Is a tree known by its fruits? The effect of the Cito-test on intergenerational educational mobility

February 1, 2018

Master thesis Policy Economics
Arie-Jan van der Toorn

Student number: 388382
Supervisor: prof.dr. Olivier Marie
Second reader: prof.dr. Dinand Webbink

Abstract

This thesis looks at the effect of the introduction of the cito-test on intergenerational educational mobility. Initiated by Professor A. De Groot, the cito-test was introduced in 1969. De Groot wanted to make the advice for a child to which level of secondary education it should go more just by having an objective test besides the assessment of a teacher. In this study, a fuzzy Regression discontinuity is exploited, trying to find a causal effect of the introduction of the cito-test on intergenerational educational mobility. The results show that the cohort just after the introduction of the test had a significant lower mobility than the cohort just before the introduction, indicating that the cito-test caused a drop in intergenerational educational mobility.

Contents

1 Introduction

‘The cito-test is outdated, please get rid of it immediately‘ (Volkskrant, 2014) ‘Teacher knows better what is good for a child than the cito-test‘ (NRC, 2014). ‘Cito-test also supports the teacher‘ (Volkskrant, 2016). Just some of the headlines in the last couple of years about the cito-test. The cito-test (since 2015 actually called Central test but most people still use the name cito-test) is a high-stake standardized test that is available for all primary schools in the Netherlands. As one can see from the headlines, the test is highly debated. In the Dutch educational system, in the last grade of primary education a child gets an advice from the primary school to which level of secondary education it should go. The cito-test is primarily being used as an indication for this advice. In the last decades the cito-test increased in popularity among primary schools in the Netherlands. It started out with 35.000 children (15%[1]) which made the test in 1969 to 100.000 children (55%) in 1990 and 157.000 (85%) in 2014 (CITO, 2014). Since 2015 the Dutch government obliged primary schools to use some kind of test for children in the last group of primary school (Rijksoverheid, 2018). Although it was allowed to use another test than the cito-test, most schools used to the cito-test. Along with the obligation of a test, the government did point out that the test should not be the most important factor for the advice to a student which level of secondary education to go to. However, in most cases, primary schools follow the advice that rolls out of the cito-test (Driessen, 2011).

There are many different views with respect to the importance of the cito-test. Some say that the cito-test should not be that important in the advice to a student. They argue that the assessment of a teacher is better than the assessment that flows out of the cito-test (Onderwijsinspectie, 2014). Others state that the cito-test must be more important in the advice to a student. They state that the assessment of teachers is too subjective and gives rise to inequality of opportunities (Van der Hoeven-van Doornum, 1990; Jungbluth, 2003). The founder of the cito-test, Professor De Groot, also thought that the assessment of a teacher was too subjective, making it not just towards children. This was one of the reasons for initiating the cito-test (de Groot, 1966). Whether the goal of the initiator of the cito-test to reduce the inequality of opportunities in the Dutch schooling system was achieved remains an unknown question. In this study I try to get an answer to this by examining the research question:

What was the effect of the introduction of the cito-test in 1969 on the intergenerational educational mobility between cohorts of students just before and just after the introduction of the test?

One way to measure equality of opportunities is to look at the intergenerational mobility of persons. Intergenerational mobility is a economic/sociologic concept that looks at the mobility between parents and their children. It measures whether a childs income/education level/social status in later life depends on the income/education level/social status of ones parent. The more mobile a child, the lower this dependency. This study uses this concept to look at the effect of the introduction of the cito-test in 1969. I perform a fuzzy Regression

[1]Percentage of total amount of children in the last group of primary school retrieved from (Statline, 2017)

Discontinuity which tries to estimate the causal effect of the introduction of the cito-test on intergenerational educational mobility. I use data from questionnaires which contain among other things the education levels of respondents and father of respondents. I use the birth-dates of the respondents to compare cohorts just before and just after the implementation of the cito-test.

This study adds to the literature in that it directly links the cito-test to a measure of equality of opportunities. As far as I know, this is not done before in the literature about the effects of the cito-test. There exists some literature on the effects of standardized testing on equality of opportunity, especially in the United states (Madaus & Clarke, 2001; Amrein & Berliner, 2002; Marchant & Paulson, 2005; Nichols et al., 2005). The conclusion in these studies is often that standardized testing does more harm than good in the chances of low social economic groups. In the Netherlands, there are some studies that look at the alternative of the cito-test: the assessment of a teacher in giving advice to a student. These studies show mixed results, in which some say that the assessment of a teacher is too subjective which could give rise to inequality to opportunities, whilst others show that the advice of a teacher is in most cases better than the advice based on the results of the cito-test (Van der Hoeven-van Doornum, 1990; Jungbluth, 2003; Feron et al., 2015). A part of the literature which is also relevant are studies about the effect of social-economic status on the scholastic achievement of students. These studies often show that a lower social economic status leads to lower achievements of students, which could be due to lower support from home (White, 1982; Gottfried et al., 1998). This study looks directly at the effect of the introduction of the cito-test on intergenerational educational mobility, and could thus say something about the preferability of a standardized test next to the assessment of the teacher with regard to equality of opportunities.

This paper succeeds as follows: section 2 will go deeper into the concept of intergenerational mobility, give some more insight in the implementation of the first cito-test and will give an overview of the literature on the effects of standardized tests. Section 3 explains which data is used in the study and which methodology is used to evaluate the data at hand. Section 4 shows the results of the study in combination with some robustness analysis. Section 5 concludes and discusses.

2 Intergenerational mobility and the cito test

2.1 Intergenerational mobility

Intergenerational mobility can be defined as the extent to which circumstances in ones childhood are reflected in outcomes in later life. In other words it is the extent to which individuals can achieve things by their own talents, motivation and luck. The level of intergenerational mobility is often seen as a measure of equality of economic and social opportunity. If a country is perfectly mobile, the income (or other characteristics) of the parents have no influence on the income of the children later on in life. In this case there is total equality of opportunity: children from poor families have the same chances as children from rich families. Many studies show that perfect mobility is not present in the real world. These studies show that the income position of the parents is, to a certain extent, carried over to the next generation

(Österberg, 2000; Corak, 2004; Piraino et al., 2006).

There exists some theory that tries to explain intergenerational persistence. Becker and Tomes (1979) created a human capital model which explains the decision of parents when it comes to investment in children and how these investments influences childrens outcomes in later life. In the model, parents maximize a Cobb-Douglas utility function when allocating their lifetime earnings between own consumption and investment in children. Previous models of human capital only explained inequality within one generation as a process of luck and ability. This model shows inequality through generations: as parents invest more or less in their children, inequality grows or shrinks in future generations. Following the Becker and Tomes model, Fox et al (2016) state that intergenerational mobility works in more ways than just through direct investment. First of all they argue that a difference between investments in children between poor and rich families causes a difference in mobility. More investment in children from poor families relative to rich families causes more mobility. Second, they show that intergenerational mobility could change due to a change in returns to endowments. To get a better understanding of this last effect one could look at height. Height is an endowment that is typically perceived to be genetically determined but environmental factors such as nutrition also play a role. As nutrition in a country improves, differences in height will likely decline and therefore the return of this endowment will decrease, which means that income inequalities based on height will be reduced and intergenerational mobility will rise.

Becker and Tomes also looked at the effect of credit constraints. They found that parents at the lower end of the income distribution have less access to credit markets. This causes that they cannot borrow enough to invest in their children, and thus there is more intergenerational persistence at the bottom of the income distribution, as parent with higher incomes also have more access to credit markets (G. S. Becker & Tomes, 1986). This hypothesis was heavenly tested but no real evidence has been found (Behrman & Taubman, 1990; Mulligan, 1997; Grawe, 2004; Mazumder, 2005).

The model of Becker and Tomes shows that, besides investment of parents, differences in endowments between different families are important for intergenerational mobility. Looking at the example of height, this is a typical thing where the government of a country plays a role. If a government tries to achieve that everyone in a country has access to the same basic nutritions, it also influences intergenerational mobility in the country. Via policy changes, governments can thus change the intergenerational persistence of a country. Various studies have tried to test these effects empirically. Black and Devereux (2010) have written an overview of studies that tried to explain the causal mechanisms underlying intergenerational mobility (or persistence). The earliest studies that look at these mechanisms, look at genetic effects in intergenerational persistence. Quantifying how much of the intergenerational persistence is due to genetics is however still very hard. More recently, the focus of explaining intergenerational mobility has moved to individual parental attributes and policy attributes on the outcomes of children (Black & Devereux, 2010). For example, Blanden, Gregg and Mcmillan (2007) tried to show which factors underlie intergenerational persistence in the United Kingdom. They look at the role of education, ability, non-cognitive skills and labour market experience in intergenerational persistence. The results show that inequalities in achievements at age 16 and in post-compulsory education by family background are extremely important in determining the level of intergenerational mobility. They also try to explain why mobility has dropped between 1958 and 1970. They show that the growing imbalance between access to higher education by family background accounts for a large part of the fall in intergenerational mobility (Blanden et al., 2007). Kotera and Seshadri (2017) try to examine how much of the variation in intergenerational mobility can be accounted for by the variation in state policies in the US. They show that a more equal distribution of public school spending between states improves intergenerational mobility (Kotera & Seshadri, 2017). In both these studies it is clear that policy changes could have an effect on intergenerational mobility within a country. In the study of Blanden et al, if a country could provide equal access to higher education for every income group, intergenerational persistence could be reduced. In the Kotera paper, if the public school spending between states is equalized, it will also reduce intergenerational persistence.

