ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Pedagogical Philosophies in Primary School: Do they increase students' cognitive knowledge?

ABSTRACT

I investigate the effect of receiving education with a specific pedagogical philosophy on (1) the final test score in primary school and (2) the secondary school advice relative to the school advice based on the final test score. I use data from a Dutch' (semi-)governmental program that measures student's test scores, information about the school a student attends, and student household information. The effects are estimated via a 'value added' model. I estimate causal effects under the identifying assumption that school selection is exogeneous, which might not hold. My estimates show that (1) receiving education from schools with specific pedagogical philosophies is not associated with higher final test scores, and (2) Montessori and Jenaplan education are associated with handing out teacher-based school advices that are higher than school advices based on the final test score. (2) is economically and statistically significant, but prone to omitted variable bias. I propose several mechanisms that may explain both findings.

Name student: Bjorn Roozenbeek

Student ID number: 532017

Supervisor: Dinand Webbink, Professor of Policy Evaluation

Second assessor: Ivo Arnold, Professor of Economic Education

Date final version: August 1, 2022

Table of Contents

I. Introduction

In recent years the number of primary schools with a specific pedagogical philosophy strongly increased (The Dutch Education Inspectorate, 2018). These schools strive to educate children on a broader scope than cognitive abilities. Given the big impact schools have on a child's development and later life outcomes (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010), this raises the question whether these schools improve the learning outcomes of students.

My study assesses the cognitive benefit of three of these pedagogical philosophies, Montessori, Dalton and Jenaplan education. This is done by investigating the effect of attending those schools on pupils' final test scores. Moreover, I investigate to what extent schools with a specific pedagogical philosophy hand out teacher-based secondary school advices that are higher relative to the secondary school advice based on the final test score. This second investigation is interesting, since the final test score is based on cognitive tests, while schools with specific pedagogical philosophies put a relatively high weight on noncognitive development. My findings will help policymakers decide whether the rise of schools with a specific pedagogical philosophy are desirable from a cognitive perspective.

Much research has been conducted towards the effect of certain school types on cognitive outcomes. Altonjì, Elder and Taber (2005) find that catholic schools affect student's grades very little. The evidence is mixed when it comes to the effect of elite schools on cognitive outcomes. Oosterbeek, Ruijs and De Wolf (2020), Abdulkadiroglu, Angrist, and Pathak (2014), Barrow, Sartain and De la Torre (2020) and Clark (2010) find little to no effect of attending elite education on cognitive outcomes. However, Deming (2014), Dobbie and Fryer (2011), and Duflo, Dupas and Kremer (2011), do find that elite schools lead to higher cognitive outcomes. Other studies by Jackson (2018) and Heckman et al. (2010) assess that there are also noncognitive benefits to education. My study will add to this literature by investigating to what extent education with a specific pedagogical philosophy influences the final test score in the Netherlands. Additionally, my study adds to the literature by investigating to what extent schools that teach based on a pedagogical philosophy hand out school advices that differ from the school advice based on the final test score.

I will do this by using a multivariate regression where I regress final test score, a standardized test all children in the Netherlands are required to take in the last year of primary education, on receiving education with a specific pedagogical philosophy. This is defined by attending either Montessori, Dalton or Jenaplan education. I control for standardized tests taken in group 5, thereby transforming my model to a 'value added' model, as is common in teacher effect research (e.g. Deming 2014; McCaffrey et al. 2003). I also control for a wide set of covariates including regional background, family background, and school religion effects. I address endogeneity concerns by performing Oster's test (2019) on the coefficients. I leverage three cohorts from the COOL5-18 program. A (semi-)governmental program consisting of triennial data, from 2007 until 2016. The program measures student's test scores, information about the school a student attends, and student household information.

My main findings are that (1) education with a specific pedagogical philosophy is not associated with higher cognitive scores, and (2) Jenaplan and Montessori education are associated with advising their students to a higher secondary school track relative to the school advice based on the final test score. (2) is economically and statistically significant, but prone to omitted variable bias.

My study is structured as follows: the context and related literature are reviewed in Section II and III respectively, data and empirical strategies are explained in Sections IV and V, results and potential mechanisms are discussed in Section VI and VII, and the discussion and conclusion are presented in Sections VIII and IX respectively. Section X consists of the references and section XI contains the appendix.

II. Context

Dutch primary education system and transition to secondary education

Dutch primary education consists of approximately eight years. Attendance is mandatory from the age of five to twelve years, but many students already start attending primary school from the age of four. Children start in group 1 and finish in group 8. During these years, they receive education from one or more teachers in a fixed weekly schedule. Group 1 and group 2 can be seen as the Dutch version of kindergarten. After these 8 years, the pupils move on to secondary education. Depending on the school advice pupils receive in group 8 of primary education, pupils either enrol in one of the pre-vocational secondary education ('vmbo') tracks, the higher general secondary education ('havo') track or the pre-university education ('vwo') track. Depending on the secondary school, students may also begin secondary school in a class that mixes two types of education. The school advice is given before the first of March to students in group 8 and is based on how the school perceives what the pupil is good at, how well the pupil learns, and regular test results (Ministerie van Algemene Zaken, 2022).

The final test

Since 2015, the final test is taken somewhere between 15 April and 15 May, thus taken after the school has handed out the school advice. Before 2015, the final test was taken before the school handed out the school advice. This means, that before 2015 a realistic scenario existed in which teachers were biased by the final test scores. All pupils in the last year of their primary education in the Netherlands are required to conduct the final test. Pupils in the year before the final year are not allowed to make the final test for training purposes. The final test is described as a supplement to the school advice and acts as an objective instrument to evaluate what type of secondary education fits the student. If the final test score indicates a higher advice than the school advice, the schools are obligated to revaluate their earlier given school advice. However, if the final test indicates a lower advice, the schools are not allowed to lower their earlier given school advice. Until 2015, the final test was called the CITO-test. Moreover, the Education Inspectorate uses the final test scores to assess schools (Ministerie van Algemene Zaken, 2021) (Ministerie van Algemene Zaken, 2022).

The final test consists at least of math and language. Facultative additional components of the final test are world orientation, geography, history, and nature.

III. Literature analysis

Before elaborating on schools with a specific pedagogical philosophy, I will briefly go over earlier economic literature that investigated the effect of specific school types on cognitive and noncognitive outcomes.

Economic literature on school types and (non)cognitive outcomes

Substantial amount of economic research has been conducted on the effects of certain schools on (non)cognitive outcomes.

By using a multivariate regression, Altonjì, Elder and Taber (2005) show for students in the United States that catholic schools affect student's grades very little. They, do, however, increase the probability of graduating high school with 8 percentage points and attending colleges with 15 percentage points.

Oosterbeek, Ruijs and de Wolf (2020) investigated a Boston mechanism assignment procedure for popular secondary education schools in Amsterdam. By examining differential effects, the authors found two interesting findings. (1) elite schools lower the probability to graduate in the nominal timespan by 22 percentage points for low achieving students, and (2) for the top achieving students it increases the possibility of graduation with 17 percentage points. The author notes, that these findings imply that other studies which focus on regression discontinuity designs should be taken precautionary, since these studies only assess a specific group of students near the threshold (the local average treatment effect).

In line with Oosterbeek et al. (2020), Abdulkadiroğlu, Angrist and Pathak (2014) conclude that elite schools have little causal effect on test scores. They came to this conclusion by exploiting a fuzzy RDD on over-subscribed exam schools in Boston and New York, that selected students based on SAT-scores.

Barrow, Sartain and De la Torre (2020) exploit a RDD on different social economic backgrounds in Chicago. They do not find that elite schools increase test scores over multiple years. However, they do notice that students attending elite schools are more positive about their high school experience relative to other students.

These findings also seem to hold for the United Kingdom. Clark (2010) uses a fuzzy RDD and finds that four years of elite school attendance generates little effect on test scores. However, just as in Oosterbeek Ruijs and de Wolf (2020), Clark notices that attending elite schools improves course-taking and university enrolment.

Contrary to the studies discussed so far, Deming (2014) does assess that elite schools positively influence student's test scores. Deming looked at student's final test scores and controlled for tests the students took before entering an elite school. He exploits data on public school choice lottery in Charlotte-Mecklenburg.

Deming is not alone in this finding, Dobbie and Fryer (2011) leverage a lottery system for elite schools and conclude that sending students with little financial means significantly increases their academic achievement. Specifically, they found that students that went to an elite school scored 0.687 standard deviations higher in math and 0.141 standard deviation reading relative to students attending non-elite schools.

