.

The sampling of grade 3 classes was done in 1992, hence most students were born in 1982 and finished grade 9 in 1998. It is important to note that all students sampled in grade 3 are followed over time, regardless of whether they move or change class, hence, regarding these data, there is virtually no attrition. The panel includes approximately 8,500 individuals.

This panel data set is combined with additional register data from the databases RAMS and LOUISE provided by Statistics Sweden. These data include socio-economic background information such as parental education and immigrant status. Most of this information is measured in 1998. We focus on students who finish grade 9 in the expected year 1998 or later.[6]

In addition, we have access to a survey with information on students, parents and teachers and their attitudes towards school-related issues. This information was collected by the Department of Education at Göteborg University when the students were in grade 6.[7] Parents were asked about their involvement in school issues and if they had actively chosen the school or simply accepted the nearest one. Teachers were asked about their work experience, whether they had a formal degree, and their attitude towards homework. Results from tests measuring verbal and reasoning abilities in grade 6 were also collected (see Appendix for a more detailed description of the tests). The predictive validity of these tests on success in higher education is high, as shown by, for example, Emanuelsson, Reuterberg, and Svensson (1993) and Svensson (1980). Also the reliability is high, see for example Emanuelsson et al. (1993).[8]

Due to non-response, survey information is only available for a sub-sample of the original sample. Of the individuals in the original sample, 85% undertook the grade 6 test, and 54% answered all of the survey questions we used. It is this reduced sample we use for our analyses. Table A3a in the Appendix shows the difference between the raw register data, data with test results available (the basic sample), and data with all survey information available (our survey sample). The differences in means are very small when comparing the raw data and the basic sample. In 6 out of 27 cases there are statistically significant

---

[4] Participation in the study was voluntarily. About 4% of the originally sampled students were not able to, or chose not to, participate in the study.

[5] For more information about how the data was collected, see Statistics Sweden (1996). Throughout the paper, we will show descriptive statistics and estimation results from unweighted data, since the number of students in MA classes is small and outliers could potentially be given large sampling weights.

[6] Sixteen students finished school one year earlier, but due to a changed grading system we do not include these in our sample.

[7] For a more detailed description of the data, see Härnqvist (2000).

[8] Emanuelsson et al. (1993) report reliability coefficients of 0.9 (Kuder-Richardson's formula 20).

differences at the 10%-level and in these cases the magnitudes of the differences are small. Comparing the raw data with the survey sample, there are some additional differences. The survey sample seems to consist of a slightly more "privileged" group of students than the raw data. For example, students in the survey sample have higher average credits and grade 6 test results, they are more seldom given special help or mother tongue education in grade 3, and their parents are better educated.

Our data are mainly collected at the individual level, but some variables are, for obvious reasons, measured at the class level (teacher information) or school level (school information). The inference from descriptive statistics previously discussed (Table A3a in the Appendix) is based on individual-level variation. This potentially implies underestimated standard deviations for the class- and school-level variables. In the Appendix we present descriptive statistics of these variables with standard deviations calculated on the school level (individual classes are unfortunately not observed in our data). At this level, there are no statistically significant differences between the basic and survey sample.

Differences Between MA Classes and Traditional Classes

Table 1 presents descriptive statistics for students in MA and traditional classes in our survey sample. First of all, we can note that students in MA classes have lower scores on the grade 6 cognitive tests. This could have two different explanations: one is that MA classes are detrimental to student achievement, and another is that we have negative selection into MA classes. Regarding parental and student characteristics, the groups are relatively similar, with two exceptions. Students in MA classes have fewer mothers with a university degree, and are more often given mother tongue education in grade 3.

Regarding teacher and class characteristics, the differences are more striking. Mixed age classes are usually smaller. The teachers in MA classes are less experienced, have spent a shorter time in each class, and are more often on leave than teachers in traditional classes. The teachers' attitudes also differ.[9] Teachers in MA classes put less emphasis on homework, basic knowledge, and formal tests than teachers in traditional classes. Mixed age class teachers also believe student influence to be more important than their colleagues in traditional classes. Hence, from these descriptive statistics it seems as if the pedagogical environment for students in MA classes differs substantially from the environment in traditional classes.

The Causal Effect of Mixed Age Classes

In this section, we describe our strategy for identifying and estimating the causal effect of MA classes on students' cognitive skills. We also discuss potential threats to the validity of our approach and how we handle them.

Identification and Estimation Strategy

The potential effect on cognitive skills of attending an MA class may stem from two types of factors: (1) the effect of interactions between students of different age and school

[9] The attitude variables are measured on a 1–5 scale; the more important a teacher regards the issue, the higher the number. See Appendix for more details.

Table 1
Descriptive Statistics of Students Attending an MA Class During Grades 4–6 Versus Others, Survey Sample