2.2 The cito-test

Until the late 20th century the division into different levels of higher education in the Netherlands was purely based on the societal position of the parents. The highest form of secondary education was for children from the intellectual elite, whilst children from the lowest social classes often did not even go to secondary school. This division was due to the judgement of the primary school. For a large part they based the decision to send a child to a level of secondary education on the social class it originated from. Since 1873 there also existed a matriculation for secondary school to ensure that only the best children went to secondary school. The fact that there was a huge difference in levels of education for different social classes changed slowly during the twentieth century as people became more alert to inequality of opportunities in society. To look for better connections between primary and secondary school people started to question the matriculation and looked for ways that did more right to the qualities of children (Regt, 2004).

One of the persons who criticized the Dutch educational system was Professor A. De Groot. De Groot was a professor of psychology and social sciences at the University of Amsterdam. In 1965 he created a test which was made by students from primary schools in Amsterdam from 1966-1969. This test had to replace the subjective judgement of the teacher in the advice of sending a pupil to a level of secondary education. A test should ensure more objectivity in this advice. Due to the minister of Education at that time a central institute for such a test was founded in 1968. This institute was called C.I.T.O. which stands for Centraal Instituut Toets en Ontwikkeling (Central Institute Test and Development). Until 2015 this institute made a socalled cito-test, which was available for every primary school in the Netherlands. Since 2015 this test is called Centrale Eindtoets (Central test).

De Groot wrote the reasons and ideas behind the cito-test down in his book “Vijven en zessen”. In this book he argues that the Dutch educational system needed a big reform. According to de Groot, grades as they were given in the Dutch system were prone to subjectivity of the teacher. He states that the prevailing notion was that the whole student had to be reflected in the overall grade of the student. So not only how someone makes a test but also how he or she acts in class. By this notion, a teacher was constantly busy to judge his pupils. This of course has consequences for the objectivity of the grades of a pupil: one teacher judges the actions of a pupil totally different than the other. In this way, one cannot know whether an overall grade of 6 for pupil A given by one teacher has the same meaning as an overall grade of 6 for pupil B given by another teacher. One could imagine that the background of a pupil could have a big influence on the judgement of the teacher. Many studies have shown that teachers have different expectations and a different attitude towards children from lower social classes (H. S. Becker, 1952; Douglas, 1964; Rist, 1970; Vandrick, 2014). This could have a negative effect on the grades of pupils from lower social classes, regardless of their capability and intelligence. De Groot states it in his book as follows: "From the start the focus of us (the Netherlands) is to create no, to preserve elites. (de Groot, 1966).

What were the solutions of de Groot to change the negatives of the educational system in the Netherlands? First of all he pleaded to have selection free periods. For example if a student succeeds his propaedeutic exam, the school has to take the responsibility to help the student succeed the whole education. In other words, when the student reasonably cooperates, there is no chance that he will not succeed the education: this is the responsibility of the school. To get to these selection free periods, the exams a student has to make have to be objective. Whether a student passes or not should only depend on the capability of the student, and this should be valued the same for every student. De Groot referred to these tests as study tests. According to de Groot, one of the biggest advantage of these study tests was that they were more just: scores and differences in scores could be justified better to students. The cito-test is an example of such a study test: a test that is equal for all students who make it. De Groot wanted the cito-test to be given on every primary school in the Netherlands, made by a National institute for development of study tests (de Groot, 1966). Such an institute became reality in the Central Institute Test and Development, but the cito-test did not become mandatory for every Dutch primary school. Nevertheless, in the first year of the cito-test almost 15% of all students in the last class of primary school made the test (Wijnstra, 1984). The cito-test also did not became the only measure that determines to which level of secondary education a child is going to. Eventually, the teacher and the primary school of a pupil give the advice. Nevertheless, the cito-test played and plays a very important role in determining the advice (Lubbe, 2005; Driessen, 2011).

Before the cito-test was implemented, there was a precursor of the test called the Amsterdamse Schooltoets (Amsterdam school test). This test was held only at primary schools in Amsterdam from 1966-1969. In a study on the effects of the Amsterdam school test on the procedure of selection to secondary education, Oosterbaan (1973) explains the role of this test in the advice of a primary school to a student. He states that the assessment of the teacher was most important in the advice. The procedure looked as follows: First, the teacher made his assessment to which level of secondary education a student should go. Then after the teacher received the results of the test, he reconsidered his first advice and made the second, final advice. Oosterbaan states that in 1967 as well as in 1968, the second advice was different from the first advice in 1 out of 4 cases (Oosterbaan, 1973). It is very likely that the procedure of the first cito-test was roughly the same as described above. This is also quite the same as how the advice is made nowadays: the advice of the school is leading, the cito-test is used as second opinion (WVO, 2018, 1 January).

2.3 Effects of the cito-test

Is the cito-test a good thing or a bad thing? That is an important and highly debated question in the Netherlands. De Regt (2004) argues that the cito-test is way too important in the selection procedure of students. He states that despite the critical attitude of many towards the cito-test as the only advisor for the level of secondary education, the cito-test became more and more important. An OECD rapport of (2014) supports the idea of De Regt that cognitive test scores gained importance in the last couple of years. Other studies focus more on the explanatory power of the cito-test. A CPB report from 2015 shows that the teachers assessment of a childs ability is a better assessment than that of the cito-test. The authors argue that when the teachers assessment dominates the assessment based on the cito-score, it could reduce switching of students between secondary education levels (because of a wrong advice) by at least ten percent (Feron et al., 2015). The other side of the debate argues that it is good that an objective test determines the allocation of students to different levels of secondary education. Professor Jaap Dronkers is a big advocate of the use of standardized tests in the educational system. In a paper with Prokic-Breuer (2012) he hypothesizes that the high performance of Dutch pupils is partly caused by standardized tests. They however do not find that these tests contribute significantly to the high performance of Dutch pupils. In 2015, Dronkers speaks about the idea of the Dutch government to reduce the importance of the cito-test by shifting the test from February to June. In this way the cito-scores of students were obtained after the advice of the school already had been put together. The cito-scores thus became less important in the advice about follow-up education of the primary school. In the Volkskrant, Dronkers states that this is a really bad idea, as this gives rise to the subjective assessment of the teacher, which leads, according to Dronkers, to more inequality of opportunities (Volkskrant, 2015). He follows the same reasoning as De Groot in saying that the subjective assessment of a teacher is not just. More researchers share the idea that a teacher has a big influence on inequality of opportunities for students (Van der Hoeven-van Doornum, 1990; Jungbluth, 2003).

2.4 Equality of oppurtunities

The question is whether the cito-test does a better job in assuring equality of opportunities than the assessment of the teacher. To my understanding, there is not a lot research with regard to the effect of the cito-test on equality of opportunities. There are some studies that look at influences of social background and situations at home on educational chances of children (Grotenhuis & Dronkers, 1989; Driessen, 2006). But the effect of taking the cito-test on the equality of opportunities is an uncovered field. In other countries there have been some studies that look at the effect of standardized test on equality of opportunities. The United States experienced a rise in standardized high-stake testing in the last decades. Both on national and state level the use of such tests became more and more important in judging students. Next to the assumptions that teachers will teach better and students will learn better due to standardized high stakes tests, an important argument for these test was also to create an equal opportunity for all students to demonstrate their knowledge (Amrein & Berliner, 2002). In the United States there is also a big ongoing debate about whether high-stake testing improves the educational system or not. When it comes to equal opportunities, a big criticism is that high stake standardized testing restricts the curricula of teachers and students. The effect of a restricted curriculum is that it does not account for student diversity. When the tests are high stake, i.e. the results have a great impact, teachers will often teach to the test which could harm other skills that students have. In a lot of cases non-testable courses like arts are pushed out of the curriculum (Rentner et al., 2006). The problem is that the curriculum of a school is now made centrally: the content of the standardized test (which is made centrally) determines the curriculum of a school. The argument is that an individual school or teacher is better in addressing different ways of learning of different students. When the way of learning is standardized by the tests, this could harm students that have a different way of learning. Multiple studies have shown that those high stake standardized tests especially have a negative effect on low-income and non-white students (Madaus & Clarke, 2001; Marchant & Paulson, 2005; Nichols et al., 2005). In the United States a big problem was also that the schools were judged based on the scores of the standardized tests. This led to the situations that schools intentionally filtered students out that had a lower chance of achieving a good score (Heilig & Darling-Hammond, 2008).

Another argument why a standardized test might not overcome the problem of a subjective assessment of a teacher could be the effect of home environment on achievements of students. A paper of White (1982) shows that there is a relationship between the home environment of a student and his scholastic achievements. In a meta-analysis examining 200 different studies on the effect of social economic status on achievement of students, the author found that especially home environment was a big predictor of future academic outcomes of a child. Moreover, Gottfried et al (1998) show in a longitudinal study on the roll of cognitively stimulating home environment that there is a relationship between a low social economic status, a less stimulating home environment, and lower intrinsic motivation of a child. In this study, the home environment of a child was measured at age 8 and the academic intrinsic motivation was measured at age 9, 11 and 13. The negative effects of a low stimulating home environment was present until the age of 13, pointing to a long term relationship. These relationships in which lower social economic status is related to less stimulation at home and lower intrinsic motivation of a child, could cause that the cito-test does not overcome the problem of a subjective assessment from a teacher. In this case the results of a cito-test could just reflect the social-economic status of a child, by giving the effects of lower intrinsic motivation due to a bad home environment.

An important measure for equality of opportunity is intergenerational mobility. As far as I know there is no study that looks directly at the effect of high-stake standardized tests on intergenerational mobility. This study will directly test this relationship by looking at the effect of the introduction of the cito-test on intergenerational educational mobility.