Duflo, Dupas and Kremer (2011) found similar results in Kenia. They compared schools where students were randomly assigned to first grade to schools where students were assigned to first grade based on their entry tests. They find that the latter group performed significantly better than the first group. Specifically, tracking schools scored 0.138 standard deviations higher relative to students in non-tracking schools overall.

Jackson (2010) uses several designs on administrative data about students in Trinidad and Tobago and concludes that attending elite schools significantly effects student's test scores. The effects are twice as big for girls as for boys. Specifically, males' and females' graduation chances rose by approximately 4.8 and 10 percent respectively when their peers have incoming test scores half a standard deviation higher.

Concludingly, the evidence seems to be mixed when it comes to differences between school types. This could be explained by the remark made by Oosterbeek et al. (2020), namely that (fuzzy) RDD designs lead to Local Average Treatment Effects (LATE), thereby overlooking the effect of a certain school type on noncompliers. Studies that leverage (fuzzy) RDD's can thus yield different results than studies that leverage a randomized control trial or multivariate regression design. The contextual setting of the discussed studies can also play a role. No educational system in the world is the same, so what in one country is seen as an 'elite school' might not be comparable to what other countries define an 'elite school'.

Non-cognitive development

Although cognitive development is important, it is merely one of the determining factors for a bright future. Hence, the development of non-cognitive skills might also be a determinant of parents in their deliberation to send their child to a specific school. Specifically, the development of five personality traits defined by Fiske (1949): openness, conscientiousness, extraversion, agreeableness, and neuroticism. It is therefore interesting to see what previous literature found about the relation between these non-cognitive aspects and later life outcomes. Jackson (2018) investigated non-test score behaviours such as absences, suspensions, course grades, and grade repetition in ninth grade in North Carolina. He finds that test scores and non-cognitive scores are weakly correlated, and that improved non-cognitive skills better predicts long run outcomes such as high school completion. More precisely, Jackson found that a one standard deviation increase in behaviour increases following year GPA, SAT taking, likelihood of going to a 4-year college after high school and graduating high school.

This is corroborated by Heckman, Stixrud, and Urzua (2006). By using longitudinal data on youth in the United States they find that schooling level can have a bigger impact on noncognitive outcomes than cognitive outcomes. A similar conclusion can be drawn from Heckman, Moon, Pinto, Savelyev, and Yavitz, (2010), where they investigated the outcomes from the popular Perry Preschool Programme.

Glaeser, Ponzetto and Shleifer (2007) add another interesting motivation for the development of noncognitive skills. They see that education raises the benefits of civic engagement, and it therefore raises participation for democracy over authoritarian regimes.

Deming (2011) assess that high-risk youth that went to an elite school, commit 50% less crime in later life in three measures that index crimes by severity.

These studies appear to inform us that non-cognitive development yields positive later life outcomes. Corroborating the idea that non-cognitive development plays an important factor in parents' decision to send their child to a specific school.

Specific pedagogical philosophies and cognitive outcomes

In the Netherlands there are five types of schools with a specific pedagogical philosophy. Montessori, Jenaplan, Dalton, Freinet and Steiner education. Montessori education aims to give children a sense of independence, self-esteem and confidence in their own personality (De Nederlandse Montessori Vereniging, 2019). Jenaplanconcept is a concept in which relations are central. The relation of the child with itself, the relation of the child with the other, and the relation of the child with the world (Nederlandse Jenaplan Vereniging). The five core values of Dalton education are responsibility, collaboration, effectivity, independence and reflection (Nederlandse Dalton Vereniging, 2021). In Freinet education, the experiences and learning questions of the children are the starting point for education. The organisation of class life is done by the children together with the teacher, thereby promoting citizenship (Vereniging voor Freinetpedagogie). Steiner pedagogy is a form of education centred around the human and developmental vision of anthroposophy, a philosophy that serves as a source of inspiration but is not taught itself. Education serves the formation of personality, including social formation. The goal is for the student to be able to develop freely and in a balanced way (Federatie Steinerscholen, 2022). A denominator for all these concepts is that they strive to educate children on a broader scope than only cognitive abilities.

Evidence about these specific pedagogical philosophies

The following studies assess the influences of these five school types with a specific pedagogical philosophy on (non-)cognitive outcomes.

Berends and Wolthuis (2014) explored to what extent Dalton education influences both cognitive as well as non-cognitive outcomes of children following this education. They found no difference when it comes to cognitive outcomes relative to traditional schools. However, they did find that students attending Dalton school in 5th grade scored higher on the Citizenship Competence Questionnaire. They conducted their study via multilevel analyses on language, math, wellbeing, self-efficacy, task-motivation, and citizenship competencies by using large scale triennial cohort data in the Netherlands (COOL).

De Bilde, van Damme, Lamote and De Fraine (2013) probed whether enjoying education from a school with a specific pedagogical philosophy increases children’s school engagement. By assessing school enjoyment and independent participation among 6.000 students that participated in the ‘School careers in primary education’ (SiBO) project, they concluded that children attending schools with a specific pedagogical philosophy act less independent than traditional school children.

Research on schools with a specific pedagogical philosophy has also been conducted outside of the Netherlands. Seegers, van Putten and de Brabander (2002) scrutinized to what extent goal orientation affects mathematics outcomes in British schools. They find via multivariate analyses that performance-oriented learning goals emphasizes the negative impact of failure experiences, whereas task-oriented learning goals have a strengthening effect on how success experiences influence students’ attitude.

Kliebard (2004) examined different Montessori programs and their effect on cognitive outcomes of a child. He concluded that education with a high degree of Montessori aspects is associated with better cognitive outcomes than education with little or non-Montessori influences.

This appears to be in line with Lilliard et al. (2017). These authors exploited a randomized admission process to Montessori education and concluded that Montessori children perform

better on academic achievement, social understanding and mastery orientation. Moreover, they found that Montessori education seems to lead to convergence in outcomes among subgroups.
These studies seem to reflect that educational outcomes are different per type of pedagogical philosophy, something that is corroborated by van der Wal and Waslander (2007). They investigated whether pedagogical philosophies are complimentary to cognitive outcomes. They found via multilevel analyses on 26 schools in the Netherlands and a literature review that it is ambiguous if a trade-off or complimentary effect occurs. This depends on specific school conditions.

Summary

Concludingly, a lot of economic research has centred on differences between religious and non-religious schools, and elite vs. non-elite schools. Some sociological and pedagogical research, but little to no economic research so far focused on differences between schools with a specific pedagogical philosophy.

This has motivated me to investigate two questions. Whether attending a school with a specific pedagogical philosophy result in higher cognitive outcomes, measured by standardized CITO-scores in the final year of primary education. And whether attending a school with a specific pedagogical philosophy leads to a school advice that is higher relative to the school advice based on the CITO-score, given that these schools focus more on non-cognitive development than regular schools.

Investigating this issue at primary school level is also socially relevant, since policymakers can leverage my results to see to what extent these pedagogical philosophies have added value on a cognitive level. This is especially important, given that early-life influences play an important role in a child's later life development. Heckman underlines the importance of early interventions targeted toward children (2006). In his literature study he concluded that the earlier an intervention takes place, the higher the returns.

IV. Data

To investigate the effect of attending a school with a specific pedagogical philosophy on cognitive results and school advice, I use cohort studies that collect data about the development of children during their school careers. The specific cohort studies I use, are deducted from the COOL5-18 program (Cohort Study of Education Careers among Students 5 to 18 Years of Age). This specific cohort study consists of triennial data, from 2007 until 2016. The COOL5-18 program is conducted by several (semi-)governmental Dutch institutions. Namely, the Kohnstamm Institute of the University of Amsterdam and the Institute for Applied Sociology (ITS) of the Radboud University Nijmegen for the data at primary school level, and Cito and Groningen Institute for Educational Research (GION) for data at the secondary education level. The study measures student's test scores, information about the school a student attends, student household information (measured by questionnaires filled in by the student's parent(s)), and information on the student's learning and classroom behaviour (measured by questionnaires filled in by the student's teacher). I use the cohorts of 2007/2008, 2010/2011, and 2013/2014, since these are all the cohorts that have student information at the primary school level. The descriptive statistics of this data can be observed in table 1. Sadly, there were no observations available for students that attend Freinet or Waldorf education. Therefore, I shall limit the scope of my study to Jenaplan, Dalton and Montessori education. A student attends 'regular' education when the teacher answered that their school is not influenced by any pedagogical philosophy. So, schools that are partly influenced by a certain philosophy are removed from the control group as well. A student attends one of the three pedagogical philosophies when the teacher answered that their school is completely influenced by the respective specific pedagogical philosophy.