	MA class in grades 4–6		Ordinary class in grades 4–6	
	Mean	SD	Mean	SD
Individual characteristics				
Grade 9 credits	51.49	28.34	52.67	28.66
Grade 6 test results	46.99	28.23	52.20***	28.80
Female student	0.48	0.50	0.50	0.50
Early start	0.01	0.11	0.01	0.08
Late start	0.03	0.16	0.02	0.15
Birth month	6.11	3.43	6.27	3.35
Help in grade 3	0.16	0.37	0.19	0.39
Mother tongue in grade 3	0.11	0.31	0.08*	0.27
Non-Nordic student	0.07	0.25	0.06	0.24
Mother sec. educ.	0.46	0.50	0.46	0.50
Mother univ. educ.	0.26	0.44	0.33***	0.47
Father sec. educ.	0.37	0.48	0.40	0.49
Father univ. educ.	0.19	0.39	0.22	0.41
Father educ. miss	0.20	0.40	0.22	0.41
Mother educ. miss	0.06	0.23	0.06	0.23
Father non-Nordic	0.09	0.29	0.10	0.31
Mother non-Nordic	0.11	0.31	0.10	0.30
Birth country miss	0.00	0.06	0.00	0.03
Father birth country miss	0.03	0.18	0.03	0.16
Mother birth country miss	0.02	0.12	0.02	0.13
Parent attitude: active school choice	0.15	0.36	0.15	0.36
Parent attitude: parent help	1.87	0.89	1.91	0.95
Parent attitude: parent active	2.39	1.05	2.34	1.03
Teacher and class characteristics[a]				
International school	0.0000	0.0000	0.0007	0.0265
Confessional school	0.0032	0.0562	0.0028	0.0530
Special school	0.02	0.14	0.03	0.18
Grade 9 students	101.56	39.25	114.31***	40.42
Few grade 9 students	0.07	0.25	0.01***	0.11
Teacher experience	18.44	10.62	20.12***	9.65
Teacher not qualified	0.04	0.20	0.04	0.20
Class size	18.31	7.19	23.67***	5.91
Small class	0.13	0.33	0.01***	0.11
Large class	0.21	0.41	0.37***	0.48
Share boys	0.55	0.12	0.51***	0.11
Share Swe2 students	0.07	0.13	0.06	0.14
Teacher not full time	0.10	0.30	0.12	0.32
Teacher on leave	0.08	0.26	0.03***	0.17
Teacher year in class	2.56	1.28	2.80***	0.84
Teacher attitude: homework	3.56	0.94	3.83***	0.91
Teacher attitude: tests	2.66	0.77	2.96***	0.94

Table 1
(*Continued*)

	MA class in grades 4–6		Ordinary class in grades 4–6	
	Mean	SD	Mean	SD
Teacher attitude: basic knowledge	4.42	0.81	4.67***	0.60
Teacher attitude: student influence	3.96	0.80	3.85**	0.86
Teacher attitude: student responsibility	4.71	0.63	4.77*	0.50
N	317		4,267	

Notes. [a]Teacher and class information are collected at the individual level (the teacher has filled in one form for each student) and are treated as individual level information when calculating standard errors. The reason is that we cannot identify class in the data set. Significance levels for the difference in means: *** 1%, **5%, *10%. Swe2 is a special course in Swedish adapted for students who do not have Swedish as mother tongue.

experience, and/or (2) effects from parent and teacher involvement (see Figure 1). In the literature on MA classes, most arguments for the beneficial effects of MA classes focus on the student interaction effects.

Our purpose is to estimate the combined effect of (1) and (2) on cognitive skills. This is the relevant question from a policy perspective and this is also what a randomized experiment would capture.[10]

Our identification strategy is a regression adjustment approach. We estimate the following model using ordinary least squares (OLS): $y = \alpha + \beta ma456 + \delta X_1 + \gamma X_2 + m + \varepsilon$ where y denotes student achievement—either percentile[11] ranked results from grade 6 cognitive

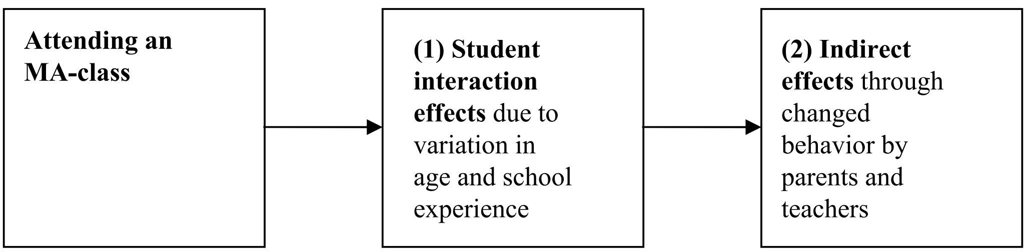

Figure 1. The different components of the MA-effect.

[10] Although it is not the purpose of our paper, measuring the effect of interactions between students of different ages only (i.e. part of effect [1]) is relatively easily achieved. Given birth dates for every student within each class, we could simply estimate the effect of age variance within a class on student outcomes. In our data set we only have information on all students within each class for the traditional classes, and our sample size is much too small for a precise estimation of this effect. In spite of that, we find that the point estimate of the age variance in traditional classes on student achievement is negative.

[11] The percentile ranks are, for transparency, calculated on the Register sample. Since the survey sample consists of a slightly positively selected group, the mean percentile ranks are slightly above 50 for the survey sample.

tests, or percentile ranked grade 9 credits. Our key explanatory variable, *ma456*, is a dummy variable for attending an MA class all years in grades 4 to 6.[12] It is important to note that a class is defined as an MA class only if it consists of students of both different ages and grades; this is not to be confused with traditional classes, where some students happen to be born a different year than the others (for example, students with learning difficulties or especially skilled students). The term X_1 denotes the covariates used to control for selection bias. These include socio-economic information such as parental education levels, immigrant status, gender, and birth month of the student, and information on whether the student was given special help or mother tongue education in grade 3. For a complete list of all variables, see the Appendix. When estimating the effect on grade 9 credits, we also control for the number of students in grade 9 at the school.[13] In addition, we have access to a variable indicating if the student attended an MA class also during grades 7–9. This variable is included as a control in a separate estimation. The term X_2 denotes the variables used to control for selection, but where there is some uncertainty about whether or not they instead represent the indirect effects of MA classes (see the next section). These variables include the attitudes and behavior of the teachers and parents. Finally, in all estimations we include municipality fixed effects, *m*.