3 Data and methodology

3.1 Data description

To answer the research question whether the cito-test has an effect on intergenerational mobility, I need some specific data. The ideal approach is to collect longitudinal data on a mobility measure (i.e. income, social status, education level). However, this data is not openly accessible. I thus need to look at cohort level data to analyze the research question. With cohort-level data it is still possible to look for a causal effect of the cito-test on intergenerational mobility. This is the case when we compare birth cohorts just before and just after the implementation of the cito-test. These people should not differ that much in characteristics (which of course can be tested). There is however one major difference between these people: the first group did not make a cito-test, and part of the second group did make the test. Part of the second group, because the cito-test was not mandatory when it was introduced. At the start in 1969 there were around 35.000 pupils who made the test (CITO, 2014). Besides this data I need information on year of birth, and measures to compute an intergenerational mobility variable. With this data I can run a (fuzzy) Regression Discontinuity (RD) to exploit the causal effect of the cito-test on the mobility measure around the cut-off point of implementation of the cito-test. First, I will describe the data. To get all the information needed for the RD, I need questionnaires which contain at least the following measures:

1. Year of birth respondent
2. Education level/income/social status respondent
3. Education level/income/social status father of respondent
4. Background variables (age, gender etc.)

Also, these questionnaires must contain a representative sample of the Dutch population. In the Netherlands there were multiple national questionnaires which contain (some of) these measures. The following table gives an overview of the best questionnaires available. None of the questionnaires contains a measure for income of the father of the respondent. I will thus not use this for my study.

Table 1: Overview datasets

Variables	NKO	Arbeidspanel	Inkomens-ongelijkheid	Levensloop en Carriere	Inkomenspolitiek	Relatievormen
Year of birth Respondent	Yes	Yes	Yes	Yes	Yes	Yes
Educ. Respondent	Yes	Yes	Yes	Yes	Yes	No
Occ. level Respondent	Yes	No	No	No	Yes	No
Educ. father	Yes	Yes	Yes	Yes	Yes	No
Occ. level father	Yes	No	No	No	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes	Yes
Place of birth	No	No	No	No	No	No
Place of Residence	Yes	Yes	No	Yes	Yes	Yes
Adress Father	No	No	No	No	No	No
Observations	3846	4020	795	475	1729	1600
Frequency	Every 4 year	Every 2 year	Once	Once	Once	Once

As one can see in Table 1, when it comes to variables, the datasets NKO and Inkomenspolitiek are the most detailed datasets. However the dataset Inkomenspolitiek does not have a lot of observations and is only performed once. NKO seems a good dataset with all the relevant variables. However there are some problems with it. There are 3846 observations, which is quite a lot, but if we filter out the missing observations of the important variables, the dataset becomes a lot smaller. Moreover, we need the most observations around the cut-off point to get strong results. And although this survey is done every four years, only one of those contains the variables which I need for this study. The best dataset is the Arbeidsaanbodpanel or labor supply panel. This questionnaire is done every 2 years since 1985. It is very well documented, containing a big overview of the variables per dataset. Also, every dataset contains a variable of the first year that a respondent participated in the survey, which prevents me from double counting. Eventually I selected 5 datasets from the labor supply panel: 1994, 1996, 1998, 2000, 2002 (KNAW, 2015). Combining these datasets and deleting double observations left me with a total dataset of 5993 observations. In Table 2 one can see the relevant descriptive statistics of the data. Some of the variables contained different values in different datasets. I bundled some values together to ensure that the variables could be compared with each other. The most important variables, the education levels, now run from 1: only primary education till 5: academic education. Appendix A shows how I bundled this education level together.

For my evaluation I need some kind of measure for mobility. I created a dummy variable which has the value 1 if the education level of the respondent differs at least 1 level from that of his/her father. At first sight, this seems not a very solid measure of mobility. However, because the evaluation is done around a cut-off point, there should be no difference between this measure before and after the cut-off point. If there is, this is most likely due to the cito-test. Therefor this measure could still say something about the effect.

Table 2: Descriptive statistics

		Mean/Count	%
Intergenerational educational mobility		0.70	
Education level respondent		2.91	
Education level father of respondent		2.25	
Year and month of birth		05-JAN-1960	
Gender	Male	2957	49.3%
	Female	3036	50.7%
	Total	5993	100.0%
Province of resident	Groningen	246	4.1%
	Friesland	182	3.0%
	Drenthe	187	3.1%
	Overijssel	378	6.3%
	Gelderland	648	10.8%
	Utrecht	349	5.8%
	Noord-Holland	721	12.0%
	Zuid-Holland	824	13.8%
	Zeeland	304	5.1%
	Noord-Brabant	808	13.5%
	Limburg	368	6.1%
	Amsterdam, Rotterdam, Den Haag	734	12.3%
	Flevoland	238	4.0%
	Total	5987	100.0%
Religion	geen	2344	48.7%
	Roman catholic	1493	31.0%
	Protestant	782	16.2%
	Other christian communities	79	1.6%
	Jewish	3	0.1%
	Islamic	23	0.5%
	Buddhist	10	0.2%
	Hindu	40	0.8%
	Other	44	0.9%
	Total	4818	100.0%
Country of origin	Netherlands	5750	96.0%
	Suriname, Antils, Dutch Indies	56	0.9%
	Marocco, Turkey	14	0.2%
	Other	167	2.8%
	Unknown	3	0.1%
	Total	5990	100.0%

To obtain the cut-off point for the analysis we need to know when the first cohort which made the test was born. The first test was made in 1969, which is the cohort that went to the last group of primary school on 1 september 1968. To be allowed in primary school, one had to be 6 before the first of October of the year he or she entered primary school (Onderwijsinspectie, n.d.). This means that a person was twelve, or became twelve before the first of October, when it entered the last group of primary school. Thus if a person was born before 1 October 1956, it most likely did not make a cito-test, and when it was born after this date, the probability of making the cito became higher than zero. By how much this probability increases was not so easy to find. There is little documentation from the early years of the cito-test. However, I did find an article by (Wijnstra, 1984) which gives an historic overview of the period 1966-1980 of the cito-test. Here I did find some figures about the amount of children that made the cito-test. Next to this, there were numbers available from the website of the cito itself (CITO, 2014). Table 3 shows these figures.

Table 3: Children which made the cito-test

Year	Year of birth	CITO	Total children primary school	Total children last class	Percentage
1969	after October 1956	35000	1438800	239800	14.60%
1976	after October 1963	40000	1453500	242250	16.51%
1980	after October 1967	90000	1379900	229983	39.13%
1990	after October 1977	100000	1432800	179100	55.83%

For my analysis, I need the percentage of children in the last group of primary school which made the cito-test. This will be the probability of making the cito-test in my sample. The value can be found in the last column of Table 3. To compute this I used the total of children in primary school in a particular year (Statline, 2017). This amount divided by 6 (because there were 6 classes in primary school) gives me an estimation of the total amount of children in the last group of primary school. This leads to a percentage of 14.6% in the first row: 14.6% of all children in the last group of primary school did make the CITO-test in 1969.

3.2 Methodology

3.2.1 Regression discontinuity

With the data I have, I can perform a fuzzy Regression discontinuity design (RDD) to look for effects of the cito-test on intergenerational mobility. A fuzzy RDD exploits the difference in the outcome variable before and after a cut-off point. The idea is that close to the cut-off point, there should be no difference between characteristics of the sample before and after this point. This cut-off point is determined by a running variable. In a sharp RD setting, when the running variable exceeds the cut-off point, the treatment goes from 0 to 1. When all other characteristics stay the same, the difference in the outcome variable gives the causal effect of going from 0 to 1 (i.e. from control to treatment). Equation 1 shows the sharp RDD formally in my setting.

$$Mobility = \beta_0 + \beta_1 T + f(birthdate) + \gamma Z + \epsilon \tag{1}$$

In this equation, T gets the value 0 if a person is in the control group and 1 if a person belongs to the treatment group. In this setting treatment and control are based on whether someone was born before (control) or after (treatment) the date of introduction of the cito-test. The treatment effect β_1 will be consistent if the running variable birthdate is the only systematic determinant of T. In that case T will not be correlated with the error term ϵ because birthdate will capture any correlation between T and ϵ. One can choose different orders for the function of birthdate (polynomials). This is because the running variable, which is birthdate in my setting, could behave in other ways than just linear. It could be a quadratic or a cubic function. In my analysis I will use different orders to check whether my results are sensitive to changes in specification. If that is the case, it generally means that the results are less reliable.