Table 1 Descriptive statistics per pedagogical philosophy for individuals that participate in group 5 and group 8.

Variables	Regular	Montessori	Jenaplan	Dalton	Total
% women	50.39	40.91	52.23	54.98	50.41
Province %					
Groningen	2.60				2.37
Friesland	2.12			7.69	2.17
Drenthe	0.82				0.75
Overijssel	9.96			16.67	9.60
Flevoland	0.91				0.83
Gelderland	10.48		54.29	75.64	12.64
Utrecht	3.68				3.36
Noord-Holland	14.12	33.64			14.34
Zuid-Holland	16.28	66.36			17.73
Zeeland	4.46				4.07
Noord-Brabant	25.77				23.50
Limburg	8.79		45.71		8.65
Urbanity (scale 1-5)	2.84	1.66	2.91	4.56	2.84
	(1.22)	(0.47)	(1.01)	(0.50)	(1.24)
Migrant background %	22.80	11.01	14.71		21.49
Social Economic Status %					
1	12.05	4.59	6.45		11.27
2	15.37		3.23	10.26	14.38
3	7.13	1.83	6.45		6.66
4	35.83	8.26	25.81	51.28	34.98
5	3.68	4.59	3.23		3.59
6	25.95	80.73	54.84	38.46	29.12
Denomination %					
Public	17.45	100	45.71	51.28	22.47
Roman Catholic	40.97			23.08	38.07
Protestant	31.44			25.64	29.46
Other	10.13		54.29		9.99
Highest education father %					
Primary education	10.66	0.94	6.67		9.87
lower vocational education	31.72	8.49	3.33	18.18	29.90
vocational education	35.70	7.55	43.33	51.95	35.07
Higher education	21.92	83.02	46.67	29.87	25.19
Highest education mother %					
Primary education	12.94	2.75	6.45	1.28	12.05
lower vocational education	24.25	3.67	9.68	23.08	23.13
vocational education	42.03	21.10	45.16	51.28	41.45
Higher education	20.77	72.48	38.71	24.36	23.37
Single parent %	14.57	8.11			13.75
Final test score (CITO)	534.4	539.2	536.8	535.4	534.7
	(10.1)	(9.0)	(9.0)	(8.2)	(10.0)
Math score group 5 (standardized)	0.134	0.333	0.200	0.142	0.144
	(0.968)	(0.889)	(0.717)	(1.047)	(0.965)
Reading score group 5 (standardized)	0.085	0.352	0.248	0.024	0.097
	(0.963)	(0.994)	(0.842)	(0.882)	(0.961)
Secondary school advice (%)					
vwo	16.19	45.79	44.12	14.10	17.79

havo/vwo	7.16	11.21		3.85	7.13
havo	19.62	24.30	23.53	21.79	19.94
vmbo/havo	6.36	10.28		3.85	6.36
vmbo	50.67	8.41	32.35	56.41	48.76
Individuals	2,309	110	35	78	2,532

Sample consists of students categorized per pedagogical philosophy. Means are denoted for every variable. The standard deviations are denoted in parentheses. There are a total of 5,004 observations on 2,502 individuals. Observations are from triennial panel data over the years 2008 – 2014. Every individual is measured twice, once in group 5 and once in group 8. Standardized test scores in group 5 are measured in terms of standard deviations. For urbanity, a lower score indicates more urbanity. Social economic status is determined by the migration background of a student and parent's education level. The Social economic status scores represent the following: (1) lower vocational education and migrant background, (2) lower vocational education and native background, (3) vocational education and migrant background, (4) vocational education and native background, (5) higher education and migrant background, and (6) higher education and native background.

The descriptive statistics differ substantially per pedagogical philosophy.

Dalton and Jenaplan education seem to appear mainly in the less urban provinces, while Montessori appears more in the more urban provinces. This is confirmed by the average urbanity of each education type. Especially Montessori education appears very often in urban areas relative to the other educational concepts.

All pedagogical philosophies have a low number of students with a migrant background relative to regular education. A student has a migrant background when at least one of his parents has a migrant background.

The differences between the schools with a pedagogical philosophy and regular schools are quite large when it comes to parental education level. Especially students who attend Montessori and Jenaplan education have parents with a high educational background. The difference is less for students attending Dalton education.

The number of single parent households is lower for all students attending schools with a specific pedagogical philosophy relative to regular education.

The final test scores seem to be, on average, higher for students attending different educational concepts, especially for Montessori students. For students attending the other educational concepts, the final test scores seem to be the same for those students attending regular education.

The number of students receiving a 'vwo' school advice for secondary education is, on average, higher for students attending Montessori and Jenaplan education. They also differ from regular education when it comes to the percentage of students receiving a 'vmbo' advice.

Concerns with the data

Due to my desired design, I need to observe individuals once in group 5 and once in group 8. Individuals whom I do not observe twice, are removed from the data. This has led to a total of 2,532 unique individuals. This sample contains a relatively small sample of individuals for the specific pedagogical philosophies, only 223 individuals. This makes my treatment group prone to selection bias.

To investigate to what extent my data is still representative of the Netherlands, I included descriptive statistics of all 35,066 individuals in group 8 that participated in the COOL study in table A2 in the appendix. Differences are observable, especially when it comes to descriptive statistics of the specific pedagogical philosophies. This gives me reason to suspect that my treatment group is no longer sufficiently representative of the total population.

These differences intensify the importance of controlling for a wide array of observable factors, to try to tackle false correlations.

V. Empirical Strategies

The effect of education with a specific pedagogical philosophy on cognitive outcome

It is my aim to investigate if there is a causal effect of education with a specific pedagogical philosophy on cognitive outcomes. I do this by conducting a multivariate analysis research design. The COOL data also consists of data on tests scores in group 5. This allows me to transform my model to a 'value added' model by standardizing these test scores in group 5 and controlling for these standardized scores. Furthermore, this allows me to control for intelligence level of student's before they receive the education in-between group 5 and group 8. In my design, I need to be aware of several factors that could influence both my outcome variable (cognitive outcome) as well as my variable of interest (attending a school with a specific pedagogical philosophy).

This yields equation (1):

$$(1)\ Y_{ij} = \alpha_0 + \alpha_1 T_{ij} + \alpha_2 M_{ij} + \alpha_3 R_{ij} + X_{ij} + \varepsilon$$

Where Y_{ij} is the outcome variable that measures the final test score for individual i that attends a school with a specific pedagogical philosophy j. j represents education types Montessori, Dalton, Jenaplan, or all three of them combined. α_0 is a constant. α_1 is the coefficient of interest and denotes the effect of education concept j on final test score. T_{ij} is a categorical variable denoting what educational concept j individual i attended. α_2 and α_3 respectively denote the coefficient of the standardized math M_{ij} and reading R_{ij} test in group 5. X_{ij} is a set of control variables, controlling for variables that are correlated with both my variable of interest as well as my dependent variable. This set includes household characteristics (Household composition, Social economic status, Education level father, and Education level mother), regional characteristics (Level of urbanity and Province), religious conviction of the school, and time variant effects. ε is the error term and is clustered at the school level.

To compare my results with results found by previous literature, my tables also show the coefficients in terms of standard deviations.

My identifying assumption is that selection on schools is exogeneous. To find a causal effect of my variable of interest on the dependant variable, four important criteria must therefore be met. Conditional mean independence assumption (CIA), identically distributed and independent means, large outliers unlikely and no perfect multicollinearity. The conditional mean independence assumption will most likely not be met, since I can only control for observable factors that might influence my variable of interest and my dependant variable. Hence, I cannot control for unobservable factors that might influence both my variable of interest as well as my outcome variable. An example of such an endogeneity concern is that I cannot control for teacher quality of certain schools, while teacher quality could influence both the school choice as well as the final test score. The identically distributed and independent random variables assumption is met. Since it is highly unlikely that a certain outcome of one sample is influenced by the outcome of another sample. Perfect collinearity will be of little concern, since Stata automatically omits perfect collinear variables. Concludingly, I will be able to find a causal effect of attending education with a specific pedagogical philosophy on final test scores, under the assumption that the conditional mean assumption holds.