We can also note that the two different measures of student outcomes, the grade 6 test results and the grade 9 credits differ in two respects. Not only do they capture short- versus long-run effects of attending an MA class, they can also reflect slightly different types of skills. While the grade 9 credits are a weighted average of grades in different subjects, and, as such, could include not only the teachers' assessment of the students' skills but also the students' behavior and diligence, the grade 6 tests are simply test results. The correlation between the two measures is relatively low, with a correlation coefficient of 0.57.

Another thing to note is that we do not have information on whether the student attended an MA class during grades 1–3. Since MA class attendance in grades 1–3 is likely to be correlated with MA class attendance in grades 4–6, it is possible that our dummy variable for MA class attendance in grades 4–6 also partly captures the long-run effects of earlier MA class attendance.

With the estimation strategy above, we implicitly assume that the effect of attending an MA class is equal for all groups of students. This may not be true—in fact, many of the arguments for or against MA classes are concerned with how they affect different kinds of students. In particular, it is usually argued that attending an MA class is especially valuable for students who do not perform as well as their peers. In many studies, it is shown that girls outperform boys in school and that immigrant students have lower school achievement than the average student. Hence, to relax the equal-effects assumption, we include interaction terms that allow the MA-effect to vary depending on gender, if the student has a non-Nordic background, and if the student was low performing in grade 3 (measured by if the student were given special help in grade 3).

[12] Using other definitions of the explanatory variable *ma456*, such as a dummy for attending an MA class only in grade 4 or at least one year during grades 4–6 or a cumulative variable capturing the number of years spent in an MA class does not change the results.

[13] We do not have information on the size of the school in grade 6.

Validity: The Selection on Observables Assumption

Our identification strategy relies on the selection on observables assumption, that is, that the control variables we condition the analysis on capture all potential selection of importance. This section discusses the validity of this assumption.

There are basically two different reasons for why a student attends an MA-class: (1) the number of students (or teachers) is too small to form an age homogeneous class, and/or (2) a belief in the MA-concept. Each reason gives rise to specific selection issues.

The first mainly occurs in smaller municipalities far from larger cities and/or in schools with special profiles, such as confessional or international schools. This is problematic if municipality size/school type is correlated with important student/teacher characteristics (for example, if students in rural areas simultaneously have relatively lower abilities and are more likely to attend an MA class.) Potential selection problems due to this is mitigated by including municipality dummies, information on the size of the school (in grade 9), and information on whether the school has a special profile (e.g., international and confessional).

The second implies selection problems if there is active sorting of students and/or teachers into MA classes. This could happen for several reasons. Parents could actively choose class type for their child, either by direct class/school choice within their area of living or by moving to/from areas offering MA classes. Principals could place different types of students/teachers in MA classes, or a different set of teachers could actively choose to work in MA classes. To handle potential selection due to this, we rely on our extremely detailed information, not only from register data but also from the survey data on students, parents, and teachers and their self-reported attitudes towards different school issues. One of our most important variables in this context is information on whether the parents have made an active school choice or not.

However, the choice of control variables is not straightforward, partly because the literature on MA classes is rather vague, and partly because some of our control variables are measured in grade 6. Although we view our variables to be controls for sorting and selection, it could be the case that some of them also reflect the indirect effect (2) of attending an MA class. One example is the variable attempting to measure how involved the parents are in school issues. Active parents may actively choose an MA class (or traditional class) for their child (in which case the variable becomes an important control for selection), but it could also be the case that parents in MA classes (or traditional classes) are forced to become more actively involved in school issues (in which case the variable represents the indirect effect of MA). We will estimate the effect of MA classes both with and without the control variables considered potentially problematic. In the list of variables in the Appendix, we have distinguished between these two types of variables. Finally, in the sensitivity analysis we test the validity of this selection on observables assumption by adding a proxy for ability.

Validity: Model Specification

Our estimation strategy relies on the assumptions underlying the classic linear regression model. First, the relationship between the explanatory variables and the outcome must be linear. In the sensitivity analysis, we relax this assumption and use an unparametric matching strategy. We have also employed different specifications of the variable of interest, MA (see footnote 13).

Second, there is the issue of spherical errors. In our case, serial correlation is not an issue since the data does not have a time dimension. Instead, the important issue is heteroscedastic errors. This is particularly important since some variables are common within classes and/or schools. For that reason, inference is based on robust standard errors that allow for heteroscedastic errors and also for correlation within schools. We cluster on school in grade 9 (as opposed to school in any earlier grade) due to data limitations—we only have information on the school in grade 9.

Results

How does attending an MA-class affect student performance? In the first section below we estimate the average effect, while in the second section we examine whether the effect varies by observed characteristics.

Main Results

Table 2 and Table 3 show the effect of MA classes on grade 6 cognitive tests and grade 9 credits, respectively. For the cognitive test results, there is a negative and statistically significant effect of attending an MA-class. The effect is also relatively large, around 5 percentile points in all columns. This can be compared to the effect of class size reductions. In the Tennessee STAR experiment, reducing class size by one student increased student performance with almost one percentile point (Krueger, 1999). A rough extrapolation suggests that MA class attendance has around the same impact as increasing class size by five students.

The estimated effect on grade 9 credits is also negative but much smaller in magnitude (around 1 percentile point in columns 2 and 3) and not statistically significant. One possible explanation is catch-up effects—in the long run, the gap between children from MA classes and traditional classes may be reduced due to more effort from children, parents, and/or teachers. However, it is also important to remember that the grade 6 tests and grade 9 credits measure slightly different types of skills, where credits to some extent also reflect the students' diligence and behavior, while the cognitive tests are simply test results.