A RDD comes close to a randomized experiment when there are no discontinuities at the cut-off point other than the discontinuity in the outcome variable. In this case, the assignment to treatment and control is close to random, hence the outcome is close to causal. The difference between fuzzy RDD and sharp RDD is that in the fuzzy design, the discontinuity between treatment and control is not sharp, i.e. from 0 to 1, but it is a change in the probability of treatment. The change in the probability is less than 100%. The fuzzy RD actually comes down to a special case of an Instrumental Variable analysis. The First Stage measures the change in the probability of treatment before and after the cut-off. The reduced form (RF) is actually the same as the sharp RD: the difference in the outcome variable between the control and the treatment group. The fuzzy RD estimate is the ratio between the reduced form and the first stage i.e. the change in the outcome variable between birth cohorts divided by the change in the probability of treatment. Equation 2, 3 and 4 show the fuzzy RDD formally in my setting (Imbens & Lemieux, 2008).

$$Cito = \beta_0 + \beta_1 T + f(birthdate) + \gamma Z + \epsilon \tag{2}$$

$$Mobility = \beta_0 + \beta_1 \hat{Cito} + f(birthdate) + \gamma Z + \epsilon \tag{3}$$

$$\beta_1 = \frac{\lim_{x \downarrow c} \mathbf{E}[Mobility|X = x] - \lim_{x \uparrow c} \mathbf{E}[Mobility|X = x]}{\lim_{x \downarrow c} \mathbf{E}[Cito|X = x] - \lim_{x \uparrow c} \mathbf{E}[Cito|X = x]} \tag{4}$$

Equation 2 shows the First-Stage estimate. Here I measure the change in the probability of getting the cito-test when we pass the cut-off date. T again shows whether a person is in the control or the treatment group. Then we add a smooth function of birthdate in the regression to capture the correlation between T and the error term. And Z is a vector of control variables. Equation 3 is the IV-estimate where we regress the predicted outcomes of the First-Stage on the outcome variable mobility. The computation of the coefficient β_1 of this regression is shown in equation 4. Here we divide the Reduced Form estimate by the First-Stage estimate. In the reduced form regression we treat the data as if it was a sharp RD design. We measure the expected value of Mobility when x, that is birthdate , approaches the cut-off value c. When we do this from both sides of the cut-off and subtract these values from each other, we get the change in mobility when we pass the cut-off point. This is the same measure as we got in the sharp RD setting. However, our discontinuity is not sharp: the change in the probability of treatment is less than 100%. Therefor we have to divide the reduced form estimate by the change in probability of getting treated i.e. the first stage. As said before, the first cito-test was made in 1969. For the analysis we assume that the probability of treatment before 1969 was 0. There exists a precursor of the cito-test, which means that it is possible that some people did make a version of the test before 1969. I will elaborate on this in the robustness analysis. In Table 3 one can see the amount of children who took the cito-test in different years, as well as the percentage of total children in the last group of primary school. In our sample, this percentage is the probability that a person gets treated. In the dataset, there is no information about whether someone has made a cito-test. This means we have to do the IV-analysis by hand, using the change in the probability of making the cito-test as the first stage estimate. For the analysis I use bootstrapping to get standard errors and significance tests for the computed IV-estimates. In Appendix B I will explain more on how bootstrapping works.

There are a couple of important assumptions in a RDD which have to be tested to make sure that the outcomes are reliable. First of all, assignment to treatment must vary discontinuous at the cut-off point. As said before, I assume that before introduction no one made a cito-test, so after introduction the assignment to treatment rises from 0 to 14.6%. The second important assumption is that there exists no sorting at the cut-off. If individuals have a great deal of control over the assignment variable and if there is a perceived benefit of treatment, one would expect individuals on one side of the threshold to be systematically different from those on the other side. In this setup this is not very likely to occur. It seems unlikely that individuals will move to another primary school because they can make a cito-test. However, this can be tested using the Mcrary density test, which will be explained later. The third assumptions states that all other characteristics of the individuals in the sample remain continuous at the point of treatment assignment. If some of the other characteristics are also discontinuous at the cut-off, the change in the outcome variable could be due to the change in this characteristic rather than due to the cito-test.

3.2.2 Linear probability model

Because the outcome variable, mobility, is binary, I need to use a probability model to come to the results. I use a Linear Probability Model (LPM), as this gives me coefficients that are easy to interpret. The expected value of the binary outcome variable mobility is shown in equation 5.

$$E[Mobility] = 1 \cdot Pr(Mobility = 1) + 0 \cdot Pr(Mobility = 0) = Pr(Mobility = 1) \tag{5}$$

Including the independent variables this gives equation 6:

$$E[Mobility|T, Z] = Pr(Mobiliy = 1|T, Z) = \beta_0 + \beta_1 T + \gamma Z \tag{6}$$

The coefficient β_1 is equal to the change in the probability that $Mobility = 1$ associated with a unit change in T (which is going from control to treatment). This is shown in equation 7.

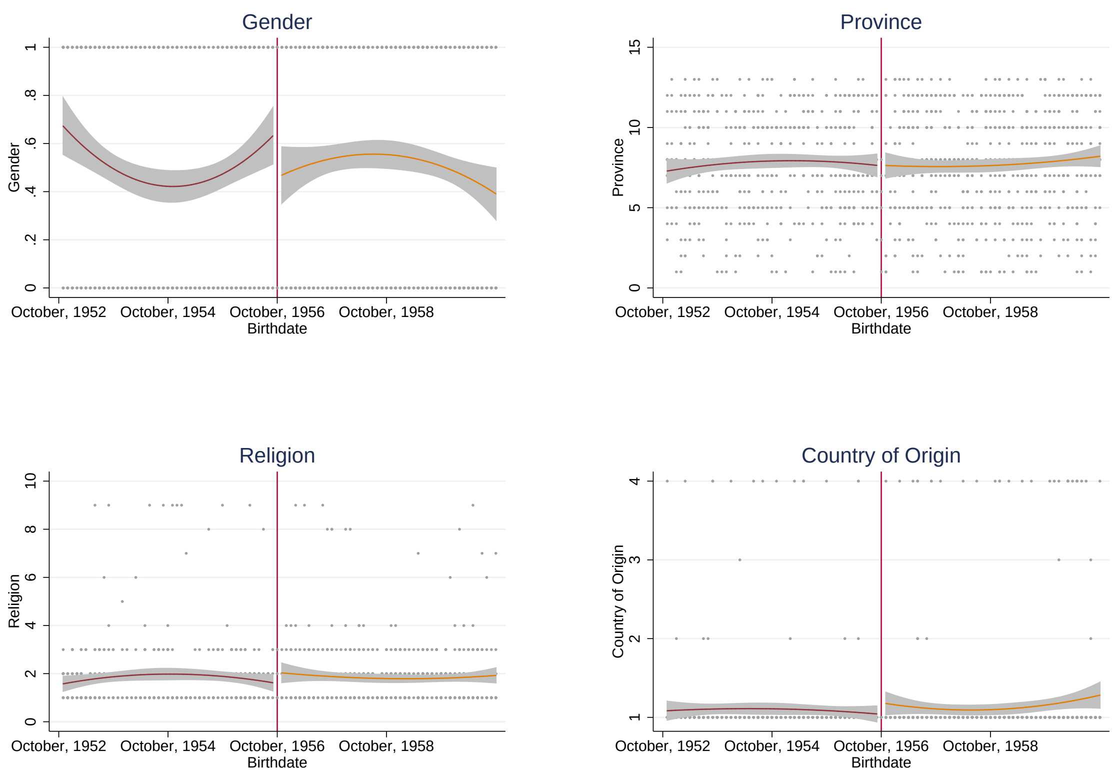
$$\frac{\partial Pr(Mobiliy = 1|T, Z)}{\partial T} \tag{7}$$

The LPM is useful in that the outcomes are easy to interpret. β_1 is equal to a percentage point change in the outcome variable. The model also has its limitations. First, the estimation imposes heteroskedastic error terms. This is solved by using robust standard errors. The second shortcoming is not so easy to solve. The estimates of the model are not constrained by the unit interval because it the LPM fits a linear line. This means that it can measure values that are greater than 1 or less than 0. Following Angrist and Pischke (2008) which state that when it comes to marginal effects, using a linear or nonlinear model matters little, I will still use the LPM in my study.

3.2.3 Continuity checks

We can test for continuity by using the characteristics as dependent variables in the regression. First, I will show a visual representation of the continuity in the control variables. Figure 1 shows the (dis)continuity at the cut-off (October 1956) for the control variables.

Figure 1: Continuity checks control variables



As one can see most of the variables are more or less continuous at the cut-off. The variable Gender shows a small discontinuity at the cut-off. We can check whether this discontinuity is significant by using the control variables as dependent variables in the regression. If the treatment variable has a significant impact on the control variable, this means that there is a significant discontinuity in this control variable at the cut-off. Table 4 shows the outcomes of these regressions.

Table 4: Continuity checks

	(1) Province	(2) Gender	(3) Country of origin	(4) Religion
Treatment	-1.58*	-0.06	0.01	0.12
	[0.806]	[0.127]	[0.131]	[0.394]
Birthdate	0.00**	0.00	0.00	-0.00
	[0.002]	[0.000]	[0.000]	[0.001]
Constant	13.02***	0.69*	1.37***	1.71
	[2.546]	[0.409]	[0.473]	[1.354]
Observations	311	311	310	256
R-squared	0.014	0.001	0.008	0.001
Robust errors	Yes	Yes	Yes	Yes

