Economics
Working Papers
2021-13

How to Cope with Dyslexia: The Effects of Special Education on Academic Performance, Personality Traits, and Well-being

Søren Albeck Nielsen

AARHUS BSS
DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

How to Cope with Dyslexia: The Effects of Special Education on Academic Performance, Personality Traits, and Well-being*

Søren Albeck Nielsen

TrygFonden's Centre for Child Research, Aarhus University, Denmark
Department of Economics and Business Economics, Aarhus University, Denmark
Centre for Integrated Register-based Research, CIRRAU, Aarhus University, Denmark

Abstract

I use Danish administrative data to study the effects of a special education intervention for pupils with severe dyslexia. My empirical approach exploits individual level panel data that allow me to track pupils before and after the initiation of treatment. Using a difference-in-difference strategy in combination with entropy balancing, I find positive and persistent effects of the intervention on pupils' reading scores, personality traits, and school well-being. The intervention reduces the reading gap to non-dyslexics with up 33 percent and the well-being gap with up to 80 percent.

JEL Classification: C21, H52, I21, I28
Keywords: Dyslexia, Special education, Assistive technology, Personality traits, Difference-in-difference

*Current version: October 26, 2021. I gratefully acknowledge helpful comments from Marianne Simonsen, Michael Rosholm, Helena Skyt Nielsen, Lars Skipper, Kristoffer Holst Ibsen, Bastien Michel, participants in the EALE conference, and seminar participants at Aarhus University and the Norwegian School of Economics. The usual disclaimers apply. Corresponding author: snielsen@econ.au.dk.

1 Introduction

This paper studies a specialized intervention targeted children with dyslexia. Dyslexia is a genetic disorder that affects 3% to 10% of the population (Snowling, 2013). It affects the ability to spell, read, and write (Hebert et al., 2018) and may even have far-reaching consequences for children's lives such as low school grades, poor educational attainment, and behavior problems (Undheim, 2009; Epnion, 2018; Einar et al., 2001). It is possible, however, that this academic underperformance and negative behavior are a sign of inefficiencies in the educational system since the existing literature finds no relationship between dyslexia and intelligence (Snowling et al., 2020).

I provide the most extensive evidence yet, regarding the impact of education programs for pupils with dyslexia. In particular, I use population-level administrative Danish individual level panel data covering the period 2010–2019 to study the effect of a 10 consecutive weeks dyslexia learning program with one and a half years follow up. I investigate, first, how the learning program affects the pupils reading abilities as well as personality traits and school well-being. Secondly, I ask whether the observed effects are persistent over multiple time-periods.

The analysis delivers notable results across all outcome groups. I find positive effects on two out of three areas of reading (language and text comprehension) with effect sizes of 22% to 25% of a standard derivation. Thus, the program participants increase their reading abilities well beyond the level of the population of dyslexics. In fact, the intervention reduces the reading gap to non-dyslexics with 22% to 33%. Likewise, I find positive effects on two out of the three personality traits investigated in this study with effect sizes between 15% to 17% of a standard derivation (conscientiousness and emotional stability). Moreover, the intervention increases the participants' school well-being by 15% to 20% of a standard derivation (school connectedness, learning self-efficacy and classroom management). In fact, participants increase their level of personality traits and school well-being beyond the level of the population of dyslexics. Compared to non-dyslexics, the program significantly reduces the gap with up to 80%. Secondly, I show that the effects on reading scores are persistent over the four post-treatment school years I am able to observe. In terms of emotional stability, the event study show short-run effects that fade away over time. For conscientiousness and learning self-efficacy, the effect increases throughout the post-intervention time-periods with effect sizes up to 40% of a standard derivation.

A large and growing number of studies estimate the causal effect of school-based policies and reforms targeting low-achieving pupils. However, there is limited evidence for the effect of special education policies. Schwartz et al. (2021) conclude that general special education in New York City improves academic performance (effect sizes around 0.11 standard deviations) for pupils with learning disabilities, which among others include pupils diagnosed with dyslexia. The impact is largest when entering special education in the earlier grades. These results are

in line with previously published studies involving general special education and its effect on academic abilities (Hurwitz et al., 2020; Morgan et al., 2010; Hanushek et al., 2002; Reynolds and Wolfe, 1999). To the best of my knowledge, no papers have estimated causal effects of a program specifically targeting pupils with dyslexia.

This study relies on population-wide register-based data for Danish pupils and their parents, which provides me with longitudinal information about relevant scholastic and well-being outcomes. From the national reading test that occurs every two years during primary schooling, I am able to measure the following three key aspects of reading: 1) language comprehension, 2) decoding, and 3) text comprehension. Additionally, I use the yearly national well-being survey to construct validated psychometric measures for personality traits and school well-being. I combine these data with a long range of socio-economic background characteristics and the membership list from the Danish Library and Expertise Center for people with print disabilities (henceforth NOTA). The national NOTA register allows me to identify pupils diagnosed with dyslexia and who have not participated in the intervention.

I use a difference-in-difference research design, where I exploit that pupils in practice receive the intervention between grade 4 and 8. Thus, I observe outcomes before and after the intervention, which enables me to estimate treatment effects using a fixed-effects analysis. Importantly, the difference-in-difference approach implicitly controls for selection on time-invariant unobserved characteristics such as genetics and intelligence. Access to the NOTA membership list allows me to construct a comparison group that is similar in terms of reading disadvantage but not exposed to the special education intervention.

A key assumption behind the difference-in-difference design is that there can be no differential trends between the treatment and comparison group in the absence of treatment. To address this, I combine my standard individual difference-in-difference strategy with matching techniques. In practice, I use entropy balancing that relies on a maximum entropy reweighting scheme that calibrates individual weights such that the reweighted group satisfy a set of pre-specified balance conditions (Hainmueller, 2012). Incorporating entropy balancing in the econometric framework ensures exact balancing between the NOTA and intervention group, not only concerning the mean but also on higher moments of the large set of observed covariates. Additionally, I use event study graphs to investigate how the effect evolves over time and to test for similar pre-trends. In the sensitivity checks, I use other specifications for being dyslexic and alternative balancing strategies in order to explore the credibility of the estimated treatment effects. The results are robust to these sensitivity checks.

I thus provide evidence on learning programs targeting dyslexics and complement the existing literature on assistive technology. The present study contributes to the literature by being the first to estimate causal effects of a learning program specifically targeting pupils with dyslexia on reading scores, personality traits, and school well-being. Especially the ability to investigate the effects on personality traits - an important prerequisite for future academic achievement – is new in the general special education literature.

The remainder of the paper is structured as follows: Section 2 describes the background and institutional settings, such as the key components of the intervention studied, the selection of a valid comparison group, and the Danish school system. I describe the data in section 3 and the empirical strategy in section 4. Section 5 presents the results and robustness tests. Section 6 provides a discussion and interpretations of the findings. The last section concludes.