Given the large sample size, the preferred models in column 2 and 3 explain a fairly large fraction—around 30% for both outcome variables—of the variation in the data. The

Table 2
OLS Estimates of the Effect of Attending an MA Class During Grades 4–6 on Percentile Ranked Grade 6 Test Results, Survey Sample

	Grade 6 test results	Grade 6 test results	Grade 6 test results
MA grades 4–6	–5.681	–4.325	–4.509
	(2.059)**	(1.786)*	(1.825)*
Including X_1	No	Yes	Yes
Including X_2	No	No	Yes
R^2	.05	.29	.30
F-test if added parameters jointly equals zero		75.83	3.17
Probability>F		(0.0000)	(0.0026)
N	4,584	4,584	4,584

Notes. All models include municipality dummies, standard errors (in parentheses) are clustered on schools. Significance levels: ** 1%, *5%.

Table 3
OLS Estimates of the Effect of Attending an MA Class During Grades 4–6 on Grade 9 Credits, Survey Sample

	Grade 9 credits	Grade 9 credits	Grade 9 credits	Grade 9 credits
MA grades 4–6	−2.579	−0.989	−1.169	−0.915
	(1.742)	(1.317)	(1.315)	(1.336)
Including X^1	No	Yes	Yes	Yes
Including X^2	No	No	Yes	Yes
Including MA grades 7–9	No	No	No	Yes
R^2	.04	.30	.30	.31
F-test if added parameters jointly equals zero		53.76	4.35	23.11
Probability>F		0.0000	0.0019	0.0000
N	4,584	4,584	4,584	4,584

Notes. All models include municipality dummies; standard errors in parentheses, clustered on schools.

variable for MA class attendance explains only a small proportion of this variation (R2-value in the fully extended model increases from .2730 to .2741 when adding the MA dummy). However, the important thing to note is that the variable has both an economically and statistically significant effect on test results.

We can also note that the negative effect of attending an MA class results remain in about the same range regardless of the set of covariates used. A comparison between column 2 and column 3 in Tables 2 and 3 shows no large differences. Hence, the variables added in column 3, that we view as good controls for selection but that potentially also could capture the indirect MA-effects, do not seem to be important in explaining the difference in achievement between students in MA and traditional classes. This is interesting since these variables include the parental and teacher attitudes towards school issues. In the Data section, above, we noted that the largest differences between MA and traditional classes were in terms of these different parental and teacher attitudes. At the same time, they seem unimportant for explaining the negative effect of MA classes.

Heterogeneous Effects

Table 4 shows the results from the heterogeneous effects estimations. None of the interactions between MA class attendance and the different subgroups are precisely estimated, most likely due to small sample sizes for each subgroup. However, if we interpret the point estimates, it seems as if girls and non-Nordic students may be less negatively, or even positively, affected by attending an MA-class, while low-performing students (as measured by the help in grade 3-variable) are more negatively affected. This is in sharp contrast to the arguments commonly used in favor of MA classes—that MA classes especially benefit low performing students.[14]

[14] We have also studied the same heterogeneous effects on the grade 9 credits, but find no statistically significant differences between groups.

Table 4
OLS Estimates of Heterogeneous Effects of Attending an MA Class During Grades 4–6 on Percentile Ranked Grade 6 Test Results, Survey Sample

	Grade 6 test results
MA grades 4–6	−6.917
	(2.907)**
Female student	−3.278
	(0.809)***
(MA grades 4–6)* (Female student)	4.730
	(3.959)
Help grade 3	−24.891
	(0.916)***
(MA grades 4–6)* (Help grade 3)	−1.555
	(3.842)
Non-Nordic student	−6.722
	(2.510)***
(MA grades 4–6)*(Non-Nordic student)	7.151
	(5.709)
Including X_1	Yes
Including X_2	Yes
R^2	.30
N	4,584

Notes. All models include municipality dummies. Significance levels: *** 1%, **5%, *10%. Standard errors in parentheses, clustered on schools.

Sensitivity Analysis

Our regression adjustment approach is based on several assumptions such as correct functional form, no omitted variables, and no selection into class type, conditional on the observed variables. In general, these assumptions are strong, although the richness of control variables increases the credibility of this study. Below, we extend the main analysis in different ways in order to investigate the robustness of our results. First, we include an additional proxy variable for ability and re-estimate the main model. Second, we re-estimate the model using a larger sample, and show calculations on the probability of entering the sample, depending on ability and class type. Finally, we also employ a propensity score matching method.[15] All in all, these sensitivity analyses do not modify our conclusion that attending an MA class is negative for student achievement.

An additional issue is potential measurement errors. Since our most important control variables stem from register information, we do not consider this to be a serious problem. MA attendance is a survey variable but it is hardly a sensitive question. Hence, we do not suspect measurement errors. In addition, if there are, it would only imply attenuation bias, yielding a smaller-than-true coefficient estimate.

[15] Matching approaches relaxes the linearity assumption. In addition, matching addresses the issue of selection on observables somewhat differently by only comparing individuals within the common support. See, for example, Black and Smith (2004) for a discussion of this in the educational context.

Adding a Proxy for Ability

If there is selection of students and/or teachers into different class types depending on unobserved ability, the estimated effect of attending an MA class is biased. To investigate this, we add the grade 6 test result for spatial ability ("metal folding")[16] to the fully extended model (Table 2, column 3). Recent evidence in Öckert (2009) suggests that spatial ability is less malleable to schooling than inductive or verbal ability, which is our reason for not using it as an outcome variable. However, it might indeed serve as a proxy for initial ability.