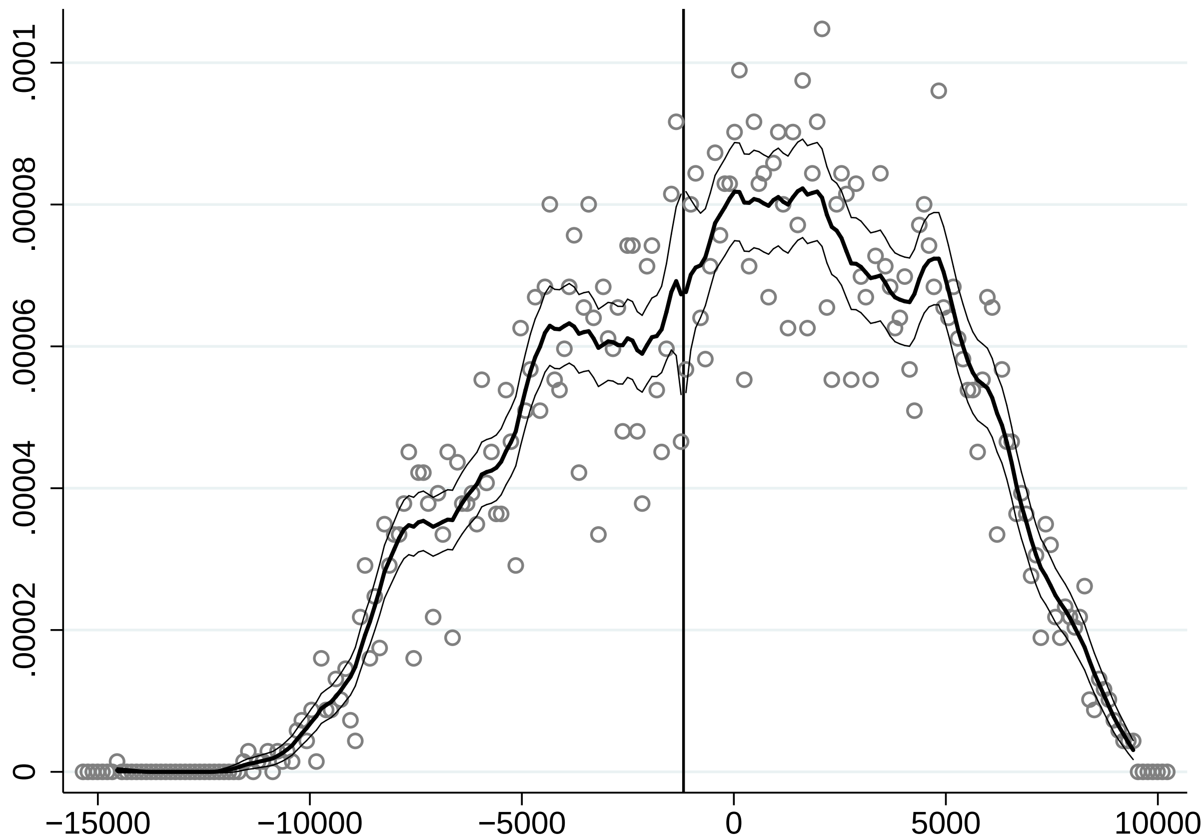
Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

As one can see from the table, the variable Gender does not show a significant change at the cut-off. However, the variable 'Province' does seem to be slightly discontinuous (at the 10% significance level). This could be a problem in our analysis if we want to interpret the effect of the cito-test on mobility. However, I can include these variables in our model to control for these effects. The other variables do not have a discontinuity at the cut-off. I will still add these variables in the model to get a better precision of the estimation.

3.2.4 Density test

To check whether people have control over assignment to treatment or control group, I will use the McCrary density test (McCrary, 2008). This test estimates whether there exists a discontinuity at the cutoff in the density function of the running variable. This discontinuity is calculated in two steps. The first step computes a histogram of the density, which is smoothed using local linear regression, separately on either side of the cutoff in the second step. In this way one obtains a difference in height of the histogram at the cut-off. Graph 2 shows the results of the density test for the cut-off of October 1956. The log difference in height (i.e. the discontinuity) is estimated to be -.0029 with a standard deviation of 0.18. A t-test points out that this is not significantly different than zero, i.e. there is no significant discontinuity in the density of the running variable at the cut-off.

Figure 2: Density test



Notes: The vertical axis represents the log difference in height. The horizontal axis equals the birthdate with 0 = 1 January 1960. The vertical line represents the cut-off date: 1 October 1956.

4 Results

4.1 Intergenerational mobility matrix

This section will show the results of the fuzzy regression discontinuity analysis for the effect of the cito-test on intergenerational mobility. First I want to give an insight on the overall intergenerational mobility within the sample. I will do this by showing the transition matrix of education levels. Table 5 shows this matrix. On the left side one can see the education level of the father. When there exists perfect mobility, the childs education level does not depend on the fathers education level and is thus equally distributed over all education levels. In other words, when the education level of the father is 1, the probability of having the same or any of the other education levels for the child is 20% with perfect mobility. When the fathers education level is 1 and the child has a probability of more than 20% to also have the first level of education, than we can speak of intergenerational persistence: the education level of the child depends on the education level of the father.

Table 5: Intergenerational mobility matrix

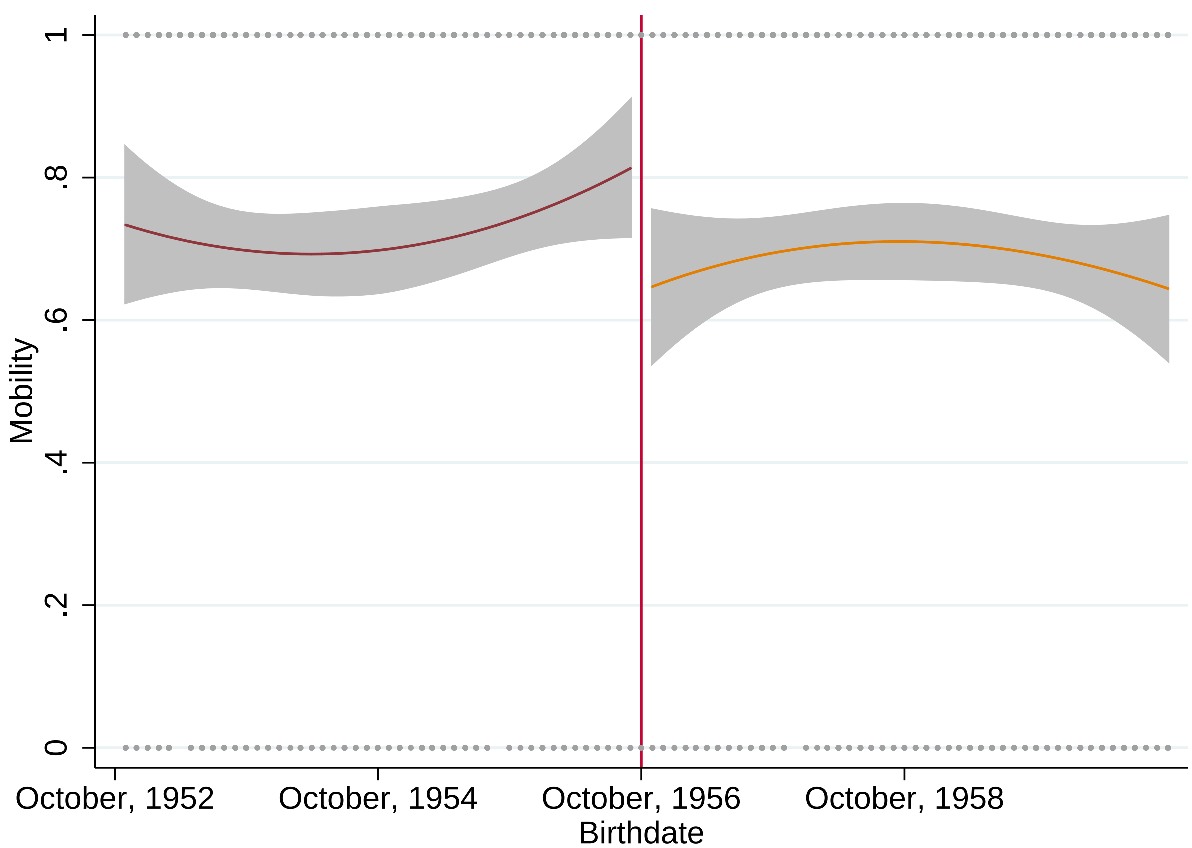
		Childs'	education	level	
Fathers' education level	1	2	3	4	5
1	14.1%	44.5%	26.0%	12.2%	3.3%
2	3.7%	36.5%	36.5%	18.3%	5.0%
3	3.6%	21.1%	37.2%	27.6%	10.6%
4	3.1%	14.3%	35.8%	32.2%	14.7%
5	1.2%	7.0%	31.3%	31.7%	29.0%

Most of the education levels show a percentage of more than 20% in the same education levels of father and child. This means that for most education levels there is no perfect mobility but there is some kind of persistence. Also for almost all education levels of the father, the child stays in the same or one level higher or lower in around 60% of the cases. With perfect mobility this should be 40%. It is also good to look at the extremes. When the father has an education level of 1, only 3.3% of the children make it to level 5. And when the father has level 5, only 1.2% of the children drop to level 1. This shows that it is really hard to get a level of education far away from ones father. As one might notice, the persisentence is not so high in the first level of education of the father. Most children go to the second level of education when their father was in the first level. The reason behind this is pretty intuitive. When the fathers followed education, it was quite normal to start working directly after primary school. This changed during the 20th century as there was more focus on the importance of education. In the course of the 20th century there came laws that obliged children to follow education for a given amount of years. This amount of years increased in every new law. That is why in my sample there are more fathers than children that had only primary school. Still there is persistence in this group as most of these children did not get further than the second level of education.

4.2 Reduced form estimates

Next, I will show the results of the reduced form estimation. These results can be found in Table 6. I use a bandwidth of 8 years in total: 4 years before the cut-off and 4 years after. This bandwidth is chosen using an optimal bandwidth selector. I also checked whether the results are consistent using different bandwidths. The estimates increase a little when I narrow the bandwidth, but the sign and significance of the estimates stay the same. One can see that I have analyzed 4 different models. The models differ in their choice of polynomial. Changing the polynomial orders have little impact on the outcome. This is a good thing because it means that the results are not sensitive to a change in the specification. This makes the model more reliable (Lee & Lemieux, 2010). Changing the polynomial order form 0 to 1 and 2 results in a small increase of the impact and a small decrease of the significance. The p-values in these models lie just above the 5 percent level (0.052 and 0.051 resepctively). When we change the specification to a cubic function, we see a much bigger estimate which is largely significant. To know which of these models is best, I looked at Aikaikes Information criterion (AIC) as suggested by Lee and Lemieux (2010). The optimal model with regard to the polynomial level, i.e. the model with the lowest AIC, turned out to be the model with polynomial of order 0. However, the differences in the AICs are really small.