The effect of education with a specific pedagogical philosophy on secondary school advice

My aim in this second question is to find the causal effect of receiving education from a school with a specific pedagogical philosophy on secondary school advice. I will run the following regression where I again control for the standardized test scores in group 5, thereby transforming my model to a 'value added' model. Moreover, this allows me to balance out intelligence level at group 5:

$$(2)\ Y_{ij} = \alpha_0 + \alpha_1 T_{ij} + \alpha_2 M_{ij} + \alpha_3 R_{ij} + X_{ij} + \varepsilon$$

Where Y_{ij} is the outcome variable that measures the size of the difference between school advice and school advice based on the final test score. It can thus also be written as $Y_{ij} = A_{ij} - F_{ij}$, with A_{ij} and F_{ij} representing the respective school advice and school advice based on the final test score guideline (see table A1 in the appendix) that individual i attending educational concept j received. A_{ij} can take on 8 values: 50 for 'vwo' (pre-university education), 45 for 'havo/vwo', 40 for 'havo' (higher general secondary education), 35 for 'havo/vmbo-tl', 30 for 'vmbo-tl' (secondary vocational education theoretical pathway), 20 for 'vmbo kader' (secondary education, intermediate pathway), 15 for 'vmbo kader/basis', and 10 for 'vmbo basis' (secondary vocational education, basic pathway). A_{ij} can take on the values 'x5', since in practice teachers can also hand out school advices that are in between two secondary school tracks because Dutch middle schools often offer mixed classes. F_{ij} can take on only singular advices, and no mixed advices, i.e., F_{ij} can take on only 'x0' values. α_1 is the coefficient of interest and denotes the effect pedagogical philosophy j has on $A_{ij} - F_{ij}$. T_{ij} is a categorical variable denoting what educational concept j individual i attended. α_2 and α_3 respectively denote the coefficient of the standardized math M_{ij} and reading R_{ij} test in group 5. X_{ij} is a set of control variables, controlling for factors that influence both my variable of interest as well as my dependent variable. This set includes the same control variables as in the previous regression. ε is the error term and is clustered at the school level.

The way A_{ij} and F_{ij} are formulated can make the coefficient of interest hard to interpret. Therefore, my tables also show the coefficients in terms of standard deviations.

Again, my identifying assumption is that selection on schools is exogeneous. There are the four assumptions that need to hold to allow me to find a causal relationship between attending education with a specific pedagogical philosophy secondary school advice. The conditional mean independence assumption cannot be guaranteed since unobserved factors could affect both my variable of interest as well as my dependent variable. The identically distributed and independent random variables assumptions will hold, since it is unlikely that the outcome of my sample will affect the outcome of other samples. There will be no outliers for school advice, since receiving a certain school advice is a categorical variable. Perfect multicollinearity is not an issue since Stata automatically omits perfect collinear variables.

Selection on observable and unobservable variables

Unfortunately, my observable data is not suitable for a (quasi-)experimental research designs. Thus, leveraging the multivariate regression design is the most reasonable thing to do. Nevertheless, this leads to selection issues resulting in a violation of the CIA; I can only control for observable characteristics in the data and not for unobservable factors that are 'hidden' in the error term and influence both the variable of interest as well as the dependent variable. To address this concern, Altonji, Elder and Taber (2005) formulated a method which entails that observable factors contain information about the bias that originates from the selection on unobservable characteristics, thereby gaining information about the sensitivity of the coefficient of interest to omitted variable bias.

Briefly described, Altonji, Elder and Taber assume that selection on the unobservables is the same as selection on the observables under the condition that the part of the dependent variable that is related to the observables and the part related to the unobservables have the same relationship with the latent independent variable of interest. Altonji substantiates this condition by referring to three types of assumptions made in Altonji et al. (2002). (1) The full set of observable factors are chosen at random from the full set of factors (observable and unobservable) that affect the dependent variable. (2) The set of observables and the full set of factors are large, so that no specific factor dominates the distribution of the variable of interest and/or the dependent variable. (3) The set of observable factors is orthogonal to the set of unobservable factors. i.e., adding unobservable control variables to the regression will not affect the coefficients of the observable variables.

Altonji et al. (2005) argues that this condition is more realistic than the conditional independence assumption, due to the collection design of many large-scale datasets. Large datasets are designed in a way to be useful to multiple disciplines, and therefore contain a wide variety of observable characteristics. The assumptions for the condition to hold thus imply that it is better to think of the set of available observable characteristics as a random subset of the full set of factors that influence the dependent variable.

Oster's test

Oster (2019) formalizes and expands on the theory introduced by Altonji et al. (2005). The main expansion of Oster is that Omitted Variable Bias (OVB) is proportional to coefficient movement when adding observable control variables, but only if such movements are scaled by the change in R-squared when controls are included. Oster creates the following model:

> Suppose we are interested in finding the clear effect of X on y, where y is the outcome variable, α_0 is the constant, α_1 the coefficient of interest for the model with both observable - W_1 - and unobservable - W_2 - control variables, X the variable of interest and ε the error term:
>
> $$y = \alpha_0 + \alpha_1 X + W_1 + W_2 + \varepsilon$$
>
> But, since I can only access observational data, Oster advices the following steps:
>
> (1) Run a regression without control variables, where $\dot{\alpha}_1$ is the coefficient of interest for the model without control variables:
>
> $$y = \alpha_0 + \dot{\alpha}_1 X + \varepsilon$$
>
> (2) Then, run a regression with only observable control variables where $\tilde{\alpha}_1$is the coefficient of interest for the model with observable control variables:
>
> $$y = \alpha_0 + \tilde{\alpha}_1 X + W_1 + \varepsilon$$

Oster introduces the proportional selection relationship, in order to explain the mechanism of the coefficient of proportionality, δ:

$\delta \frac{\sigma_1 x}{{\sigma_1}^2} = \frac{\sigma_2 x}{\sigma_2^2}$, where $\sigma_i x = cov(W_i, X)$, and $\sigma_i^2 = var(W_i)$, with $i \in \{1, 2\}$

$i = 1$ denotes the values for only the observed set of control variables W_1, and $i = 2$ denotes the values for the full set of observed and unobserved control variables W_2. X denotes the treatment variable. Thus, if $\delta > 1$, we may infer that selection on observables has less strength than selection on unobservables. Vice versa for when $\delta < 1$. Orthogonality between W_1 and W_2 is assumed here once more, implying that adding unobservable control variables to the regression will not affect the coefficients of W_1.

(3) Oster then explains that one can leverage the results from these relationships via two methods: Calculate (a) bounding set for the coefficient of interest, and (b) calculate the 'coefficient of proportionality', δ. These take the explanatory power of the regressions in steps (1) and (2) into account, $\dot{R}$ and $\tilde{R}$ respectively. The methods are available in Stata and called *psacalc*. Through this tool, I will try to approximate to what extent my coefficients of interest are robust to omitted variable bias.

Calculating the bounding set for the coefficient of interest

This will lead to a lower and upper bound estimate of the coefficient of interest $[\tilde{\alpha}, \alpha^*]$ under the following assumptions: Coefficient of proportionality, $\delta = 1$, and direction of $cov(W_1, X)$ is unchanged by including unobservable control variables. If these lower and upper bounds differ little, then one may assume that the estimated coefficient of interest is robust for OVB.

Calculating the 'coefficient of proportionality'

This will generate the δ, under the assumption that if the unobservables are added to the model, the coefficient of interest will be zero. If this leads to an unrealistic δ, then one may assume that the estimated coefficient of interest is robust for OVB. This method is especially useful when one questions the accuracy of $\delta = 1$.

VI. Results

The effect of receiving education with a specific pedagogical philosophy on final test score

Table 2 Regression results of final test scores on attending education with a specific pedagogical philosophy

Final test scores	(1)	(2)	(3)	(4)	(5)	(6)
Pedagogical philosophy	3.1**	1.9*	0.6	0.2	0.8	0.5
	(1.3)	(1.1)	(0.9)	(0.8)	(0.9)	(0.9)
COI in st. dev.	0.306	0.191	0.063	0.022	0.085	0.053
Test scores group 5	No	Yes	Yes	Yes	Yes	Yes
Household background characteristics	No	No	Yes	Yes	Yes	Yes
Regional characteristics	No	No	No	Yes	Yes	Yes
School denomination	No	No	No	No	Yes	Yes
Time variant effects	No	No	No	No	No	Yes
Constant	534.4	533.6	533.4	532.8	532.2	532.7
Individuals	2,532	2,467	2,300	2,300	2,300	2,300
Montessori	110	103	100	100	100	100
Jenaplan	35	35	30	30	30	30
Dalton	78	77	76	76	76	76

*Multivariate regression on final test score, which is measured on a scale of 500-550, 550 and 500 being the respective highest and lowest score possible. The row 'COI in st. dev.' shows the coefficient of interest in terms of standard deviations. The columns differ in the amount of control variables included. 'Yes' indicates that the respective observable factors have been controlled for in the corresponding regression, 'No' indicates that this has not been the case. Observations for the years 2008 – 2013 are included. The columns show estimated coefficients and standard errors are denoted in parentheses. Standard errors are robust and clustered at the school level. The last three rows show the number of individuals per specific pedagogical philosophy that are included in the regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

The naive with-and-without comparison in column (1) yields a coefficient of interest of 3.1 points on the final test score, which is significant at the 5% significance level.