2 Institutional Settings

In this section, I present the structure, components and objective of the intervention under study. Secondly, I provide a description of the comparison group and discuss treatment as usual. Finally, I briefly introduce the Danish compulsory school system.

2.1 Reading Competency Center for Dyslexics

The intervention under study - Reading Competency Center for Dyslexics (RCCD henceforth) - is a specialized learning program targeted dyslexics at public schools grade 4 to 8. It takes place in the municipality of Aarhus, the 2nd largest municipality in Denmark. The intervention last for approximately one and a half years with a total (annual) cost of 18,000USD (12,000USD) per pupil. The cost of the program seems high but with an annual cost of 29,500USD for special needs teaching 9 hours per week (Nørgaard et al., 2018) and a total cost of 3,600USD for 10 weeks of regular school teaching the program has potential to be cost-effective. Also in the short-run, if the program improve academic abilities to a level where special needs teaching is no longer necessary.

The objective of RCCD is to enhance dyslexic pupils academic skills by focusing on qualifications in general use of assistive technology and at the same time practice conventional reading and writing. Through this, the intention is to enhance the pupil's abilities to allow for participation in age-appropriate teaching in their local classroom, and thereby acquire age-appropriate knowledge. Additionally, the intervention has a broader pedagogical aim. The pupils should not only develop their academic skills but also regains faith in own abilities, getting good independent work habits and achieve greater self-esteem in relation to academic learning such that the pupils become self-reliant in relation to continuing education and training.

The structure of the RCCD program consist of four steps. First, the pupil is assigned an RCCD consultant who is responsible for the pupil's learning throughout the program. The consultant meets with the local teachers and forms an overview of the opportunities at the pupil's local learning environment in order to design an individual action plan. Secondly, the pupil attends a 10 consecutive weeks learning camp at RCCD's location. The teaching takes place in groups with approximately five pupils and consists of 30 weekly lectures in all compulsory school subjects. During the 10 weeks, the pupil trains proper use of assistive

technology combined with additional focus on non-cognitive skills such as mindset and self-awareness. After the 10 weeks learning camp, the RCCD consultant facilitates the transfer of the pupil's new acquired learning strategies and methods of learning to the pupil's local learning environment. Finally, the consultant regularly follows up and provides advice and guidance for the pupil, parents, teachers and the management team at the local school to ensure that the pupil continues to make progress.

The 10 weeks learning camp consist of three key components, 1) small group instructions, 2) Non-cognitive skills training, and 3) practice the use of assistive technology, which combined have the potential to enhance academic performance, personality traits, and school well-being for this disadvantaged group of pupils.

Small group instructions. Dietrichson et al. (2020a,b) conduct two systematic reviews that combine different types of instructional methods[1] for pupils at risk of academic difficulties in grade 0-6 and grade 7-12 that are evaluated using standardized tests in reading and mathematics. They conclude that small group instruction (1-5 pupils per teacher) has large positive effects on test scores. In fact, the effect of small group instruction is nearly double the effect of the second best instructional component. Few of the included studies investigate the effect more than three months after the intervention, and therefore, little evidence for the persistence of these effects exists.

Non-cognitive skills training. Non-cognitive skills are widely recognized to play an important role for academic performance (Andersen et al., 2020). Duckworth et al. (2007), for example, show no correlation between grit[2] and intelligence. Nonetheless, the authors demonstrate the importance of grit for future academic performance. Their findings suggest that achievement of academic targets not only involves genes and socio-economic background but also personality traits.

Importantly, non-cognitive skills are malleable. A large review by Kautz et al. (2014) summarizes the literature on interventions targeting cognitive and non-cognitive skills. The authors find interventions targeting both cognitive and non-cognitive skills to be superior and argue that it is important to consider non-cognitive in addition to cognitive skills when evaluating interventions. For example, recent studies by Alan and Ertac (2018) and Alan et al. (2019) show, using randomized controlled trials, large lasting effects on both cognitive and non-cognitive outcomes of training non-cognitive skills related to patience (the former study) and perseverance (the latter study) for 12 weeks on grade 2 pupils in Turkey.

[1] Instructional methods include among others Coaching of personnel, Incentives, Computer-assisted instruction, Peer-assisted, Progress monitoring and Small and Medium group instruction.

[2] Grit is a positive non-cognitive trait on passion and perseverance for long-term goals. Individuals high in grit are able to maintain their motivation and determination over long periods despite failures and adversities (Duckworth et al., 2007).

Training the use of assistive technology Assistive technologies are alternatives to traditional reading and writing for individuals with dyslexia. Currently, dyslexics have permanent access to assistive technology apps through their devices, which enable them to better understand text and, thus, ease their participation in regular teaching on equal terms (Lindeblad et al., 2017).

One of the issues that assistive technology does not solve yet is to ensure that the pupil does not become a passive listener who just makes the technology read words or text aloud. The pupils must still be able to understand the individual words and understand the context of the text. Thus, there are still important didactic challenges in the application of assistive technology (Svendsen, 2017). The RCCD intervention teaches the use as well as the pros and cons of each assistive technology tool such that the tools become a natural part of the dyslexic's everyday life selection. Thus, the dyslexics must be able to identify the problem and then select the best assistive tool.

SBV (2014) and Perelmutter et al. (2017) conclude in their systematic literature reviews on assistive technology that causal evidence is limited. The existing studies rely on few observations or low quality methods, and thus, there is not enough evidence to conclude that assistive technology in general affects academic performance and well-being. New studies should exploit more comprehensive, systematic, longitudinal, and in-depth investigation methods (Perelmutter et al., 2017; Haßler et al., 2016).

2.2 Danish Library and Expertise Center for people with print disabilities

There exists no national register for dyslexia prior to the implementation of the national dyslexia test in 2015 by the Ministry of Education. Instead, I exploit the membership list from NOTA to construct a valid comparison group. NOTA is an institution under the Danish Ministry of Culture that produce, buy and sell audio books and e-books and develop synthetic speech for individuals with reading disabilities. Consequently, the NOTA membership list is the best Danish database of pupils with dyslexia during the RCCD intervention period.

Membership of NOTA does not occur automatic after a dyslexia diagnose. To become a member, the pupil or the parents must document that the pupil cannot read ordinary printed text. Thus, members of NOTA are by themselves or with the help of their parents actively seeking help to cope with their disability. In practice, either the school principal or a professional with the competence to test for dyslexia must sign the registration form as evidence for the dyslexia diagnose.