Figure 3: Mobility



Let us interpret the results. Graph 3 shows the reduced form results graphically. One can see that at the cut-off there is a small decrease in the probability of being mobile. After the cut-off, people are less likely to be mobile. When we look at the actual regression estimates in Table 6, we see that, as expected, most of the control variables do not have a significant effect on the outcome variable. The only variable that has a small effect is the variable province, which we also could expect from the continuity analysis. We see that living in Flevoland lowers the chance of having a different education level than ones father after the cut-off date. Because of the low amount of observations in this province (45), this effect has no real meaning. The treatment effect shows that, for any person within the sample, being born after the cut-off date lowers the chance of having mobility equal to one, i.e. having a different education level than his or her father. This means that on average intergenerational mobility slightly decreased due to introduction of the cito-test. In our most preferred model we see an estimate of -0.07. This means that the probability of having mobility equal to one decreased with 7 percentage points when we pass the cut-off date. People born after 1 October 1956 thus have a lower chance of having a different education level than ones father than people born before 1 October 1956. The probability of being mobile before the cut-off was 72.7% on average. This means that, holding everything else constant, being born after 1 October 1956 reduces the chance of being mobile to 65.7%.

Table 6: Reduced form estimates

	(1) RF 1952-1960	(2) RF 1952-1960	(3) RF 1952-1960	(4) RF 1952-1960
Treatment	-0.07**	-0.11*	-0.11*	-0.22***
	[0.029]	[0.058]	[0.058]	[0.078]
Birthdate		0.00	0.00	-0.00
		[0.000]	[0.000]	[0.000]
Birthdateˆ2			-0.00	-0.00**
			[0.000]	[0.000]
Birthdateˆ3				-0.00**
				[0.000]
Gender (Female=1)	0.02	0.02	0.02	0.01
	[0.029]	[0.030]	[0.030]	[0.030]
Province (reference=Groningen)				
Friesland	-0.03	-0.03	-0.03	-0.01
	[0.118]	[0.118]	[0.118]	[0.121]
Drenthe	0.05	0.05	0.05	0.06
	[0.101]	[0.102]	[0.102]	[0.101]
Overijssel	-0.09	-0.10	-0.09	-0.09
	[0.103]	[0.103]	[0.103]	[0.103]
Gelderland	-0.03	-0.03	-0.03	-0.02
	[0.088]	[0.089]	[0.089]	[0.088]
Utrecht	-0.08	-0.08	-0.08	-0.08
	[0.101]	[0.101]	[0.101]	[0.101]
Noord-Holland	-0.08	-0.08	-0.08	-0.08
	[0.090]	[0.090]	[0.090]	[0.090]
Zuid-Holland	-0.07	-0.07	-0.07	-0.07
	[0.087]	[0.088]	[0.088]	[0.087]
Zeeland	-0.04	-0.04	-0.04	-0.03
	[0.102]	[0.103]	[0.103]	[0.102]
Noord-Brabant	-0.03	-0.03	-0.03	-0.02
	[0.087]	[0.088]	[0.088]	[0.087]
Limburg	-0.12	-0.12	-0.12	-0.12
	[0.095]	[0.096]	[0.096]	[0.096]
Flevoland	-0.22*	-0.22*	-0.22*	-0.21*
	[0.120]	[0.121]	[0.121]	[0.120]
Three large cities	-0.04	-0.04	-0.04	-0.04
	[0.093]	[0.094]	[0.094]	[0.093]
Religion (reference=no religion)				
Roman catholic	-0.03	-0.03	-0.03	-0.03
	[0.035]	[0.035]	[0.035]	[0.035]
Protestant	0.06	0.05	0.05	0.06
	[0.043]	[0.043]	[0.043]	[0.043]
Other christian communities	-0.00	0.00	-0.00	0.00
	[0.109]	[0.108]	[0.108]	[0.107]
Hindu	0.05	0.05	0.05	0.04
	[0.157]	[0.158]	[0.158]	[0.157]
Other	-0.21	-0.21	-0.21	-0.19
	[0.141]	[0.142]	[0.141]	[0.142]
Country of origin (reference=Netherlands)				
Suriname, Antils, Dutch Indies	-0.01	-0.01	-0.01	-0.01
	[0.167]	[0.168]	[0.169]	[0.172]
Marocco, Turkey	0.05	0.05	0.05	0.09
	[0.078]	[0.079]	[0.079]	[0.080]
Other countries	0.04	0.04	0.04	0.04
	[0.080]	[0.079]	[0.080]	[0.078]
Constant	0.80***	0.86***	0.85***	0.95***
	[0.084]	[0.109]	[0.112]	[0.120]
Observations	978	978	978	978
R-squared	0.032	0.033	0.033	0.037
Model	Constant	Linear	Quadratic	Cubic

Robust standard errors in brackets *** p<0.01, ** p<0.05, * p<0.1

4.3 IV-estimates

The reduced form just compares mobility before and after the cut-off dates. However, we have a situation where not everyone did make the cito-test after introduction. This is why we need to divide the reduced form by the first stage to get to the true effect of the introduction of the cito-test. Naturally, these effects will be larger as we only look at a fraction of all the people in the sample. Table 7 shows the results of the IV estimates. One can see that the estimate are indeed larger. In fact the estimates should be larger by a factor 6.8 which is one divided by the increase in probability of making the cito, i.e. 14.6%. If one looks carefully, the estimates are not exactly the same as the reduced form estimates divided by the first stage. This is because in the reduced form I used a weighting scheme, giving higher weights to the observations closer to the cut-off and lower weights to observations further away. However, bootstrapping does not allow to put in these weight, as it computes its own weights. Therefor the estimates are a bit different than one would expect. This difference is however very small, so keeping in mind this difference, I use the outcomes as shown in Table 7.

Table 7: IV-estimates cito-test on intergenerational mobility

	(1) IV 1952-1960	(2) IV 1952-1960	(3) IV 1952-1960	(4) IV 1952-1960
Treatment	-0.50**	-0.75*	-0.75*	-1.47***
	[0.198]	[0.409]	[0.405]	[0.541]
Observations	978	978	978	978
Replications	2230	2231	2240	2252
Model	Constant	Linear	Quadratic	Cubic

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Notes: The number of replications stands for the amount of models the bootstrap method estimates. Generally the more replications, the more reliable the standard errors become (Efron & Tibshirani, 1994).

When we look at our most preferred model (the constant model), the cito-test lowers the probability of having a different education level than his or her father with 50% percentage points. This means that this probability goes from 72.7% to 22.7%. This results is quite big and surprising. If we look at the cubic model, this gives an unreasonably high estimation, outside the range of probabilities (which is between 0 and 1). This is a drawback of the Linear Probability Model as explained in section 3. It seems that the IV-estimates are a bit overestimated. Nevertheless, it is clear that there is a negative effect of the introduction of the cito-test on intergenerational mobility. Looking at the intentions of the cito-test this is a contrary effect. The cito-test intended to reduce the subjective judgements of the primary schools in sending a child to secondary education. The objective test had to help the primary school in making the advice more just for the children. The advice had to be independent of which social background a child had, making the chances for children more equal. The results point into the direction that the introduction of cito-test did not achieve this equality; it made it even worse.

4.4 Different education levels

In this subsection I will analyze whether there exists difference between education levels. The question is whether the effect of the cito-test differs for persons which father had a lower level compared to persons which father had a higher level of education. First, I will estimate a model using a dummy for higher/lower education. Lower education means that the highest education a person has had was intermediate vocational education (MBO)[2]. Higher education is everything above this education level. Second, I will estimate whether there is a difference in the change of the probability of being mobile for the different education levels. This is done by restricting the sample to persons whose father had lower and higher education respectively. Table 8 reports the results. I used the model which was most preferred in the previous section, which was the constant model.

Table 8: Different education levels

	(1) Higher vs Lower education	(2) Lower education	(3) Higher education
Treatment	-0.53***	-0.77***	-0.21
	[0.028]	[0.037]	[0.039]
Higher education	0.24***		
	[0.027]		
Constant	0.74***	0.77***	0.94***
	[0.083]	[0.111]	[0.101]
Observations	978	673	305
R-squared	0.091	0.045	0.085
Robust errors	Yes	Yes	Yes

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Notes: the models contain all the other relevant control variables, but for practical purposes these are not shown in this table. All the models use the bandwidth of 4 years before and 4 years after.

These outcomes all point in the direction that the results are worse for lower education segments. When we look at the first model, we can see that there exists a positive and significant effect of higher education. This means that within the sample the probability of being mobile is 24 percentage points higher for persons which father had higher education than for persons which father had lower education. Looking at the second and third model we see that this gap in mobility between higher and lower education becomes even bigger after introduction of the cito test. Persons which father had higher education do not see a significant change in their probability of being mobile due to the cito-test, while people

[2]This includes having done pre-university education (VWO) but no successive education thereafter

with lower educated fathers see a drop in their probability of being mobile of 77 percentage points. This means as a child it is quite easy to get a different education level than ones father when the father had higher education, while it is harder to get a different level of education when the father had lower education.