Including spatial ability in the model reduces the MA-coefficient slightly (from −4.509 to −3.169), but the difference is not statistically significant. Hence, unobserved heterogeneity in terms of ability does not seem to bias our results.

In addition, and in order to explicitly allow the MA-effect to vary depending on spatial ability, we also include an interaction term between this ability measure and the MA-coefficient. The estimate of this interaction term is not statistically significant and the parameter of interest (the MA-coefficient) is about the same as in the main estimations (in fact slightly larger but less precisely estimated).

Sample Selection Analysis

Sample selection is a concern for both external and internal validity. If the ability distribution differs between students who respond to the survey and the total population of students, external validity is violated. If the ability distribution of survey respondents differs between MA classes and traditional classes, internal validity is threatened.[17] Some calculations can help shed light on this issue.

Unconditional on covariates, there is a small positive correlation between ability (as measured by the variable "metal folding," our proxy for initial ability) and the propensity to answer the survey (the response rate). This means that our sample consists of a slightly more high-ability group of students than the general population, which may violate external validity. Moreover, the magnitude of this correlation also differs between students in different class types—the correlation is 0.036 for MA-students, and 0.049 for students attending traditional classes. In practice, this means that we draw comparison students slightly more from the upper part of the ability distribution which may violate internal validity.

One way to examine this issue is to re-estimate the model on a larger sample closer to the total population of students in Sweden. We use the basic sample, which only requires register covariates to be available, and hence is a more representative sample than the survey sample.[18] The results are shown in Table A4 in the Appendix. Clearly, the estimated effect of attending an MA class is still negative, although slightly less so, in the basic sample compared to the survey sample.

In addition, we estimate how the probability of being included in the survey sample depends on MA class attendance, ability, and the interaction between MA-class and ability,

---

[16] See the Appendix for more information about this test. Ideally, we would want our measure of ability to be collected prior to school start or at least prior to our period of observation but such a variable is not available.

[17] Among MA-students 55% answered the survey while the corresponding number for comparison students is 64%.

[18] The basic sample contains 7,234 observations, compared to the 8,531 individuals in the raw data.

Table 5
Sample Selection Into Survey Sample, on Basic Sample

	Prob (included in survey sample)
MA grades 4–6	–0.092
	(0.054)*
Spatial Ability	0.000
	(0.000)*
Spatial Ability*MA	–0.000
	(0.001)
R^2	.07
N	7,228

Notes. The basic sample is reduced by 6 observations due to missing information on the ability measure. The model includes register covariates and municipality dummies. Significance levels: *10%. Standard errors in parentheses, clustered on schools.

plus the register covariates.[19] The results are shown in Table 5. Clearly, the propensity to answer the survey is slightly lower among MA students; however, this in itself does not necessarily violate external or internal validity. The coefficient for spatial ability is close to zero (although statistically significant), which indicates that conditional on covariates, our survey sample is representative of the total population in terms of ability. The interaction term between MA-status and ability is also zero and statistically insignificant, indicating that, conditional on covariates, we have no selection in terms of ability into different types of classes. In sum, the register control variables seem to handle the potential problems of sample selection well.

Propensity Score Matching

Finally, we also estimate the effect of MA-class attendance by propensity score matching. First, we estimate a probit regression model for the probability of attending an MA class. In the probit regression, we use the fully extended model (column 3, Table 2). Second, we match (nearest neighbor without replacement) these predicted values so that for each MA student, we get one comparable individual who has attended a traditional class. In this matched sample, there are no statistically significant differences between MA and comparison students. Using this more homogeneous sample, we estimate the effect of MA without any covariates, that is, we simply compare the mean values. The results are presented in Table 6. The point estimate of attending an MA class is about the same as in the linear regression model, but, as expected due to a smaller sample size, the precision is lower. Hence, using a matching model which relaxes the linear functional form of OLS and also

[19] The model becomes $P(\textit{included in the survey sample})_i = a + bMA_i + cABILITY_i + dMA_i * ABILITY_i + e * X_i + f_i$, where X contains the register covariates. The coefficients of interest are *c* (to investigate sample selection threatening external validity) and *d* (to investigate sample selection threatening internal validity).

Table 6
Matching Approach, Survey Sample

	Grade 6 test results	Grade 9 credits
MA grades 4–6	–3.532	–0.479
	(3.412)	(2.658)
R^2	.00	.00
N	608	608

Notes. The sample is slightly reduced (by 2*15 individuals) since the probit regression perfectly predicts success/failure for 15 MA-students. Standard errors in parentheses, clustered on schools.

excludes non-comparable individuals (outside the common support) does not alter our main conclusions.

Concluding Remarks

Despite ambiguous scientific evidence, MA classes are a common phenomenon in schools around the world. In some cases it is because of demographic necessity, in other cases it is because MA classes are claimed to enhance student achievement. In Sweden, these types of classes have been rapidly re-introduced and, nowadays, around one fourth of all children attend an MA-class during grades 4–6.

In this paper, we present evidence that MA classes have a negative effect on students' cognitive skills. The short-run effect is relatively large—attending an MA class during grades 4–6 lowers the results on cognitive tests in grade 6 by almost 5 percentile points. Also the effect on grade 9 credits is negative, but smaller in magnitude (around 1 percentile point) and not precisely estimated. The smaller and statistically insignificant effect on grade 9 credits should not be interpreted as the choice of class type is irrelevant from a policy perspective. First, since the point estimate is smaller, we might need a larger sample to estimate it precisely. Second, even if the effect is zero in the long run, this could be due to catch-up effects that potentially require additional effort and resources from the student and/or the school.