This application process causes a potential selection problem because not all dyslexics become NOTA members. This selection, however, does not affect the internal validity of the study since every RCCD participant is also a member of NOTA. A NOTA membership provides access to learning materials that is only relevant for pupils who are using assistive technology. Thus, NOTA members have access to and uses assistive technology but do not receive the

same extensive training in its application and potential. Therefore, I am estimating the effects of RCCD relatively to the average intervention for NOTA members. A recent report concludes using survey data that interventions targeted dyslexics are widespread used throughout the Danish municipalities. However, the structure of the interventions varies across municipalities from providing assistive technology tools to comprehensive learning programs such as RCCD (NOTA, 2019). In section 6, I discuss the implication of *treatment as usual* in further details.

2.3 Primary and lower secondary education in Denmark

The Danish school system consists of 10 mandatory grades and an optional grade 10. The typical school starting age is the year when the child turns six in which the child enrolls in grade 0. Grade 0 is a transition year taking place at the school, where pupils learn to go to school, but there is no explicit instruction in any academic subjects. Grade 1 to 9 consist of nine years of primary and lower secondary education. Grade 10 is optional and designed for pupils in need of one additional year to be ready for upper secondary education.

Dyslexia is difficult to diagnose with certainty at a young age and there exists no test for kindergarten children. However, signs of dyslexia are if a child begins to speak at a late age, have problems with rhyming, and slowly learn new words. In primary and lower secondary school, the possibilities for detecting dyslexia and implementing interventions are easier. Therefore, the Ministry of Education introduced in 2015 the risk of dyslexia test (grade 0 to 1) and the national dyslexia test (grade 3 to 9). The risk of dyslexia test is a test for early identification of pupils at risk of developing severe decoding difficulties, including dyslexia. It indicates risk of dyslexia but not necessarily dyslexia such that the teachers can implement a preventive intervention. If there is still suspicion of dyslexia at the end of grade 3 the teachers test using the national dyslexia test and make a definitive diagnosis.

3 Data and descriptive statistics

The sample consists of 46,348 NOTA pupils in public schooling, of which 513 pupils were enrolled in RCCD at some point in between 2010/2011 and 2018/2019. To investigate how RCCD affects pupil outcomes, I leverage Danish administrative register data that covers the entire population of children in elementary schooling and is available through Statistics Denmark. Focal to this study is the NOTA membership list that is the best national-wide register for individuals with dyslexia and the Danish Student Register, which is a unique longitudinal dataset. This enables me to follow dyslexics schooling information such as school and classroom movements, type of school as well as special needs teaching from the school year 2009/2010 to 2018/2019. I augment this data with rich socio-economic information describing demographics, ethnicity, marital status, and education, just as I exploit information regarding reading abilities from the Danish Ministry of Education. Finally, for the period 2015-2019, I construct measures of personality traits and school well-being using the Danish Ministry of

Educations national well-being survey.

On average, 57 pupils participate in the RCCD program each school year, half of them starting in September and the other half in February. Prior to being considered a candidate for RCCD, the pupils must have been through prolonged or repeated special educational interventions at their local schools such as reading courses, group or one-to-one tuition, additional classroom support, etc. without any or minor positive effects. Around half of the RCCD participants enrolled in the program at grade 5 and 6. This reflects the fact that dyslexia is usually determined at grade 4 due to the timing of the national dyslexic test.

Next, I provide an overview of the outcome variables. The outcome variables cover three areas: Reading abilities, personality traits, and well-being. Table 1 provides an illustration of the timeline for the outcomes. Finally, I characterize the RCCD pupils against NOTA members and non-dyslexics.

Table 1: Timeline of outcome variables

	Grades	**Years**
Reading abilities	2nd, 4th, 6th, and 8th	2010-2019
Personality traits	4th-9th	2015-2019
School Well-being	4th-9th	2015-2019

Notes: All variables are measured in the spring semester of the year.

3.1 Reading Abilities

Reading performance variables originate from the mandatory national test in reading that takes place in the spring of grade 2, 4, 6 and 8. The national reading test was introduced at a national scale in 2009/2010. It simultaneously tests three cognitive domains of reading, called profile areas: 1) Language Comprehension, 2) Decoding, and 3) Text Comprehension. A clear advantage of the national test is that it is IT-based, self-scoring and adaptive. Thus, the computer does the scoring automatically such that it reflects objective reading abilities and not influenced by teacher opinions such as classroom misbehavior. Instead of giving all pupils the same questions, the national test program re-estimates a new ability level after each question and adjusts the difficulty level of the next question. Therefore, the final ability measure for each profile area is a function of the difficulty level of the questions and the ability of the pupil. The final scores are measured on a continuous logit scale distributed from -7 to 7. See Beuchert and Nandrup (2017) for a thorough description of the Danish national tests. On the individual level, national test scores explain 48% to 51% of the variation in average Danish and math exam performances. Across all subjects, a 1 SD increase in test scores is associated with approximately 2 grade points in the GPAs, a 19% higher probability of enrolling in upper secondary education, and a 16% higher probability of completing general

upper secondary education (Beuchert and Nandrup, 2017).

I standardize the reading ability measures on the full population within school years and profile areas to mean zero and standard deviation of one to render the results comparable to effect sizes of other studies.

Table 2 shows summary statistics of RCCD participants, NOTA and non-NOTA members. The reading abilities are measure one or two school years prior to RCCD participation. It is evident that RCCD participants reading abilities are below other dyslexics and especially non-dyslexics. Particularly for decoding abilities, where they perform 0.35 SD below other dyslexics and 1.6 SD below non-dyslexics. This is in line with the literature as dyslexia involves problems identifying speech sounds and learning how they relate to letters and words, i.e. dyslexia makes the decoding process difficult, regardless of the pupil's level of academic skills. This lack of decoding abilities influences also *language comprehension* and *text comprehension*, where RCCD scores approximately 0.1 and 0.9 SD below respectively other dyslexic and non-dyslexics. Thus, if pupils cannot decode words, they cannot understand what the text means, even if they understand the meaning of words in conversations.