4.5 Robustness analysis

4.5.1 Placebo

It has already been shown that the results are quite robust to changing the bandwidths and polynomial orders. In this section I will give some more attention to the robustness of the results. I will first show some placebo results. These are the results when we randomly take a date and put the cut-off at that birthdate. Naturally, we do not expect being born after that randomly chosen date to have an effect on the outcome variable. If this is the case, then there might be something wrong with the data. In Table 9 I estimate the effects around 3 different cut-offs. The estimations are the reduced form estimates, as there is no first stage to divide by in these placebo tests. The first cut-off is set at 8 years before the cut-off that I use in the actual regression. Here, I thus estimate the effect of being born after 1 October 1948. The second cut-off is set at 1 January 1957. This cut-off is set in the same school year as our real cut-off (the school year runs from September 1956- August 1957). The idea for setting the actual cut-off at 1 October 1956 is that when a person is born after this date, it has to go to school on year later than when it was born before this date. By setting a placebo cut-off later in the same school year, I can check whether the effect I estimate is really due to being in a different school year. The last date is set one year before the actual cut-off date. This is not entirely randomly chosen. It is chosen because in 1968 the socalled Mammoetwet (Mammothlaw) was introduced. This law was introduced to change the whole structure of secondary education in the Netherlands. Although the law did not intent to do something about the inequality between social classes, such a large reform could of course have an effect. As said before, the law was executed in 1968, which means that the students who went to the first class of secondary education in 1968 were the first cohort that experienced this reform. Looking at birth cohorts, this was the birth cohort one year before the birth cohort in my analysis: born after 1 October 1955. In the last placebo test I thus use this cut-off date. As one can see in the table, all the effects are not significantly different than zero. This means that for all the placebo cut-offs, being born after the cut-off does not have an effect on mobility. Obtaining no effects at the placebo cut-offs adds to the robustness of the outcome of the actual analysis.

Table 9: Placebo test

	(1) Placebo 1 October 1948	(2) Placebo 1 January 1957	(3) Placebo 1 October 1955
Treatment	0.00	-0.04	0.01
	[0.032]	[0.030]	[0.030]
Constant	0.55***	0.78***	0.75***
	[0.085]	[0.083]	[0.084]
Observations	821	945	937
R-squared	0.028	0.027	0.026
Robust errors	Yes	Yes	Yes

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Notes: the models contain all the other relevant control variables, but for practical purposes these are not shown in this table. All the models use the bandwidth of 4 years before and 4 years after.

4.5.2 Continuous probabilities

In this subsection I will look at the effect of adding the different probabilities of making the cito-test as a continuous explanatory variable. In Table 3 in Section 3 I show the different probabilities of making the cito-test for different years. To look for some overall effect of making the cito test for people in the dataset, I will estimate a model in which these probabilities explain mobility. I made a variable which contains the different probabilities for the different birth cohorts. I treat the probabilities as if they were continuous by putting them in the model as a continuous variable. Table 10 shows the outcome of this model.

As one can see, the independent variable does not have a significant effect on mobility. This could be expected, because there is not a lot of variation in the probability of making the cito-test (there are only 5 known probabilities). However, also here we see a negative sign. An increase in the probability of making the cito-test has, if anything, a negative effect on mobility. This again reinforces the results from the IV-estimation.

Table 10: Continuous probabilities

	(1) Continuous probabilities
Probability of making the cito-test	-0.04
	[0.034]
Constant	0.71***
	[0.035]
Observations	4,815
R-squared	0.008
Robust errors	Yes

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Notes: the model contains all the other relevant control variables, but for practical purposes these are not shown in this table.

4.5.3 Amsterdam school test

As said before, the cito-test was not the first standardized test for children in the last group of primary school in the Netherlands. The precursor of the cito-test was the socalled Amsterdamse Schooltest (Amsterdam school test). This test was held only at primary schools in Amsterdam from 1966-1968 (see section 2.2 for more information). This precursor of the cito-test causes that some children made some kind of cito-test before the introduction in 1969. The percentage of children which made the test before introduction is thus a bit higher than zero. To see whether this could change the results, I look at two things. First, I check whether the results change when I leave people from the three big cities (including Amsterdam) out of the regression. Table 11 shows that this is not the case. The IV estimate stays roughly the same. Second, I check whether the effect I found for the cito-test, can also be found for the Amsterdam school test. Appendix C contains the table where all regions have been regressed separately. Because the First Stage is unknown in this case, I only show the Reduced Form estimates. The outcomes of these models strongly substantiate my results. As one can see, when we compare the cohort just before the introduction of the Amsterdam school test with the cohort just after implementation, we see a drop in intergenerational educational mobility. This effect is only present for the three big cities, which makes it more likely that the effect is due to the introduction of the precursor of the cito-test.

Table 11: IV without 3 big cities

	(1) IV without Amsterdam
Treatment	-0.50**
	[0.212]
Observations	897
Robust errors	Yes
Model	Constant

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1. Notes: the model contains all the other relevant control variables, but for practical purposes these are not shown in this table.

5 Conclusion and discussion

This paper sets out to study the effect of the cito-test on intergenerational educational mobility. As stated in the introduction, the cito test is a highly debated topic in the Netherlands. In the course of history there have been multiple changes to make the test either more important or less. My understanding is that no study before has looked directly at the effect of the cito-test on intergenerational mobility. This study tries so by using the introduction of the cito-test as some kind of natural experiment to compare cohorts just before and just after the introduction of the test. A fuzzy regression discontinuity analysis was used to come to the results. The results are quite remarkable. The founding father of the cito-test, Professor A.D. de Groot, wanted a standardized test to make sure that the advice to a child to which level of secondary education became more just. My results show that the cito-test did not fulfill this goal, by all means not when it comes to intergenerational mobility. The results show that the introduction of the cito-test dropped the intergenerational mobility of persons significantly. This negative effect especially hit people with parents which had a low level of education. This means that the cito-test caused people which father had a low level of education to be more depended on their fathers than before the cito-test was introduced. The robustness analyses strengthen the results in showing that a placebo effect was not present, and that the precursor of the cito-test also had a negative effect on intergenerational mobility. The overall conclusion that can be drawn from these results is that the introduction of the cito-test worsened the equality of opportunities, especially for people which father had a low level of education. This conclusion might support the argument that a lower social economic status leads to less support from home and worsens the intrinsic motivation to make tests (Gottfried et al., 1998). Also, it seems that the advice of the teacher (which was the situation before the introduction of the cito-test) is better with regard to intergenerational mobility than the cito-test. This could support the results of the CPB discussion paper on whether the teacher beats the test (Feron et al., 2015). Despite the results being quite strong, one should be careful extrapolating these results to draw conclusions about the cito-test nowadays. The content as well as the importance of the cito-test has changed during the years. However, the statements of Oosterbaan (1973) which explain the procedure of the formation of the advice in 1968 indicates that not so much changed after all. One should be cautious in following the founder of the test in saying that a standardized test makes the advice to children more just. My study has of course some shortcomings. The data could be much better had it contained a variable with information about whether or not a respondent made the cito-test. With that information, one could perform the IV-estimation without the bootstrap-method giving somewhat more reliable results. Further research should look at whether the negative effect that I have shown in this study is also present nowadays. The results of this study certainly raise some questions about whether a standardized test is good for everyone. Maybe further research could use more detailed data, e.g. school-level data, to look whether the introduction of the cito-test on specific schools had an effect on the intergenerational mobility of children on that school. If those results coincide with the results of this study, I think we could add some thorough opinions to the dispersed debate about the cito-test.

References

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education policy analysis archives*, *10*, 18.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, *87*(6), 1153–1189.

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of labor economics*, *4*(3, Part 2), S1–S39.

Becker, H. S. (1952). Social-class variations in the teacher-pupil relationship. *The Journal of Educational Sociology*, *25*(8), 451–465.

Behrman, J. R., & Taubman, P. (1990). The intergenerational correlation between childrens adult earnings and their parents income: Results from the michigan panel survey of income dynamics. *Review of Income and Wealth*, *36*(2), 115–127.

Black, S. E., & Devereux, P. J. (2010). *Recent developments in intergenerational mobility* (Tech. Rep.). National Bureau of Economic Research.

Blanden, J., Gregg, P., & Macmillan, L. (2007). Accounting for intergenerational income persistence: noncognitive skills, ability and education. *The Economic Journal*, *117*(519).

CITO. (2014). De geschiedenis van de cito-toets.

Corak, M. (2004). *Generational income mobility in north america and europe*. Cambridge University Press.

de Groot, A. D. (1966). *Vijven en zessen: cijfers en beslissingen: het selectieproces in ons onderwijs*. JB Wolters.

Douglas, J. W. (1964). The home and the school: a study of ability and attainment in the primary school.

Driessen, G. (2006). De totstandkoming van de adviezen voortgezet onderwijs: Invloeden van thuis en school. *Pedagogiek*, *25*(4), 279–298.

Driessen, G. (2011). Over-en onderadvisering bij de overgang naar het voortgezet onderwijs.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Feron, E., Schils, T., & ter Weel, B. (2015). Does the teacher beat the test?

Fox, L., Torche, F., & Waldfogel, J. (2016). Intergenerational mobility. In *The oxford handbook of the social science of poverty.*

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1998). Role of cognitively stimulating home environment in children's academic intrinsic motivation: A longitudinal study. *Child development*, *69*(5), 1448–1460.

Grawe, N. D. (2004). Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints. *Journal of Human Resources*, *39*(3), 813–827.

Grotenhuis, H. t., & Dronkers, J. (1989). Enkele gevolgen van werkloosheid en arbeidsongeschiktheid in de verzorgingsstaat. ongelijke onderwijskansen van kinderen. *Amsterdams Sociologisch Tijdschrift*, *15*(4), 634–651.