Controlling for standardized tests in group 5 reduces the coefficient of interest to a significant value of 1.9 points on the final test score at the 10% significance level.

Adding control variables to control for household characteristics sizably lowers the association found in column (2) to 0.6 points on the final test scores. This corroborates the suggestion made by the Dutch education Inspectorate (2018), that wealthier and highly educated households send their children relatively more often to schools with a specific pedagogical philosophy compared to less wealthy and less educated parents. Moreover, this suggests that having a wealthy and/or highly educated family background positively affects one's cognitive development.

Controlling for regional effects sizably lowers the coefficient of interest.

Motivated by Altonji et al. (2005), I added controls for the religious conviction of schools, to take away bias caused by religious schools. Interestingly, adding these control variables sizably raises the coefficient of interest to an insignificant value of 0.8 points on the final test score. Implying that religious conviction of schools is associated with cognitive outcomes.

In the last column, a control variable is included to control for time variant effects. This lowers the already insignificant coefficient of interest.

Results per specific pedagogical philosophy

Table A3 in the appendix shows the results for regressions on the full set of observable control variables for all three types of education with a specific pedagogical philosophy, Montessori, Jenaplan and Dalton. For all these pedagogical philosophies, the coefficient of interest is insignificant. These results show that the results in table 1 are predominantly driven by Montessori and Dalton education, while Jenaplan education is associated with lower final test scores.

These results seem to indicate that attending schools that hold on to specific pedagogical philosophies does not lead to higher final test scores relative to regular schools.

Oster's test

Table 3 Oster test for stability of coefficient of interest for regression (6) in Table 2

Pedagogical philosophy	$\dot{\alpha}$	$\tilde{\alpha}$	α^* $(\delta = 1)$	$\alpha^* = 0$
Coefficient	2.7	0.5	-6.4	0
R-squared	0.006	0.525	1	n.a.
δ	n.a	n.a	1	0.15079

$\dot{\alpha}$, $\tilde{\alpha}$, and α^* represent respectively the uncontrolled, controlled, and bias-adjusted coefficient of interest. The first and second columns show the input for the Oster test. The third column shows results of the first method of the Oster test, the upper bound estimate of the coefficient of interest. The coefficient of proportionality here is assumed to be 1, thus indicating that the effect of observables is as big as the effect of unobservables. The fourth column shows the results of the second method of the Oster test.

The bounding set for the coefficient of interest

Table 3 shows the results of the Oster test to investigate the stability of the coefficient of interest. If the coefficient of proportionality, δ, is equal to one, this test yields an upper bound of the coefficient of interest of -6.4 points on the final test score. In other words, given that Oster's assumptions hold, the latent coefficient of interest is within the interval [0.5 ; -6.4].

The coefficient of proportionality

The last column shows the second Oster test, what the coefficient of proportionality would have to be given that adding the unobservables would lead to a bias-adjusted effect of 0. It takes on the value 0.15079 here, indicating that even the slightest addition of unobservables will have a powerful effect on the coefficient of interest.

Assumptions Oster's test

These tests appear to show evidence of an unrobust coefficient of interest. Still, it is important to critically assess the underlying assumptions of these two Oster tests:

The assumption that $\delta = 1$ is a strict assumption that is substantiated by the assumptions that (1) no specific factor dominates the variable of interest or dependent variable, and (2) that the full set of observable factors are chosen at random.

Given the many factors that influence a child's (cognitive) development, (1) can be assumed to hold. (2) is difficult to substantiate, the COOL study specifically serves the purpose to describe and explain the development of children during their school careers. Thus, the set of observable factors that I have at my disposal are not random but rather purposely documented to investigate what drives educational attainment. Therefore, a lower value of δ would perhaps be more realistic.

But, as stated in the methodology, Oster's second test can be used as another robustness test to overcome my concern regarding the validity of the value of δ in Oster's first test.

Oster's last assumption, orthogonality, will most likely hold. Since my coefficient of interest did not change much anymore after I controlled for household characteristics. This gives me reason to believe that this assumption holds.

The effect of receiving education with a specific pedagogical philosophy on school advice

Table 4 Regression results of difference in school advice on attending education with a specific pedagogical philosophy

Difference in advice	(1)	(2)	(3)	(4)	(5)	(6)
Pedagogical philosophy	2.26**	2.27**	1.96*	1.83	1.42	1.61*
	(1.05)	(1.08)	(1.065)	(1.09)	(1.09)	(0.975)
COI in st. dev.	0.339	0.343	0.295	0.276	0.214	0.243
Test scores group 5	No	Yes	Yes	Yes	Yes	Yes
Household background characteristics	No	No	Yes	Yes	Yes	Yes
Regional characteristics	No	No	No	Yes	Yes	Yes
School denomination	No	No	No	No	Yes	Yes
Time variant effects	No	No	No	No	No	Yes
Constant	0.58	0.54	-0.04	0.60	0.69	0.44
Observations	2,467	2,404	2,242	2,242	2,242	2,242
Montessori	107	101	98	98	98	98
Jenaplan	34	34	29	29	29	29
Dalton	78	77	76	76	76	76

Multivariate regression on difference in school advice. Difference in school advice is measured as the size of the difference between school advice and school advice based on the final test score. Observations for the years 2008 – 2013 are included. The row 'COI in st. dev.' shows the coefficient of interest in terms of standard deviations. The columns differ in the amount of control variables included. 'Yes' indicates that the respective observable factors have been controlled for in the corresponding regression, 'No' indicates that this has not been the case. Observations for the years 2008 – 2013 are included. The columns show estimated coefficients and standard errors are denoted in parentheses. Standard errors are robust and clustered at the school level. The last three rows show the number of individuals per specific pedagogical philosophy that are included in the regression. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$.

The coefficients of interest in table 4 are quite sensitive for selection on observables. Note that the coefficients of interest are also included in terms of standard deviations.

Column (1) shows the naive with and without comparison. This yields a significant result of 2.26 difference points in school advice at the 5% significance level.

In column (2) control variables are added to control for intelligence level in group 5 and transform the model to a value-added model. These control variables affect the outcome variable very little, indicating that intelligence level at group 5 is not correlated with the difference in school advice.

After controlling for household characteristics in column (3), the coefficient of interest lowers to a significant value of 1.96 difference points in school advice at the 10% significance level.
In column (4) control variables are added to the model to control for regional effects. The coefficient of interest lowers to an insignificant value 1.83 difference points in school advice.
Religious conviction is added to the model in column (5) to control for religious conviction. This leads to a reduction in the coefficient of interest, an insignificant value of 1.42 difference points.
In the last column, a control variable is included to control for time variant effects. This leads to a significant coefficient of interest of 1.61 difference points in school advice at the 10% significance level.

Results per specific pedagogical philosophy

Table A4 in the appendix shows that different pedagogical philosophies differ in the difference that arises between school advice and school advice based on the final test score. Including all control variables, students attending Montessori, Jenaplan and Dalton education receive a value of difference in school advice of respective 3.31 (1% significance level), 4.34 (1% significance level) and -1.54 (insignificant) relative to students attending regular schools. Assuming that the conditional independence assumption holds, these results indicate that Montessori and Dalton education lead to higher school advices relative to the school advice based on the final test score.
These results are interesting and economically significant. The result for Jenaplan education indicates a 0.434 higher school advice than regular schools. This is almost completely the difference between receiving a 'vmbo' advice or a 'vmbo/havo' advice.

Oster's test

Table 5 Oster test for stability of coefficient of interest for regression (6) in Table 4

Pedagogical philosophy	$\dot{\alpha}$	$\tilde{\alpha}$	$\alpha^* \ (\delta = 1)$	$\alpha^* = 0$
Coefficient	2.25	1.61	-461.89	0
R-squared	0.010	0.054	1	n.a.
δ	1	1	1	0.04123

$\dot{\alpha}$, $\tilde{\alpha}$, and α^* represent respectively the uncontrolled, controlled, and bias-adjusted coefficient of interest. The first and second columns show the input for the Oster test. The third column shows results of the first method of the Oster test, the upper bound estimate of the coefficient of interest. The coefficient of proportionality here is assumed to be 1, thus indicating that the effect of observables is as big as the effect of unobservables. The fourth column shows the results of the second method of the Oster test.