Table 2: Descriptive Statistics

	(1) RCCD		(2) NOTA		(3) Non-NOTA		(4)	(5)
	Mean	SD	Mean	SD	Mean.	SD.	(1)-(2)	(1)-(3)
Demographics								
- Boy (1/0)	0.600	(0.490)	0.561	(0.496)	0.505	(0.500)	*0.040*	**0.095**
- 1st or 2nd generation immigrant (1/0)	0.103	(0.305)	0.060	(0.237)	0.110	(0.313)	**0.044**	-0.007
- Preventive personalized interventions (no.)	0.035	(0.239)	0.053	(0.293)	0.033	(0.236)	-0.017	0.002
- Siblings (no.)	1.380	(0.979)	1.328	(0.928)	1.318	(0.913)	0.052	0.062
- Living with both parents (1/0)	0.661	(0.474)	0.647	(0.478)	0.671	(0.470)	0.014	-0.010
- School changes (no.)	0.232	(0.537)	0.431	(0.695)	0.328	(0.611)	**-0.199**	**-0.097**
- Class changes (no.)	0.055	(0.236)	0.092	(0.311)	0.038	(0.203)	**-0.038**	*0.017*
- Relocations (no.)	1.175	(1.382)	1.735	(1.575)	1.629	(1.472)	**-0.560**	**-0.454**
- Special needs teaching (1/0)	0.056	(0.229)	0.106	(0.308)	0.033	(0.178)	**-0.051**	**0.023**
- Mother's age at birth (years)	30.907	(4.864)	29.715	(4.870)	30.163	(4.825)	**1.193**	**0.744**
- Father's age at birth (years)	33.254	(5.320)	32.379	(5.615)	32.845	(5.675)	**0.875**	0.409
Mother's marital status								
- Cohabiting couple (1/0)	0.107	(0.310)	0.098	(0.297)	0.082	(0.275)	0.010	**0.025**
- Divorced (1/0)	0.181	(0.386)	0.161	(0.367)	0.152	(0.359)	0.021	*0.029*
- Married (1/0)	0.620	(0.486)	0.658	(0.474)	0.686	(0.464)	*-0.038*	**-0.066**
- Single (1/0)	0.080	(0.271)	0.072	(0.259)	0.067	(0.250)	0.008	0.013
Father's marital status								
- Cohabiting couple (1/0)	0.105	(0.307)	0.095	(0.293)	0.081	(0.273)	0.010	**0.024**
- Divorced (1/0)	0.136	(0.344)	0.152	(0.359)	0.140	(0.347)	-0.015	-0.004
- Married (1/0)	0.643	(0.480)	0.651	(0.477)	0.681	(0.466)	-0.008	*-0.038*
- Single (1/0)	0.078	(0.268)	0.066	(0.248)	0.057	(0.231)	0.012	**0.021**

Mother's highest educational degree								
- No degree or primary school (1/0)	0.170	(0.376)	0.216	(0.412)	0.166	(0.372)	**-0.047**	0.004
- High School (1/0)	0.060	(0.239)	0.045	(0.206)	0.061	(0.239)	*0.016*	0.000
- Vocational traning (1/0)	0.343	(0.475)	0.443	(0.497)	0.349	(0.477)	**-0.100**	-0.006
- Academy higher education (1/0)	0.060	(0.239)	0.047	(0.211)	0.051	(0.220)	0.014	0.010
- College (1/0)	0.240	(0.427)	0.196	(0.397)	0.249	(0.432)	**0.044**	-0.009
- University (1/0)	0.127	(0.333)	0.053	(0.225)	0.125	(0.330)	**0.073**	0.002
Father's highest educational degree								
- No degree or primary school (1/0)	0.226	(0.419)	0.278	(0.448)	0.219	(0.414)	**-0.052**	0.007
- High School (1/0)	0.047	(0.211)	0.029	(0.168)	0.054	(0.226)	**0.018**	-0.007
- Vocational traning (1/0)	0.378	(0.485)	0.497	(0.500)	0.395	(0.489)	**-0.119**	-0.016
- Academy higher education (1/0)	0.109	(0.312)	0.063	(0.244)	0.076	(0.264)	**0.046**	**0.034**
- College (1/0)	0.127	(0.333)	0.077	(0.266)	0.123	(0.329)	**0.050**	0.003
- University (1/0)	0.113	(0.317)	0.056	(0.230)	0.133	(0.340)	**0.057**	-0.020
Nationale reading test								
- Language Comprehension (std.)	-0.823	(1.175)	-0.697	(1.125)	0.091	(0.935)	**-0.126**	**-0.914**
- Decoding (std.)	-1.475	(0.894)	-1.127	(0.917)	0.143	(0.902)	**-0.348**	**-1.618**
- Text Comprehension (std.)	-0.862	(1.053)	-0.770	(0.984)	0.101	(0.944)	**-0.093**	**-0.963**
School well-being								
- School Connectedness (std.)	-0.115	(0.931)	-0.100	(1.040)	0.089	(0.975)	-0.015	**-0.204**
- Learning Self-Efficacy (std.)	-0.494	(0.936)	-0.429	(1.013)	0.101	(0.975)	-0.066	**-0.595**
- Learning Enviroment (std.)	-0.077	(1.007)	0.067	(1.012)	0.133	(0.982)	*-0.143*	**-0.210**
- Classroom Management (std.)	-0.246	(1.048)	-0.155	(1.059)	-0.007	(0.988)	-0.091	**-0.239**
Personality traits								
- Conscientiousness (std.)	-0.387	(0.960)	-0.367	(1.052)	0.073	(0.990)	-0.020	**-0.461**
- Agreeableness (std.)	-0.220	(1.101)	-0.223	(1.076)	0.035	(0.977)	0.003	**-0.255**
- Emotional Stability (std.)	-0.111	(0.912)	-0.148	(1.076)	0.052	(0.986)	0.037	**-0.163**
Absence								
- Sick (percent)	2.899	(3.397)	3.275	(3.886)	3.093	(3.838)	*-0.376*	-0.194
- Illegal (percent)	0.719	(1.948)	0.721	(2.436)	0.725	(2.517)	-0.002	-0.006
- Legal (percent)	1.263	(1.799)	1.446	(2.477)	1.473	(2.327)	-0.183	*-0.211*
Number of individuals	513		45,839		431,009			

Notes: The table shows the descriptive statistics for RCCD participants against NOTA and non-NOTA members. Column 4 (5) shows difference between RCCD and NOTA (non-NOTA) from a regressing each covariate on the RCCD indicator. All variables are measured the school year prior to treatment. Except National tests that is measured up to two school years prior. The table is based on non-missing data.

3.2 School Well-being and Personality traits

The Danish Ministry of Education implemented a 40-item well-being questionnaire in 2014/2015, which is a yearly national survey for all public school pupils in grade 4 to 9. The Danish Well-being Survey is electronically distributed to all pupils. The survey is part of the regular teaching, and all pupils in the class must respond to the survey during the same lesson. Teachers are to tell the pupils that they should respond honestly and stress that results are anonymous to their parents, teachers, or other employees at the school. The teachers are

encouraged to read questions aloud if a pupil has difficulties understanding them. Additionally, the teachers have the authority to exempt pupils from the survey if they do not feel well answering the survey or if they are not capable of answering the questionnaire. This could for example be special needs pupils (Andersen et al., 2020; Niclasen et al., 2018). Even though the survey is a mandatory part of the general curriculum, there are no consequences if the pupil is not answering due to sickness or other types of absences. Table A.1 shows the average response rate for RCCD participants, NOTA, and Non-NOTA members. It is clear from the table that the RCCD group is a special group where the average response rate is 76%, which is 4.6% (7.5%) lower than the NOTA (non-NOTA) group.