The analysis is based on a regression adjustment approach, which relies on a number of assumptions such as the selection on observables assumption and correct functional form. However, the negative effect of MA class attendance is robust to a rigorous sensitivity analysis, including adding a proxy for ability and relaxing the functional form by using a matching estimator.

We have not been able to distinguish between MA classes introduced out of pedagogical beliefs and MA classes introduced out of economic and/or demographic necessity. Since the effect of MA class attendance could differ between these two groups, this would be an interesting topic for future research.

References

Andrae Thelin, A. (1991). *Undervisning i åldersblandade grupper: En granskning i samarbete mellan skolöverstyrelsen och länsskolnämnderna läsåren 1989/1990 och 1990/1991* [Teaching in mixed-age groups: A review in cooperation with the former Swedish National Agency for

Education and the county school councils] (Skolöverstyrelsens rapport R91:32). Stockholm: Skolöverstyrelsen.

Black, D.A., & Smith, J. (2004). How robust is the evidence on the effect of college quality? Evidence from matching. *Journal of Econometrics, 121*, 99–124.

Edlund, A.C., & Sundell, K. (1999). *Åldersintegrerat eller åldersindelat? En jämförande studie av 1111 elever i årskurs 2* [Mixed-age or age-separated? A comparative study of 1111 students in grade 2] (FoU-rapport 1999:9). Stockholm: Socialtjänstförvaltningen.

Emanuelsson, I., Reuterberg, S-E., & Svensson, A. (1993). Changing differences in intelligence? Comparisons between groups of 13-year-olds tested from 1960 to 1990. *Scandinavian Journal of Educational Research, 37*, 259–277.

Hanushek, E. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*, 143–163.

Härnqvist, K. (2000). Evaluation through follow up: A longitudinal program for studying education and career development. In C.-G. Jansson (Ed.), *Seven Swedish longitudinal studies in the behavioral sciences* (Report no 2000:8) (pp. 76–114). Stockholm: Swedish Council for Planning and Coordination of Research.

Krueger, A. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*, 497–532.

Krueger, A. (2003). Economic considerations and class size. *Economic Journal, 113,* 34–63.

Little, A.W. (2001). Multigrade teaching: Towards an international research and policy agenda. *International Journal of Educational Development, 21*, 481–497.

Lloyd, L. (1999). Multi-age classes and high ability students. *Review of Educational Research, 69,* 187–212.

Mason, A., & Burns, R. B. (1996). 'Simply no worse and simply no better' may simply be wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research, 66*, 307–322.

Mason, A., & Burns, R.B. (1997). Reassessing the effects of combination classes. *Educational Research and Evaluation, 3*, 1–53.

Sandqvist, K. (1994). *Åldersintegrerad undervisning: En kunskapsöversikt* [Mixed-age teaching: A survey]. Stockholm: HLS.

Statistics Sweden (1996). *Elevpanel för longitudinella studier Elevpanel 4* [Student panel for longitudinal studies. Student panel 4] (Statistiska meddelanden U73, SM 9601). Stockholm: Statistics Sweden.

Sundell, K. (1995). *Åldersindelat eller åldersblandat? Forskning om ålderssammansättningens betydelse i förskola och grundskola* [Mixed-age or age separated? Research on the importance of age-mix in preschools and schools]. Lund, Sweden: Studentlitteratur.

Sundell, K. (2002). *Är åldersblandade klasser bra för eleverna? En jämförande studie av 752 elever i årskurs 2 och 5* [Are mixed-age classes good for the students? A comparative study of 752 students in grade 2 and 5] (FoU-rapport 2002:7). Stockholm: Socialtjänstförvaltningen.

Svensson, A. (1964). *Sociala och regionala faktorers samband med över- och underprestation i skolarbetet: Pedagogisk-sociologiska studier jämte en beskrivning av skolstatistikens individuppgifter* [The correlation between social and regional factors and over- and under-achievement in school: pedagogical-sociological studies and a description of the individual variables in the school data]. Göteborg, Sweden: Department of Education, Göteborg University.

Svensson, A. (1980). On equality and university education in Sweden. *Scandinavian Journal of Educational Research, 24*, 79–92.

Veenman, S. (1995). Cognitive and non-cognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research, 65*, 319–381.

Vinterek, M. (2001). *Åldersblandning i skolan – elevers erfarenheter* [Students' experiences of mixed-age grouping] (Doctoral thesis, Umeå University).

Vinterek, M. (2003). *Åldersblandade klasser: Lärares föreställningar och elevers erfarenheter* [Mixed-age classes: Teachers' views and students' experiences]. Lund, Sweden: Studentlitteratur.
Öckert, B. (2009). Does going to high school make you smarter? (Unpublished manuscript). Uppsala, Sweden: Institute for Labour Market Policy Evaluation, Uppsala University.

Appendix

The Cognitive Tests From Grade 6

There are three test scores from grade 6 available. The tests represent verbal, reasoning, and spatial abilities and are called: Opposites (motsatser), Number series (talserier), and Metal folding (platvik). In the test called Opposites, the child is asked to find the opposite of a given word among four choices (40 items, 10 minutes). In the Number series test the child is instead asked to complete number series (40 items, 18 minutes). In the last test, Metal folding, the child is asked to find the three-dimensional object among four choices that can be made from a flat piece of metal (40 items, 15 minutes).[20] The results on each of these tests are measured on a scale ranging from 0 to 40.

In this paper, we use the percentile ranked sum of two of the tests, Opposites and Number series. The correlation coefficient between each of these tests and the grade 9 credits is 0.51. The third test, Metal folding, involves tasks not regularly practiced in schools, and its correlation coefficient to the grade 9 credits is 0.36.