The bounding set for the coefficient of interest

Table 5 shows the results of the Oster test to investigate the stability of the coefficient of interest. If the coefficient of proportionality, δ, is equal to one, this test yields a lower bound of the coefficient of interest of -461.89. In other words, In other words, given that Oster's assumptions hold, the latent coefficient of interest is within the interval [1.61; -461.89].

The 'coefficient of proportionality'

The last column shows the second Oster test, what the coefficient of proportionality would have to be given that adding the unobservables would lead to a bias-adjusted effect of 0. It takes on the value 0.04123 here, indicating that the only the slightest addition of unobservable factors will already have a huge impact on my coefficient of interest.

These tests show evidence of an unrobust coefficient of interest. This is mostly due to the low values of $\dot{R}$ and $\tilde{R}$, which scale the upper bound coefficient to a very large negative value. The assumptions of Oster's test in Table 5 also hold, for the same reasons that the assumptions of Oster's test in Table 3 hold.

VII. Potential mechanisms

Several interesting findings have occurred. (1) Attending schools with a specific pedagogical philosophy does not lead to higher final test scores. (2) attending Montessori and Jenaplan schools leads a school advice that is higher than the school advice based on the final test score, given that the conditional independence assumption holds.

The findings will first be combined with previous literature, after which several mechanisms are proposed.

Attending schools with a specific pedagogical philosophy does not lead to higher final test scores

Previous literature

The effect of 0.5 points on the final test score can be translated to an increase in the final test score of 0.053 standard deviations when a student receives education with a specific pedagogical philosophy. This is in line with the results found by Oosterbeek et al. (2020), Abdulkadiroglu et al. (2014), Barrow et al. (2020) and Clark (2010). These studies found little to no effect of attending elite education on cognitive outcomes. They are in line with Berends and Wolthuis (2014) as well, who found no added value in Dalton pedagogy.

Other studies did find increases in cognitive outcomes of attending elite schools. Specifically, Deming (2014), Fryer (2011), Duflo et al. (2011) and Altonji et al (2005). My result is lower than the added values found in Fryer – 0.687 and 0.141 standard deviations higher for respective math and reading tests – and in Duflo et al. – 0.138 standard deviations higher. Two important sidenotes should be made here. The cited studies focused on the effect of elite schools. Elite schools generally do not distinguish themselves by pedagogical philosophies, but by having more means and only admitting a selective pool of students. Secondly, my study investigated the effect pedagogical philosophies have at the primary school level, while the other studies investigated the effect of attending elite schools at the secondary school level.

Table 3 indicates that my results for this relation are not robust for omitted variable bias. Moreover, table A3 indicates that no effect can be found by either of the three philosophies.

Potential mechanisms

Why do these specific pedagogical philosophies not influence the cognitive outcomes? I propose two potential mechanisms. The first one being that mostly wealthy, highly educated households send their children to these schools. In these households, children may have some advantage over other children since their cognitive development can be influenced by their parents' educational attainment. Moreover, there may be a better home situation for wealthy children since there is little to no financial distress relative to less wealthy children. This mechanism can explain why I observe such a large decrease in my coefficient of interest after controlling for the standardized test scores in group 5. Indeed, the scores of the standardized tests in group 5 may be largely correlated with household characteristics, since a child's cognitive attainment in group 5 can, to a large extent, already be influenced by household characteristics. The influence of household characteristics that is not explained by the standardized test scores in group 5, is captured after separately controlling for household characteristics. This also led to a sizable reduction in the coefficient of interest.

The second mechanism I propose, is that due to the stringency of the final test score, schools with a specific pedagogical philosophy may have altered their curriculum in such a way that the children are prepared sufficiently for the final test score. As discussed in the context, the final test score is also used as a tool to assess the quality of a primary school by the Educational Inspectorate. Thus, schools with a specific pedagogical philosophy may be incentivised to pay less attention to the noncognitive developmental aspects with which they advertised and more attention to the cognitive development of the child. Table A5 in the Appendix shows the average scores students in group 8 scored on noncognitive traits, as reported by their teachers. This table shows that the pedagogical philosophies score similarly to regular education on almost every noncognitive trait, thereby strengthening this second mechanism. This second mechanism can explain why, after controlling for standardized test scores in group 5 and household characteristics, the final test scores of these schools are roughly the same as schools that are not influenced by any specific pedagogical philosophy.

Jenaplan and Montessori schools over advice their students to higher secondary school tracks relative to the school advice based on the final test score, under the CIA.

Potential mechanisms

After investigating the effect of specific pedagogical philosophies on difference in school advice, I find that that receiving Montessori or Jenaplan education is associated with a respectively 3.31 and 4.34 difference points higher school advice relative to receiving regular education, both at the 1% significance level. This is equal to a respective 0.496 and 0.650 standard deviations higher difference relative to students not attending these specific pedagogical philosophies. Contrarily, attending Dalton education is associated with 1.54 points lower school advice (not significant), which is equal to 0.231 standard deviations lower difference in school advice.

This difference between Jenaplan and Montessori on the one hand and Dalton on the other hand, might be explained by the fact that Montessori and Jenaplan education characterize themselves more by focusing on non-cognitive development relative to Dalton education. Thereby also putting more weight on noncognitive factors in their assessment of the child relative to Dalton education.

Indeed, Jenaplan and Montessori education focus on traits that are less associated with cognitive development. Jenaplan education is defined as a concept in which relations are central. The relation of the child with itself, the relation of the child with the other, and the relation of the child with the world. And Montessori education focusses on the development of independence, self-esteem and confidence of children. Dalton education on the other hand, focusses more on traits that are more associated with cognitive development: responsibility, collaboration, effectivity, independence and reflection.

However, these coefficients are very prone to omitted variable bias, which should be kept in mind when interpreting these results.

VIII. Discussion

There are a few limitations to my study. I will discuss below what these limitations are and how they could have influenced the results. Furthermore, I will discuss the policy implications of my findings.

Data quality

For my study I used data from the COOL5-18 program. To find credible estimates, I need to observe individuals twice; once in group 5 and once in group 8. In the COOL5-18 program however, schools do not necessarily participate each cohort. Thus, each cohort some schools drop out of the study while other schools join in. I ended up with generous information on 2,532 individuals. However, of those 2,532 only 223 were individuals that attended education with a specific pedagogical philosophy. This makes it difficult to claim that my subset of schools with a specific pedagogical philosophy is substantially representative of all schools with a specific pedagogical philosophy. Specifically, it makes my treatment group prone to omitted variable bias and selection issues. Additionally, my 'value-added' design creates a selection issue in the sense that only those students are observed who did not have to retake a year or switched schools between group 5 and group 8. This is not an issue if the treatment and control group were affected equally, however, I cannot be sure that this is the case. Moreover, the codebook which contains variable names in the COOL5-18 program as well as numerical values and their corresponding meanings, differed between cohorts. I do believe I 'cleaned' the data up sufficiently to align all waves, but it does worry me that some misalignments might have slipped through and corrupted the data.

Identifying assumption

I estimated the causal effect of receiving education with a specific pedagogical philosophy on the final test score and on the school advice relative to the school advice based on the final test score, under the identifying assumption that selection on schools is exogeneous. That is, uncorrelated with other variables that could affect both the final test score and school selection. Unfortunately, I cannot be sure that this assumption holds. My data did not give me the means to conduct a (quasi-)experimental design. Therefore, I opted for a multivariate regression design, due to which I cannot be sure that the Conditional Independence Assumption holds. To address this concern, I control for a set of observable factors that could influence both my coefficient of interest as well as the outcome variable. Moreover, inspired by Deming (2014), I control for standardized tests in group 5 to try and create an unbiased baseline level at group 5 which gives my model a value-added interpretation between group 5 and group 8. Lastly, I make use of Oster's test (2019) via the *psacalc* command in Stata to assess to what extent my coefficients of interest are robust to omitted variable bias. Oster's test results weaken the already insignificant effect of receiving education with a specific pedagogical philosophy on the final test score, and weakens the causal interpretation of attending education with a specific pedagogical philosophy on the difference between school advice and advice based on the final test score.

Therefore, it is very much likely that unobservable factors influenced both my variable of interest as well as my outcome variables. For instance, teacher quality of schools is something on which I had no data, while it is good possible that this could have affected both the decision to attend a certain school as well as the cognitive development of a child.

One could point out that my design is prone to mechanism concerns, since the standardized tests in group 5 might be influenced by the specific pedagogical philosophy. While this is true, it is important to notice that these tests are mainly added to the model to observe the value added between group 5 and group 8. A practice that is common in educational economics literature (e.g., Deming 2014; McCaffrey et al. 2003).