One way of exploiting the national well-being survey is to construct three personality trait scales from the Big Five Model (Andersen et al., 2020). Agreeableness relates to how pupils tend to treat relationships with others. Conscientiousness describes pupils' ability to be responsible and work carefully to get things done. Emotional Stability measures absences of characteristics such as anxiety, insecurity and self-pitying. Importantly, the authors find a high correlation between conscientiousness and future academic performance.

Another way of exploiting the information in the national well-being survey is the one recommended by Niclasen et al. (2018). They propose a four-factor model that includes 27 of the 40 items from the national well-being survey. The four-factor model measures School Connectedness, Learning Self-Efficacy, Learning Environment and Classroom Management. According to the authors, these four scales measure central aspects of school well-being. School Connectedness is the belief held by the pupil that teachers and peers in the school care about their well-being and learning. Learning Self-Efficacy describe pupils beliefs and attitudes toward their capabilities to achieve academic success. Learning Environment deals with pupils' experience of motivation and co-determination, as well as the help and support of teachers and the surroundings in order to complement future learning. Classroom Management measures pupils' experience of the classroom as well as classroom management by the teachers.

Table A.2 shows the items used to construct each scale. The scales are standardized, first by standardizing each item, then calculating the average across all standardized items in each scale and finally standardizing the overall scales.

The school well-being as well as personality traits of the RCCD are at the same level as other dyslexics across all scales. However, when comparing RCCD pupils with non-dyslexics it is evident that they are far less satisfied with their schooling and their personality traits are significantly below the average population (see table 2). Especially, for *Learning self-efficacy* and *Conscientiousness* in which RCCD participants scores up to 0.6 SD below the non-dyslexics. Furthermore, *School Connectedness*, *Classroom Management*, *Learning Environment*, *Agreeableness* and *Emotional stability* all show a difference of approximately 0.2 standard derivations. This indicates that dyslexics do not thrive in school.

3.3 Covariates

Table 2 also presents a summary statistics for rich set of socioeconomic variables, recorded the school year prior to RCCD enrollment. The descriptive statistics show that there is an overrepresentation of boys participating in the RCCD program with 4% fewer boys in the NOTA register and 10% fewer in the non-NOTA (non-dyslexic) group. There are 4% fewer immigrants in the NOTA group compared to both the RCCD and non-NOTA groups, indicating that not all dyslexic immigrants are member of NOTA. Additionally, the RCCD pupils have relocated approximately one time during their life whereas the NOTA and non-NOTA group is closer to two times. This is also visible in the number of school movements where the RCCD pupils have significantly fewer. Interestingly, I observe that parents of the RCCD pupils are higher educated than the NOTA group parents and in line with the non-NOTA parents, i.e. RCCD pupils are inclined to live in a household of high socioeconomic status (SES henceforth) compared to other dyslexics. For the mothers (fathers) 43% (36%) have acquired an academy higher education, College or University degree where these numbers are 30% (20%) for the NOTA group and 43% (33%) for the non-NOTA group. Related to the educational level, the parents of the RCCD pupils are on average one year older when their child is born.

4 Empirical Strategy

The overarching goal of this paper is to estimate the causal effect of RCCD on reading performance, personality traits, and school well-being. Clearly, the key challenge in such an analysis is to estimate outcomes in the absence of RCCD participation. One might worry that non-random selection into RCCD participation challenge the identification. Thus, pupils participating in RCCD comprise a different population than pupils who do not participate.

In order to address this potential endogeneity problem, I use a difference-in-difference strategy corresponding to a fixed effect analysis. My strategy first compares one pupil's outcomes after RCCD participation with the same pupil's outcomes before RCCD participation. This first difference accounts for time-invariant individual outcomes. However, it is unlikely that individual level outcomes do not change over time. To account for this, I exploit information not only for the RCCD pupils, but also for other dyslexics (NOTA members) and the ability to follow their outcomes over multiple pre and post treatment time-periods. The comparison group of NOTA pupils is assigned a synthetic (random) participation grade for practical reasons.

The analyses begin with the following event study equation:

$$y_{it} = u_i + \sum_{j=-5, j\neq 0}^{4} \delta_j \cdot 1(j = t)_{it} \cdot 1(RCCD = 1)_i + \sum_{j=-5, j\neq 0}^{4} \gamma_j \cdot 1(j = t)_{it} + \epsilon_{it} \qquad (1)$$

where y is the outcome of interest, 1(RCCD) indicates RCCD participation and $(j = t)$ are

time indicators relatively to participation. δ_j are the effects of RCCD participation by school year relative to the school year prior to participation ($j = 0$). γ_j are the effects for the NOTA pupils relative to their "random" participation school year. u_i is individual level fixed effect and ϵ_{it} is the error term. Thus, the coefficient of δ measures the average treatment effect of the treated (ATT henceforth) and is the parameters of interest. Standard errors are clustered at the individual level.

Ideally, I would observe the counterfactual, i.e. what would have happened to the RCCD group in absence of treatment. However, the counterfactual is unobservable, and instead I exploit the panel structure of the data that enables within pupils estimation with a comparison group, which for identification rely solely on the parallel trends assumption. Thus, the fixed effect estimator produces causal effects if and only if the RCCD and NOTA groups would have had the same trends in the post-periods in absence of treatment. The parallel trend assumption is a much weaker identifying assumption than models based on the selection on observables assumption. Unobserved individual time-fixed heterogeneity will not bias the estimations. Only differing time-trends in the treatment and comparison groups will bias the fixed effect estimation. In the robustness section 5 below, I also report the results when using alternative comparison groups and discuss the parallel trend assumption.

4.1 Difference-in-Difference Strategy with Entropy Balance

Since the parallel trends assumption is critical for identification, I extend the fixed effect framework to account for any observed difference between the RCCD and the NOTA groups prior to treatment. I combine the fixed effect estimator with Entropy Balancing. The reason for combining fixed effect with a weighting strategy is to reduce bias due to different distributions of covariates in the RCCD and NOTA groups (Heckman et al., 1997; Blundell et al., 2004; Abadie, 2005). Entropy Balancing is a data processing method that obtain covariate balance with a binary treatment variable (Hainmueller, 2012). Hainmueller and Xu (2013) explain entropy balance as a generalization of the propensity score adjustment method suggested by Rosenbaum and Rubin (1983) that addresses its limitations. The propensity score is typically calculated using a logistic regression, and the resulting balance is assessed to see if the individual assigned weights equalizes the covariates between the two groups. Whereas, Entropy Balance directly calculates weights by integrating covariate balance directly into the weights. See Hainmueller and Xu (2013) and Hainmueller (2012) for a thorough description of the entropy balance method.