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, *30*(2), 75–110.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, *142*(2), 615–635.

Jungbluth, P. (2003). De ongelijke basisschool. *Nijmegen, ITS, te verschijnen*.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (Tech. Rep.). National Bureau of Economic Research.

KNAW. (2015). Arbeidsaanbodpanel.

Kotera, T., & Seshadri, A. (2017). Educational policy and intergenerational mobility. *Review of Economic Dynamics*, *25*, 187–207.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.

Lubbe, M. (2005). van der, verhelst, n., heuvelmans, t., & staphorsius, g. *Verslag van een onderzoek naar de toelating van leerlingen in het voortgezet onderwijs*.

Madaus, G. F., & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data.

Marchant, G. J., & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and sat scores. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, *13*.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. *The Review of Economics and Statistics*, *87*(2), 235–255.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, *142*(2), 698–714.

Mulligan, C. B. (1997). *Parental priorities and economic inequality.* University of Chicago Press.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2005). High-stakes testing and student achievement: Problems for the no child left behind act. appendices. *Education Policy Research Unit*.

NRC. (2014). Leraar weet beter wat goed is voor kind dan citotoets.

Onderwijsinspectie. (n.d.). Vragen en antwoorden.

Onderwijsinspectie. (2014). De kwaliteit van het basisschooladvies.

Oosterbaan, J. (1973). De amsterdamse schooltoets 1969 in relatie tot enige conclusies van'het verborgen talent'. *Sociologische Gids*, *20*(2), 88–97.

Österberg, T. (2000). Intergenerational income mobility in sweden: What do tax-data show? *Review of Income and Wealth*, *46*(4), 421–436.

Piraino, P., et al. (2006). *Comparable estimates of intergenerational income mobility in italy.* Universita degli Studi di Siena.

Prokic-Breuer, T., & Dronkers, J. (2012). The high performance of dutch and flemish 15-year-old native pupils: Explaining country differences in math scores between highly stratified educational systems. *Educational Research and Evaluation*, *18*(8), 749–777.

Regt, A. d. (2004). 'welkom in de ratrace.'over de dwang van de cito-toets. *Amsterdams Sociologisch Tijdschrift*, *31*(3), 297–320.

Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S., & Zabala, D. (2006). From the capital to the classroom: Year 4 of the no child left behind act. *Washington, DC: Center on Education Policy*.

Rijksoverheid. (2018). Verplichte eindtoets basisonderwijs.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard educational review*, *40*(3), 411–451.

Singh, K. (1981). On the asymptotic accuracy of efron's bootstrap. *The Annals of Statistics*, 1187–1195.

Singh, K., & Xie, M. (2008). Bootstrap: a statistical method. *Unpublished manuscript, Rutgers University, USA.*.

Statline, C. (2017). Leerlingen, deelnemers en studenten; onderwijssoort, vanaf 1900.

Van der Hoeven-van Doornum, A. (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen: een studie naar leerkrachtverwachtingen, leerkrachteffectiviteit en schooleffectiviteit.* Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Vandrick, S. (2014). The role of social class in english language education. *Journal of Language, Identity & Education*, *13*(2), 85–91.

Volkskrant. (2014). De citotoets is achterhaald, schaf alfjeblieft direct af.

Volkskrant. (2015). Je koopt makkelijker een leerkracht om dan het cito.

Volkskrant. (2016). Citotoets is ook voor docenten steun in de rug.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological bulletin*, *91*(3), 461.

Wijnstra, J. (1984). *Verantwoording eindtoets basisonderwijs 1984*. Cito.

WVO, I. (2018, 1 January). Consulted on 16 January 2018, from http://wetten.overheid.nl/BWBR0005946/2018-01-01.

A Adjustment education levels

Table 12: Adjustments education levels

Comprehensive Education level	Adjusted education level
Lagere school	Primary education
Ambachtsschool Huishoudschool Lagere land- en /of tuinbouwschool LBO: Lager Beroepsonderwijs ULO: Uitgebreid Lager Onderwijs MULO: Meer Uitgebreid Lager Onderwijs HULO: Handels ULO MAVO: Midelbaar Algemeen Voortgezet Onderwijs	Lower vocational education
HAVO:Hoger Algemeen Voortgezet Onderwijs HAVO-top /-afdeling op pedagogische afdeling Eerste leerkring, kweekschool Handelsdag of avondschool MMS: Middelbare meisjesschool HBS: hogere burgerschool Lyceum Gymnasium Atheneum VWO:Voorbereidend Wetenschappelijk Onderwijs Kweekschool Uitgebreid Lager Beroepsonderwijs (tot 1968) KBMO: Kort Middelbaar Beroepsonderwijs MBO: Middelbaar beroepsonderwijs	Intermediate vocational education/ higher general secondary education/ pre-university education
HBO: Hoger Beroepsonderwijs	Higher Vocational Education
Wetenschappelijk onderwijs-kandidaats Wetenschappelijk onderwijs-doctoraal	University education

B Bootstrap method

The bootstrap method was first introduced by Efron (1979). The name originates from the phrase 'Pull yourself up by your bootstraps'. This is something which is actually not possible to do: you cannot lift yourself up by pulling your shoes. The bootstrap method in statistics also seems to be impossible. Bootstrapping gets more information out of a sample by resampling it with replacement. To understand this better, suppose we have a population from which we want to know some statistics (e.g. mean, standard deviation etc.). If we could draw a large number of samples from this population, we could get a reasonably good idea about the distribution of the statistics of interest. Of course, in real life we cannot draw a large number of samples from a population. Most of the time we just have one sample to work with. Bootstrapping allows us to still get a fairly good idea about the distribution of the statistics, by drawing a large number of random samples (with replacement) from the original sample. Let us look at some basic theoretical support for bootstrapping.

Say that we want to know something about the population parameter θ which could be for example the mean wage of a region. Suppose we have a random sample of size n of the population which gives us the date $(X_1, X_2, , X_n)$. The sample statistic which we get out of this data equals $\hat{\theta}$. Here, the Central Limit Theory tells us that for large n the sampling distribution of $\hat{\theta}$ is bell shaped with center θ and standard deviation $\left(\frac{a}{\sqrt{n}}\right)$, where is a positive number which depends on the population. Bootstrapping takes a random sample out of the original sample. So let us say the first bootstrap sample is $(X_1, X_1, X_4, X_6, X_n)$. Note that in this sample the observation X_1 occurs twice while the observation X_2 does not occur at all. This is because the bootstrap takes a sample with replacement: it takes a value, puts it back and then takes another value. Suppose that the sample statistic which we get out of this sample equals $\hat{\theta_B}$. It turns out that when $n \to \infty$ the sampling distribution of $\hat{\theta_B}$ is also bell shaped and has the sample statistic from the original sample, $\hat{\theta}$, as its center. It also has the same standard deviation $\left(\frac{a}{\sqrt{n}}\right)$. We thus get that the bootstrap distribution of $\hat{\theta_B} - \hat{\theta}$ comes close to the sampling distribution of $\hat{\theta} - \theta$ (Singh & Xie, 2008). This phenomenon is called the bootstrap Central Limit Theory. The proof for this is too comprehensive for this moment, but it can be found in Singh (1981).

In this study I use the bootstrap technique to compute the IV-estimate. As shown in equation 4, the IV-estimate equals $\beta_1 = \frac{\lim_{x \downarrow c} \mathbf{E}[Mobility|X=x] - \lim_{x \uparrow c} \mathbf{E}[Mobility|X=x]}{\lim_{x \downarrow c} \mathbf{E}[Cito|X=x] - \lim_{x \uparrow c} \mathbf{E}[Cito|X=x]}$, which is the Reduced Form divided by the First Stage. The Reduced Form can be estimated with the data I have, however the First Stage cannot. Although I cannot get the First Stage out of the data, I do have a good indication of how large it could be. It namely is the percentage of children which made the cito-test after the introduction minus before the introduction. Assuming that the percentage of children which made the cito-test before introduction is 0, the First Stage is 14.6%. Using the bootstrap method, I compute the IV-estimate by hand by first estimating the Reduced Form and then bootstrapping the Reduced From divided by the First Stage. This gives me standard errors for the IV-estimate, by which I can assess the significance of the results.

C Amsterdam school test

Table 13: Reduced form estimates Amsterdam school test

	(1) Groningen	(2) Friesland	(3) Drenthe	(4) Overijssel	(5) Gelderland	(6) Utrecht	(7) Noord-Holland	(8) Zuid-Holland	(9) Zeeland	(10) Noord-Brabant	(11) Limburg	(12) Amsterdam, Rotterdam, The Hague	(13) Flevoland
Treatment	0.03	0.49*	0.14	0.06	-0.06	0.04	-0.09	-0.01	0.14	0.04	0.02	-0.20**	-0.19
	[0.209]	[0.262]	[0.203]	[0.143]	[0.089]	[0.124]	[0.091]	[0.094]	[0.148]	[0.080]	[0.115]	[0.096]	[0.198]
Constant	0.67***	0.53*	0.80***	0.66***	0.76***	0.60***	0.71***	0.72***	0.71***	0.60***	0.67***	0.80***	0.82***
	[0.206]	[0.290]	[0.133]	[0.148]	[0.093]	[0.128]	[0.096]	[0.103]	[0.143]	[0.094]	[0.115]	[0.075]	[0.187]
Observations	31	22	33	52	121	56	109	112	43	125	83	89	31
R-squared	0.043	0.297	0.342	0.082	0.022	0.171	0.045	0.057	0.061	0.127	0.070	0.134	0.103
Robust errors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1