Policy implications

My findings give rise to several policy implications. Firstly, from a cognitive development perspective, my findings imply that it is unprofitable to exploit the noncognitive traits that are central to the discussed pedagogical philosophies in regular education as well. However, there are also other gains to noncognitive development that might be beneficial to individuals, or even society, even if this does not immediately translate to higher cognitive results (Jackson 2018; Heckman, Stixrud, and Urzua 2006; Glaeser, Ponzetto and Shleifer 2007; and Deming 2011). For instance, better citizenship might decrease polarization, whilst citizenship development is not measured in the final test.

Moreover, from an equity perspective, policymakers should discuss to what extent it is desirable that students from Montessori and Jenaplan schools receive a higher school advice relative to the final test score, assuming that the conditional independence assumption holds. If children enrol in secondary school tracks that are too challenging for them, this may lead to public costs in terms of redoing a publicly funded school year. In addition, students may become emotionally distressed if they enrol in a school track that is not tailored to their cognitive abilities.

IX. Conclusion

Much research has been conducted towards the effect of certain school types on cognitive outcomes. Little economic research, however, has focused on the effect of schools with a specific pedagogical philosophy on cognitive development. My research contributes to this literature by (1) investigating the effect of education with a specific pedagogical philosophy on the final test score in the Netherlands, and (2) the effect of education with a specific pedagogical philosophy on the difference between the teacher-based school advice and the school advice based on the final test score.

My main findings are that (1) education with a specific pedagogical philosophy is not associated with higher cognitive scores, and (2) attending Montessori and Jenaplan schools is associated with receiving a school advice that is higher than the school advice based on the final test score. (2) is economically and statistically significant, but prone to omitted variable bias. My findings suggest that policymakers should not stimulate the development of noncognitive traits, if policymakers want to improve cognitive development. However, there may be other gains to the development of noncognitive traits such as good citizenship which should not be overlooked. Moreover, policy makers should assess to what extent it is desirable that students at Montessori and Jenaplan schools receive a higher school advice relative to the school advice based on the final test score.

Due to certain limitations, my findings should be read with caution. Firstly, the representativity of my data is questionable. Secondly, my identifying assumption does not hold due to endogeneity concerns.

Given my findings, I suggest that future studies further investigate the effect of pedagogical philosophies on cognitive outcomes, particularly for Waldorf and Freinet education as my data did not provide the resources to examine these pedagogical philosophies. Moreover, I motivate future studies to further investigate the consequences of receiving a different school advice relative to the school advice based on the final test score on a student's future career.

X. Reference list

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, *82*(1), 137-196.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of political economy*, *113*(1), 151-184.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2002a. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools." Working Paper no. 9358 (December), NBER, Cambridge, MA. ———. 2002b. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." Manuscript (April), Dept. Econ., Northwestern Univ.

Barrow, L., Sartain, L., and De la Torre, M. (2020). Increasing access to selective high schools through place-based affirmative action: Unintended consequences. American Economic Journal: Applied Economics, Forthcoming.

Berends, R., & Wolthuis, H. (2014). *Focus op dalton* [Focus on Dalton]. Saxion Dalton University Press.

Beuermann, D. and Jackson, K. (2018). Do parents know best? The short and long-run effects of attending the schools that parents prefer. NBER Working Paper No. 24920.

Clark, D. (2010). Selective schools and academic achievement. The B.E. Journal of Economic Analysis & Policy, 10(1).

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. Econometrica, 74(5):1191–1230.

De Montessori visie. (2019, 29 november). Nederlandse Montessori Vereniging. Consulted on 24 May 2022, van https://montessori.nl/veelgestelde-vragen/montessori-visie/

Deming, D. J. (2011). Better schools, less crime?. *The Quarterly Journal of Economics*, *126*(4), 2063-2115.

Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. NBER Working Paper 19803, National Bureau of Economic Research.

De Bilde, J., Van Damme, J., Lamote, C., & De Fraine, B. (2013). Can alternative education increase children's early school engagement? A longitudinal study from kindergarten to third grade. *School Effectiveness and School Improvement*, *24*(2), 212–233. https://doi.org/10.1080/09243453.2012.749795

Dobbie, W. and Fryer, R. G. F. (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. American Economic Journal: Applied Economics, 3(3):158–187

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. American Economic Review, 101(5):1739–1774.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, *44*(3), 329.

Freinet Vereniging - uitgangspunten. Freinet Vereniging. Consulted on 24 May 2022, van https://www.freinetvereniging.eu/nl/uitgangspunten.html

Glaeser, E. L., Ponzetto, G. A. M., and Shleifer, A. 2007. Why does democracy need education? Journal of Economic Growth, 12(2): 77–99.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, *94*(1-2), 114-128.

Heckman, J. J., Stixrud, J., and Urzua, S. 2006. The Effects of Cognitive and Noncognitive 33

Abilities on Labor Market Outcomes and Social Behavior. Journal of Labor Economics, 24(3): 411–482.

Inspectie van het Onderwijs. (2018, April). *De Staat van het Onderwijs 2018 | Onderwijsverslag over 2016/2017*. https://www.onderwijsinspectie.nl/documenten/rapporten/2018/04/11/rapport-de-staat-van-het-onderwijs

Jackson, K. (2010). Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. The Economic Journal, 120(549):1399–1429.

Jackson (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. Journal of Political Economy

Jenaplanessenties | NJPV. (z.d.). Nederlandse Jenaplan Vereniging. Consulted on 24 May 2022, van https://www.jenaplan.nl/nl/jenaplan/wat-is-jenaplan/essenties

Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. Routledge.

Kohnstamm Instituut, Universiteit van Amsterdam; ITS, Radboud Universiteit Nijmegen (2012): Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2010/2011. DANS. https://doi.org/10.17026/dans-xck-b8kn

Kohnstamm Instituut, Universiteit van Amsterdam; ITS, Radboud Universiteit Nijmegen (2015): Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2013/2014. DANS. https://doi.org/10.17026/dans-xtk-uz6k

Lillard, A. S., Heise, M. J., Richey, E. M., Tong, X., Hart, A., & Bray, P. M. (2017). Montessori preschool elevates and equalizes child outcomes: A longitudinal study. *Frontiers in psychology, 8*, 1783.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). Evaluating value-added models for teacher accountability.

Ministerie van Algemene Zaken. (2022, 3 March). *Schooladvies toelating voortgezet onderwijs*. Schooladvies en eindtoets basisschool | Rijksoverheid.nl. Consulted on 10 mei 2022, from https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitief-schooladvies#:%7E:text=Basisscholen%20geven%20v%C3%B3%C3%B3r%201%20maart,opnieuw%20over%20het%20schooladvies%20nadenken

Ministerie van Algemene Zaken. (2021, 27 July). *Verplichte eindtoets basisonderwijs*. Schooladvies en eindtoets basisschool | Rijksoverheid.nl. Consulted on 10 May 2022, from https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/verplichte-eindtoets-basisonderwijs#:%7E:text=Alle%20leerlingen%20van%20groep%208,speciaal%20onderwijs%20en%20speciaal%20basisonderwijs

Oosterbeek, H., Ruijs, N., & De Wolf, I. (2020). Using admission lotteries to estimate heterogeneous effects of elite schools.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.

SCO-Kohnstamm Instituut Amsterdam; ITS Radboud Universiteit Nijmegen (2007): Cohortonderzoek Onderwijsloopbanen van 5-18 jaar - COOL 5-18 - Basisonderwijs 2007/08. DANS. https://doi.org/10.17026/dans-xxs-4h2a

Seegers, G., Putten, C. M., & Brabander, C. J. (2002). Goal orientation, perceived task outcome and task demands in mathematics tasks: Effects on students' attitude in actual task settings. *British Journal of Educational Psychology*, *72*(3), 365–384. https://doi.org/10.1348/000709902320634366

S. (2022, 7 april). *Pedagogie - Federatie Steinerscholen*. Federatie Steinerscholen. Consulted on 24 May 2022, van https://steinerscholen.be/pedagogie/

Van der Vaart, T., Van der Vaart, T., & Van der Vaart, T. (2021, 11 november). *Home*.

Nederlandse Dalton Vereniging. Consulted on 24 May 2022, van https://dalton.nl/

Van der Wal, M., & Waslander, S. (2007). Traditional and Non-Traditional Educational Outcomes: Trade-off or complementarity? *School Effectiveness and School Improvement*, *18*(4), 409–428. https://doi.org/10.1080/09243450701712502

XI. Appendix

Table A1 Guideline to determine the objective school advice based on the final test score.