Table A.3 shows descriptive statistics for pupils and parents background characteristics measured in the school year prior to RCCD such that they are not affected by the intervention. Column 2 to 5 reports the unadjusted mean, standard deviation and difference in means between the RCCD and NOTA groups. It is evident that pupils of high SES parents have a higher propensity to participate in the RCCD program. Thus, pupils of mothers (fathers) with either a college or a university degree are 12% (11%) more likely to be in the RCCD

group. Column 5 in table A.3 also shows that pupils in the RCCD group have fewer school, classroom, and place of residence movements as well as older parents. Additionally, 5% fewer are receiving additional hours of special needs teaching and 4% more are 1st or 2nd generation immigrants.

The ex ante differences on observed characteristics discussed above are a threat to the parallel trend assumption. For example, if we believe that additional hours of special needs teaching or low SES flattens the pupils learning curve this will violate the parallel trends assumption simply because the RCCD pupils will accumulate skills faster, and the fixed effect estimate is upwards biased. Entropy Balance deals with this threat to identification by ensuring perfect balance between the RCCD and NOTA groups on all observed covariates prior to treatment. Ryan et al. (2018) illustrate using simulations that a combination of a balancing strategy and the difference-in-difference framework does well at dealing with non-parallel trends in a context of health care policy interventions.

Table A.3 shows the covariates used in the balancing specification where the first moment of the covariates are balanced[3]. In section 5.3 I test alternative specifications of the balancing strategy to ensure the credibility of the results. The last column in table A.3 shows that the entropy balance ensures no difference in the means between the RCCD and NOTA groups prior to treatment on observed factors. Thus, Entropy Balancing ensures that the NOTA group is on average similar to the RCCD group.

5 Results

This section presents the results. First, I show the event study representation of the effects on reading abilities, personality traits, and school well-being. Additionally, I supplement the event-study model with a minimalist difference-in-difference setup that do not allow effects to vary with distance to RCCD participation. Then, I perform a set of robustness checks to ensure the validity of the estimates and finally, conduct a set of heterogeneity analyses.

5.1 Event study representation

I start by analyzing the effects of RCCD using a event study representation. Figure 1 presents the ATT at each time-period, with the time-period leading up to RCCD (t=0) as baseline for the three profile areas of the national reading test. The national reading test takes place every second school year starting at grade 2 and ending at grade 8 as illustrated in table 1. Thus, it is important to notice that only pupils treated in grade 4 or 5 are included in periods 3 and 4 simply because the older pupils would attend grade 9 or have left lower secondary schooling at that time-point. Critically, I observe, for all three graphs, no differential pre-

[3]Balancing on the pupils basic demographics such as gender, age, ethnicity, number of siblings, living arrangement, school and classroom changes and special needs teaching, as well as parents marital status, parents educational level and one period pre-trend in reading performance, personality traits, and school well-being.

treatment trends, i.e. the differences are not significantly different from zero. Note that slight differences do appear in the earliest time-periods, however, with large uncertainty do to data constraints. Only grade 7 and 8 pupils have reading tests four or five school years prior to participation. Both the language comprehension (a) and text comprehension (c) graphs show large immediate treatment effects of 24% and 20% of a standard derivation already in the period after the RCCD camp. For language comprehension there is a small decease in the effect size in the later periods whereas there is a similar increase for text comprehension. These changes are small and not significantly different from the previous period, and thus, the effect on reading shows signs of being persistent over time. In terms of the effect on decoding (b), the line is close to being flat indicating that RCCD has no impact on future decoding abilities. This is not surprising since RCCD concentrates on providing the pupils with the necessary tools to participate in normal classroom teaching. Thus, RCCD teaches pupils the use of assistive technology and reading strategies such that they are able to read and write text at an age-appropriate level.

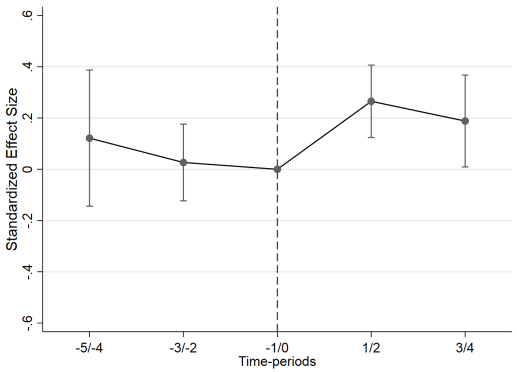
Figure 2 and 3 show similar graphs for the effect on personality traits and school well-being. All scales originate from the national well-being survey, which takes place once a year from grade 4 and started in the spring of 2015 as explained in section 3 and table 1. Thus, fewer pupils are included in these analyses and hence, the standard errors are larger and pre-RCCD investigations are only possible for the pupils treated in grade 6 to 8.

In figure 2 (a) the RCCD pupils have a higher level of conscientiousness two years prior to treatment. However, the difference is not significantly different from zero. For agreeableness (b) the two groups' pre-treatment difference is smaller and for emotional stability (c) it is close to zero. When investigating the effect across time-periods, I first note that RCCD affects conscientiousness positively and lifts the pupils to a new higher level of 25% of a standard derivation. Interestingly, the effect appears to be self-reinforcing since the conscientiousness level keeps increasing over time. However, this increase diminishes over time and seems to stabilize in the fourth time-period. In figure (b) there is no effect on agreeableness. However, there is a spike the year after the camp that is significant. This effect is gone already the time-period after and presumably spurious. Figure (c) show that RCCD has a positive effect on emotional stability in the time-periods after the intervention, however, in the following time-periods the effect fades away. This is particularly interesting since the kink on the curve occurs in the time-period where the pupils are on their own and no longer in contact with the RCCD consultants.

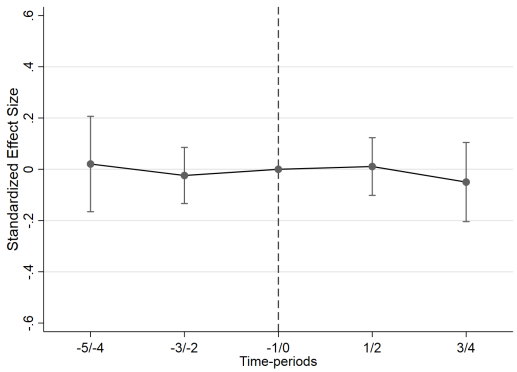
Next, in figure 3 I show how the effect of RCCD evolves in the time-periods following treatment as well as discussing whether the RCCD and NOTA pupils would have followed the same trend in the absent of treatment by investigating their pre-treatment behavior. It is evident from figure 3 that the pre-RCCD trends are similar to the previous two figures. Thus, the RCCD pupils have statistically the same behaviors as the NOTA pupils in the two time-periods leading up to treatment across all well-being scales. Similar to conscientiousness, I observe