Table A1
List of Variables: Key Variables and Register Data Variables

Variable name	Definition
Key variables	
Grade 6 test results	Percentile rank of the sum of the scores on the tests in number series and opposites given in grade 6
Grade 9 credits	Percentile rank of a summary measure of the student's 16 best credits in grade 9
MA grades 7–9	A dummy that equals 1 if the students attends an MA-class in any grade between grades 7–9
MA grades 4–6	A dummy that equals 1 if the students attends an MA-class during grades 4-6
Register variables	
Municipality dummies	One dummy for each municipality
Female student	A dummy that equals 1 if the student is female
Early start	A dummy that equals 1 if the student is born after 1982
Late start	A dummy that equals 1 if the student is born before 1982
Birth month	The student's month of birth
Help in grade 3	A dummy that equals 1 if the student has been given any form of special education intended for low performing students ("särundervisning", "anpassad studiegång" or "specialundervisning på annat sätt") in grade 3
Mother tongue in grade 3	A dummy that equals 1 if the student attended mother tongue education in grade 3
International school	A dummy that equals 1 if the school has an international profile

[20] A more detailed description of the test scores are given by Svensson (1964).

Table A1
(*Continued*)

Variable name	Definition
Register variables	
Confessional school	A dummy that equals 1 if the school has a confessional profile
Special school	A dummy that equals 1 if the school is not ordinary, for example schools at hospitals
Grade 9 students	The number of students in grade 9 at the school, collected in grade 9
Few grade 9 students	A dummy that equals 1 if the number of students in grade 9 at the school is smaller than 30
Non–Nordic	A dummy that equals 1 if the student is born in a non-Nordic country (missing values are set to 1)
Mother secondary education	A dummy that equals 1 if the mother of the student has secondary education, at most 5 years in addition to compulsory schooling (missing values are set to 0)
Mother university education	A dummy that equals 1 if the mother of the student has university education, more than 5 years in addition to compulsory schooling (missing values are set to 0)
Father secondary education	A dummy that equals 1 if the father of the student has secondary education, at most 5 years in addition to compulsory schooling (missing values are set to 0)
Father university education	A dummy that equals 1 if the father of the student has university education, more than 5 years in addition to compulsory schooling (missing values are set to 0)
Mother non-Nordic	A dummy that equals 1 if the mother of the student is born in a non Nordic country (missing values are set to 1)
Father non-Nordic	A dummy that equals 1 if the father of the student is born in a non Nordic country (missing values are set to 1)
Birth country missing	A dummy that equals 1 if information about the student's country of birth is missing
Mother education missing	A dummy that equals 1 if information about the mother's education is missing
Father education missing	A dummy that equals 1 if information about the father's education is missing
Father birth country missing	A dummy that equals 1 if information about the father's country of birth is missing
Mother birth country missing	A dummy that equals 1 if information about the mother's country of birth is missing

Table A2
List of Variables: Survey Data Collected in Grade 6

Variable name	Question	Definition
	To teachers:	
Teacher experience	What is your teacher experience in years?	A variable ranging from 1 to 43 (measured in years)
Teacher not qualified	Do you have a certificate qualifying you to teach at this level?	A dummy that equals 1 if the answer is no
Class size	What is the number of girls and boys in the class?	The sum of boys and girls in the class – ranging from 0 to 60
Small class	Constructed from Class size	A dummy that equals 1 if the class size is smaller than 10
Large class	Constructed from Class size	A dummy that equals 1 if the class size is larger than 25
Share boys	Constructed from Class size	The share of boys in the class

Table A2
(*Continued*)

Variable name	Question	Definition
	To teachers:	
Share Swe2 students	What is the number of students in your class that take Swe2? Constructed from Class size	The share of students in the class taking a special course in Swedish adapted for students who do not have Swedish as mother tongue
Teacher not full time	Do you work full time?	A dummy that equals 1 if the teacher work part time, 0 if full time
Teacher on leave	Have you been on leave during the last year?	A dummy that equals 1 if the teacher has been on leave full time or part time
Teacher year in class	How many years have you taught this class?	A variable ranging from 1 to 8 (measured in years)
Teacher attitude: homework (belongs to X_2, the extended set of covariates)	How important are homework and oral tests?	A variable ranging from 1 to 5 in the following way: Very important 5 Rather important 4 In between 3 Rather unimportant 2 Not at all important 1
Teacher attitude: tests (belongs to X_2, the extended set of covariates)	How important are formal tests?	See above
Teacher attitude: basic knowledge (belongs to X_2, the extended set of covariates)	How important is the emphasis of basic skills?	See above
Teacher attitude: student influence (belongs to X_2, the extended set of covariates)	How important is student influence during planning?	See above
Teacher attitude: student responsibility (belongs to X_2, the extended set of covariates)	How important is it that the student takes own responsibility?	See above
	To parents:	
Parent attitude: active school choice (belongs to X_2, the extended set of covariates)	Have you chosen another than the closest school to your child?	A dummy that equals 1 if the answer is *yes* or *yes we are going to* and 0 if *no or doubtful (probably not)*
Parent attitude: parent help (belongs to X_2, the extended set of covariates)	Do you participate in your child's school work?	A variable ranging from 1 to 5 in the following way: Very often 5 Rather often 4 Sometimes 3 Rarely 2 Almost never 1
Parent attitude: parent active (belongs to X_2, the extended set of covariates)	How much do you participate in school activities?	See above

Descriptive Statistics

Table A3a
Register Data, Basic Sample and Survey Sample, Variation at Individual Level