Interval test score	School advice (Dutch)	English equivalent	Assigned value
501-523	vmbo-bbl	Secondary vocational education basic level	10
524-528	vmbo-kbl	Secondary vocational education intermediate level	20
529-536	vmbo-tl	Secondary vocational education theoretical level	30
537-544	havo	Higher general secondary education	40
545-550	vwo	Pre-university education	50

Official guideline to help interpreting the final test score and determine what track to advice a student. The score for the final test ranges from 501 to 550. The values in the column 'Assigned value' denote the value for each secondary school track. By labelling a value to each secondary school track, I can quantitatively compare the given school advice to the advice based on the final test score. The values 'x5' are school advices that are in between two secondary school tracks, since many Dutch middle schools allow for mixed tracks. The intervals and their respective school advices are based on the Cito-guidelines, which until 2015 was the main supplier of the final test in the Netherlands.

Table A2 Descriptive statistics of all individuals in group 8 that participated in the COOL5-18 program.

Variables	Regular	Montessori	Jenaplan	Dalton	Total
% female	50.22	49.74	52.50	53.72	50.40
Province %					
Groningen	2.16		4.76	13.11	2.32
Friesland	3.65		6.67	3.69	3.36
Drenthe	3.25			15.73	3.40
Overijssel	7.18		2.10	11.65	7.34
Flevoland	1.66			2.72	1.85
Gelderland	9.67	5.39	19.24	20.87	10.80
Utrecht	2.14	2.96	10.10	10.00	2.36
Noord-Holland	17.00	58.09	29.14	8.45	17.95
Zuid-Holland	20.20	27.83	9.52	8.74	19.60
Zeeland	4.71				4.10
Noord-Brabant	20.98	1.74		5.05	19.69
Limburg	7.40	4.00	18.48		7.22
Urbanity (scale 1-5)	2.87 (1.34)	1.67 (0.81)	3.34 (1.32)	3.45 (1.20)	2.88 (1.34)
Migrant background %	22.39	13.40	6.64	7.85	21.73
Social Economic Status %*	3.89	5.31	4.73	4.65	3.94
1	12.19	5.20	2.63	3.44	11.79
2	14.42	3.95	8.08	10.12	13.68
3	7.11	4.16	2.22	2.23	6.91
4	36.25	17.46	39.60	38.16	36.10
5	3.47	4.16	2.02	2.23	3.38
6	26.56	65.07	45.45	43.83	28.13
Denomination					
Public	32.16	75.13	56.38	75.63	34.16
Roman Catholic	23.88		8	10.97	23.51
Protestant	33.02		23.81	11.94	31.09
Other	10.94	24.87	11.81	1.46	11.23
Highest education father %					
Primary education	10.40	5.36	4.22	3.50	10.06
lower vocational education	28.88	10.54	20.04	19.32	27.61
vocational education	36.75	20.88	36.71	40.13	36.79
Higher education	23.97	63.22	39.03	37.05	25.54
Highest education mother %					
Primary education	12.61	5.44	3.27	4.22	12.16
lower vocational education	23.14	6.90	14.52	16.77	22.21
vocational education	44.60	27.22	46.22	45.37	44.49
Higher education	19.65	60.44	35.99	33.64	21.14
Single parent %	13.44	17.18	11.60	12.41	13.54
Final test score (CITO)	533.54 (10.12)	538.81 (9.36)	534.39 (9.92)	534.86 (9.82)	533.68 (10.15)

Secondary school advice (%)					
VWO	13.09	37.25	16.81	20.24	14.04
HAVO/VWO	8.88	12.94	10.84	7.86	9.15
HAVO	16.76	20.59	15.49	20	16.92
VMBO/HAVO	8	5.69	8.19	8.10	7.67
VMBO	53.27	23.53	48.67	43.8	52.22
Observations	28,240	575	525	1,030	35,066

Sample consists of students categorized per educational type. Means are denoted for every variable. The standard deviations are denoted in parentheses. There are a total of 35,066 observations and individuals. Observations are from triennial panel data over the years 2008 – 2014. Every individual is measured once in group 8. For urbanity, a lower score indicates more urbanity. Social economic status is determined by the migration background of a student and parent's education level. The Social economic status scores represents the following: (1) lower vocational education and migrant background, (2) lower vocational education and native background, (3) vocational education and migrant background, (4) vocational education and native background, (5) higher education and migrant background, and (6) higher education and native background.

Table A3 Regression results of final test scores on attending education either Montessori, Jenaplan or Dalton education.

Final test scores	(1) Montessori	(2) Jenaplan	(3) Dalton
Specific pedagogical philosophy	0.7	-1.4	0.3
	(1.9)	(1.2)	(1.2)
COI in terms of st. dev.	0.068	-0.008	0.034
Test scores group 5	Yes	Yes	Yes
Household background characteristics	Yes	Yes	Yes
Regional characteristics	Yes	Yes	Yes
School denomination	Yes	Yes	Yes
Time variant effects	Yes	Yes	Yes
Constant	532.7	532.8	532.8
Observations	2,055	2,055	2,055
Montessori	100		
Jenaplan		30	
Dalton			76

*Multivariate regression on final test score, which is measured on a scale of 500-550, 550 and 500 being the respective highest and lowest score possible. The row 'COI in st. dev.' shows the coefficient of interest in terms of standard deviations. The columns do not differ in the amount of control variables included. 'Yes' indicates that the respective observable factors have been controlled for in the corresponding regression, 'No' indicates that this has not been the case. Observations for the years 2008 – 2013 are included. The columns show estimated coefficients and standard errors are denoted in parentheses. Standard errors are robust and clustered at the school level. The last three rows show the number of observations per specific pedagogical philosophy that are included in the regression. * p < 0.10, ** p < 0.05, *** p < 0.01.*

Table A4 Regression results of difference in school advice on attending education with a specific pedagogical philosophy

Difference in school advice	(1) Montessori	(2) Jenaplan	(3) Dalton
Specific pedagogical philosophy	3.31***	4.34***	-1.54
	(1.24)	(0.85)	(1.3)
COI in terms of st. dev.	0.496	0.650	-0.231
Test scores group 5	Yes	Yes	Yes
Household background characteristics	Yes	Yes	Yes
Regional characteristics	Yes	Yes	Yes
School denomination	Yes	Yes	Yes
Time variant effects	Yes	Yes	Yes
Constant	0.35	0.83	0.85
Observations	2,013	2,013	2,013
Montessori	98		
Jenaplan		29	
Dalton			76

*Multivariate regression on difference in school advice. Difference in school advice is measured as the size of the difference between school advice and school advice solely based on the final test score. Observations for the years 2008 – 2013 are included. The row 'COI in st. dev.' shows the coefficient of interest in terms of standard deviations. The columns differ in the amount of control variables included. 'Yes' indicates that the respective observable factors have been controlled for in the corresponding regression, 'No' indicates that this has not been the case. Observations for the years 2008 – 2013 are included. The columns show estimated coefficients and standard errors are denoted in parentheses. Standard errors are robust and clustered at the school level. The last three rows show the number of observations per specific pedagogical philosophy that are included in the regression. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A5 Noncognitive outcomes per specific pedagogical philosophy in group 8

Noncognitive trait	Regular	Montessori	Jenaplan	Dalton
Behaviour	3.80	3.613	4.25	3.901
	(0.879)	(0.939)	(0.767)	(0.745)
Reflection	2.258	2.307	2.311	2.388
	(0.578)	(0.586)	(0.494)	(0.586)
Task orientation	4.01	3.90	3.96	3.96
	(0.617)	(0.602)	(0.516)	(0.586)
Cognitive self-confidence	3.74	3.75	3.70	3.71
	(0.63)	(0.66)	(0.48)	(0.61)
Work attitude	3.54	3.40	3.64	3.48
	(0.97)	(1.04)	(1.13)	(0.96)
Relationship with other children	4.18	4.12	4.04	4.17
	(0.67)	(0.59)	(0.49)	(0.73)
Relationship with teacher	3.71	3.56	3.84	3.89
	(0.68)	(0.76)	(0.66)	(0.42)
Observations	2,285	110	35	78
Montessori		110		
Jenaplan			35	
Dalton				78

Sample consists of students categorized per educational type. Means are denoted for every variable. The standard deviations are denoted in parentheses. Observations are from triennial panel data over the years 2008 – 2014. Every individual is measured once in group 8. The data on noncognitive traits are derived from questionnaires filled in by teachers of the individuals. The teacher could respond with the values 1, 2, 3, 4, and 5. With 5 being the highest and 1 the lowest. The last three rows show the number of observations per specific pedagogical philosophy that are included in these statistics.