Figure 1: Event study on Reading performance

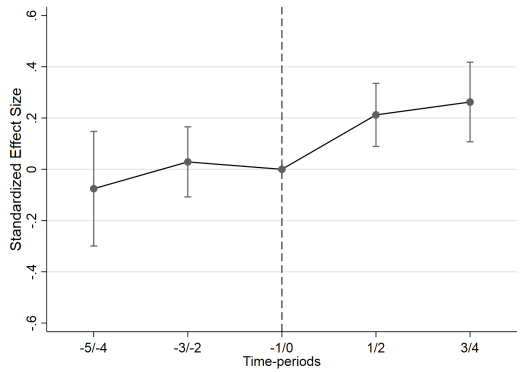
(a) Language Comprehension



(b) Decoding



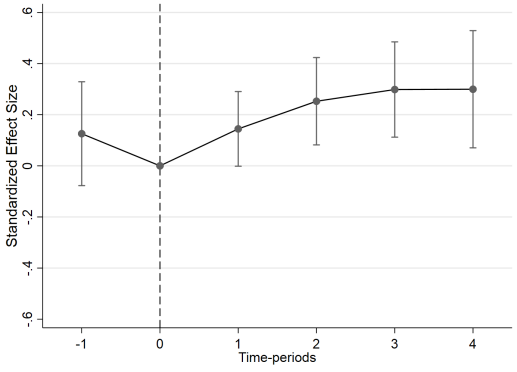
(c) Text Comprehension



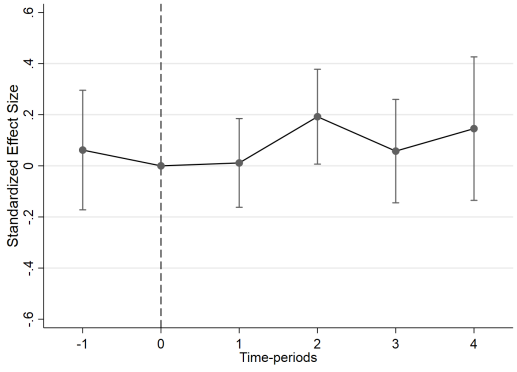
Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. Time-period, t=1, is the school year of (synthetic) RCCD participation. The vertical lines show the 95% confidence interval.

Figure 2: Event study on Personality traits

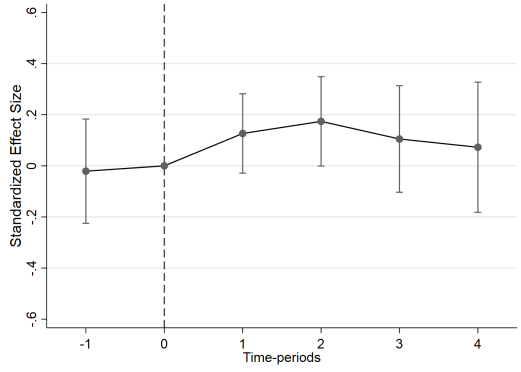
(a) Conscientiousness



(b) Agreeableness

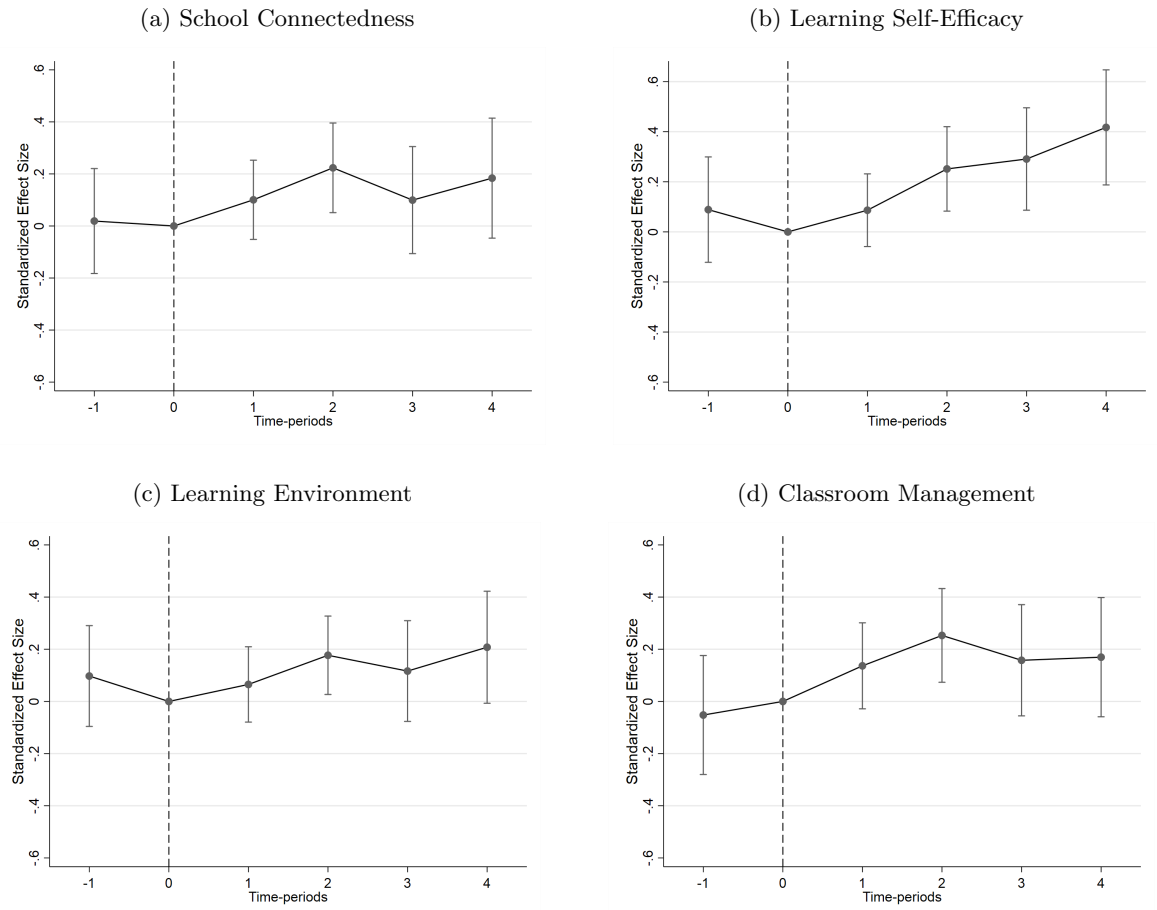


(c) Emotional Stability



Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. Time-period, t=1, is the school year of (synthetic) RCCD participation. The vertical lines show the 95% confidence interval.