	Register data			Basic sample *n* = 7,234		Survey sample *n* = 4,584	
	Mean	SD	*n*	Mean	SD	Mean	SD
Key variables							
Grade 9 credits (1)	202.17	59.88	8,490	203.93*	58.56	209.16***	56.77
Grade 6 test results (2)	44.24	12.68	7,420	44.32	12.66	45.13***	12.55
MA grades 4–6	0.08	0.26	8,531	0.08	0.27	0.07	0.25
MA grades 7–9	0.03	0.17	8,531	0.02**	0.15	0.02***	0.15
Covariates capturing selection							
Female student	0.49	0.50	8,515	0.49	0.50	0.50	0.50
Early start	0.01	0.09	8,515	0.01	0.09	0.01	0.09
Late start	0.03	0.16	8,515	0.02	0.15	0.02	0.15
Birth month	6.28	3.36	8,515	6.27	3.36	6.26	3.35
Help in grade 3	0.21	0.40	8,531	0.20	0.40	0.19***	0.39
Mother tongue in grade 3	0.10	0.30	8,531	0.09	0.29	0.08***	0.27
International school	0.0011	0.03	8,360	0.0007	0.0263	0.0007	0.0256
Confessional school	0.0054	0.07	8,360	0.0043	0.0653	0.0028**	0.0532
Special school	0.04	0.19	8,360	0.04	0.19	0.03*	0.18
Grade 9 students	113.80	41.58	8,331	113.52	41.09	113.42	40.47
Few grade 9 students	0.03	0.16	8,331	0.02*	0.14	0.02***	0.13
Non-Nordic student	0.07	0.26	8,531	0.07	0.25	0.06**	0.24
Mother sec. educ.	0.45	0.50	8,531	0.45	0.50	0.46	0.50
Mother univ. educ.	0.30	0.46	8,531	0.30	0.46	0.32***	0.47
Father sec. educ.	0.38	0.48	8,531	0.39	0.49	0.40***	0.49
Father univ. educ.	0.20	0.40	8,531	0.20	0.40	0.22**	0.41
Father educ. miss	0.26	0.44	8,531	0.24*	0.43	0.22***	0.41
Mother educ. miss	0.07	0.26	8,531	0.07	0.25	0.06***	0.23
Father non-Nordic	0.12	0.33	8,531	0.11*	0.32	0.10***	0.30
Mother non-Nordic	0.12	0.32	8,531	0.11*	0.31	0.10***	0.30
Birth country miss	0.0014	0.04	8,531	0.0011	0.0332	0.0011	0.0330
Father birth country miss	0.03	0.18	8,531	0.03	0.17	0.03	0.16
Mother birth country miss	0.02	0.13	8,531	0.02	0.13	0.02	0.13
Teacher experience						20.00	9.73
Teacher not qualified						0.04	0.20
Class size						23.30	6.16
Small class						0.02	0.14
Large class						0.35	0.48
Share boys						0.51	0.11
Share Swe2 students						0.06	0.14
Teacher not full time						0.12	0.32
Teacher on leave						0.03	0.18
Teacher year in class						2.78	0.88
Parent attitude: active school choice						0.15	0.36

Table A3a
(*Continued*)

	Register data			Basic sample n = 7,234		Survey sample n = 4,584	
	Mean	SD	n	Mean	SD	Mean	SD
Covariates capturing selection and/or indirect MA-effects							
Teacher attitude: home works						3.82	0.91
Teacher attitude: test						2.94	0.93
Teacher attitude: basic knowledge						4.65	0.62
Teacher attitude: student influence						3.86	0.85
Teacher attitude: student responsibility						4.76	0.51
Parent attitude: parent help						1.90	0.95
Parent attitude: parent active						2.34	1.04

Notes. (1) and (2) are not percentile ranked. Significance levels for comparisons against register data: *** 1%, **5%, *10%.

Table A3b
Basic Sample and Survey Sample, Variation at the School Level

	Basic sample n = 484 (schools)		Survey sample n = 375 (schools)	
	Mean	SD	Mean	SD
International school	0.01	0.09	0.01	0.07
Confessional school	0.02	0.14	0.01	0.11
Special school	0.05	0.21	0.04	0.20
Grade 9 students	98.27	45.81	102.50	44.43
Few grade 9 students	0.08	0.28	0.06	0.24
Small class			0.02	0.13
Large class			0.34	0.41
Teacher experience			19.51	8.46
Teacher not qualified			0.04	0.15
Class size			23.29	5.89
Share boys			0.50	0.09
Share Swe2 students			0.09	0.15
Teacher not full time			0.12	0.26
Teacher on leave			0.05	0.18
Teacher year in class			2.72	0.71
Teacher attitude: homework			3.89	0.76
Teacher attitude: test			2.98	0.78
Teacher attitude: basic knowledge			4.62	0.54
Teacher attitude: student influence			3.89	0.72
Teacher attitude: student responsibility			4.77	0.44

The MA-Effect in the Basic Sample Versus the Survey Sample

Table A4
OLS-Estimates of the Effect of Attending an MA Class During Grades 4–6 on Percentile Ranked Grade 6 Test Results, Controlling For Register Covariates Only

	Basic sample	Survey sample
	Grade 6 test results	**Grade 6 test results**
MA grades 4–6	–1.171	–2.141
	(0.558)*	(0.795)**
Including register covariates	Yes	Yes
Constant	44.518	44.672
	(0.533)**	(0.643)**
R^2	.31	.30
N	7,234	4,584

Notes. All models include municipality dummies. Significance levels: ** 1%, *5%. Standard errors (in parentheses) clustered on schools.