Figure 3: Event study on School Well-being



Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. Time-period, t=1, is the school year of (synthetic) RCCD participation. The vertical lines show the 95% confidence interval.

a slight decrease in the RCCD pupils learning self-efficacy and learning environment scales in the time up to the intervention start. In the upper left corner I plot the effects on school connectedness. RCCD has a positive effect on the pupil's own beliefs regarding their teachers and peers' care for them. The effect reaches its highest level the year after the camp and shows signs of stabilizing with an increase of around 20% of a standard standard derivation. The upper right corner shows the effect on the pupil's beliefs and attitude toward academic achievements. The slope of the curve is particularly interesting since it keeps increasing throughout all post-RCCD time-periods. Thus, RCCD significantly improves learning self-efficacy, and the effect appears to be self-reinforcing up to 40% of a standard derivation. In the lower left corner I show the effect on RCCD on the pupil's attitude to the learning environment such as motivation and support of the teachers. The graph is similar to the school connectedness i.e. there is a borderline significant effect the year after RCCD, and this effect seems to remain. However, there is a larger, though insignificant pre-trend, and therefore, this effect must be interpreted carefully. Finally, in the lower right corner I show that RCCD changes the pupil's

experience of the classroom. Again, I observe that the effect peaks the year after RCCD and decreases slightly and becomes borderline significant when the pupil is returned to the local learning environment and is no longer supervised by the RCCD consultant. However, the post treatment effects are insignificantly different from each other.

5.2 Simple Difference-in-Difference

For the sake of clarity and simplicity in presentation, I supplement the main event study model with a minimalist difference-in-difference analysis that do not allow treatment effects to vary with time to RCCD participation. I also conduct robustness test and heterogeneity analyses using the simpler model.

Table 3-5 show that the average treatment effect on the treated (ATT henceforth) on reading abilities, personality traits and school well-being are robust to using a gradually richer specification. The ATT should be interpreted as effect sizes since all outcomes are standardized with mean zero and standard deviation of one for the full population. Column 1-3 show the results from the fixed effect regression where column 4-6 implements the Entropy Balancing strategy and reports the weighted fixed effect regression results. Column 1 and 4 include individual and time-period fixed effects, column 2 and 5 add grade and school year fixed effect whereas column 3 and 6 include time-varying covariates[4] in the regressions. Generally, the ATT estimates are robust to the inclusion of covariates, i.e. the variation in the ATT estimates in column 1 to 3 and 4 to 6 is small and statically insignificant. My preferred specification is the weighted fixed effect regressions, displayed in column 5, without the time-varying covariates because RCCD participation may affect the covariates. Thus, I exploit this specification for the remaining sections of this paper. This is not critical; including covariates in the main model yields similar findings.

Table 3 presents the effects of RCCD on reading abilities. Panel A shows the results on language comprehension, panel B shows results on decoding, and panel C the results on text comprehension. RCCD has significantly positive effects on language comprehension and text comprehension both with and without reweighing and across all covariate specifications. The impact on decoding is insignificant and very close to zero, i.e. RCCD does not affect the pupils decoding abilities. Combining fixed effect and entropy balancing decreases the effect sizes compared to the simple fixed effect framework by 0.06 SD for text comprehension and 0.04 SD for language comprehension. Thus, pupils enrolled in RCCD improves their language comprehension by 0.23 SD and text comprehension by 0.22 SD.

In table 4, I display the ATT on personality traits constructed from the national well-being survey, which was introduced in the school year 2014/2015 and are, therefore, not available for the first RCCD cohorts. Additionally, only pupils treated in grade 5 or older affect the

[4]The following covariates are included in the model: school movements, retaken grade, received special needs teaching, timing of ADHD, OCD and anxiety diagnosis, whether the pupil is living with both parents, parents employment status, SES and marital status.

Table 3: Main effects on Reading abilities

	Fixed Effect (FE)			FE & Entropy Balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Language Comprehension						
ATT	**0.276**	**0.269**	**0.269**	**0.230**	**0.229**	**0.229**
	(0.061)	(0.060)	(0.060)	(0.061)	(0.060)	(0.060)
R-squared	0.004	0.023	0.023	0.020	0.031	0.031
Number of observations	134,171	134,171	134,171	134,171	134,171	134,171
Number of individuals	45,901	45,901	45,901	45,901	45,901	45,901
Panel B: Decoding						
ATT	0.037	0.034	0.034	0.003	0.001	0.001
	(0.051)	(0.051)	(0.051)	(0.051)	(0.050)	(0.050)
R-squared	0.001	0.013	0.013	0.002	0.018	0.018
Number of observations	134,171	134,171	134,171	134,171	134,171	134,171
Number of individuals	45,901	45,901	45,901	45,901	45,901	45,901
Panel C: Text Comprehension						
ATT	**0.278**	**0.277**	**0.277**	**0.219**	**0.221**	**0.221**
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
R-squared	0.010	0.013	0.013	0.064	0.071	0.071
Number of observations	134,171	134,171	134,171	134,171	134,171	134,171
Number of individuals	45,901	45,901	45,901	45,901	45,901	45,901
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	No	Yes	Yes	No	Yes	Yes
School year, FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes

Notes: The table presents the results from separate fixed effect models comparing treated pupils to untreated. Bold (italic) indicates significance at the 5% (10%) level. The covariates are imputed with the value zero and a missing indicator equal to one is added to the conditioning set if data on the covariates are missing. Standard errors in parentheses are clusted at the individual level

Table 4: Main effects on Personality traits

	Fixed Effect (FE)			FE & Entropy Balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Conscientiousness						
ATT	**0.182**	**0.177**	**0.183**	**0.164**	**0.164**	**0.173**
	(0.065)	(0.064)	(0.064)	(0.064)	(0.064)	(0.064)
R-squared	0.002	0.005	0.008	0.014	0.020	0.022
Number of observations	95,991	95,991	95,991	95,991	95,991	95,991
Number of individuals	37,085	37,085	37,085	37,085	37,085	37,085
Panel B: Agreeableness						
ATT	0.067	0.070	0.069	0.054	0.060	0.064
	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.078)
R-squared	0.001	0.002	0.002	0.005	0.011	0.013
Number of observations	96,881	96,881	96,881	96,881	96,881	96,881
Number of individuals	37,106	37,106	37,106	37,106	37,106	37,106
Panel C: Emotional Stability						
ATT	**0.152**	**0.150**	**0.157**	**0.143**	**0.144**	**0.152**
	(0.066)	(0.066)	(0.066)	(0.066)	(0.067)	(0.067)
R-squared	0.000	0.002	0.005	0.006	0.008	0.012
Number of observations	92,945	92,945	92,945	92,945	92,945	92,945
Number of individuals	36,741	36,741	36,741	36,741	36,741	36,741
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	No	Yes	Yes	No	Yes	Yes
Schoolyear, FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes

Notes: The table presents the results from separate fixed effect models comparing treated pupils to untreated. Bold (italic) indicates significance at the 5% (10%) level. The covariates are imputed with the value zero and a missing indicator equal to one is added to the conditioning set if data on the covariates are missing. Standard errors in parentheses are clusted at the individual level

ATT estimates since grade 4 pupils do not have a pre-RCCD personality trait measure. Panel A presents the ATT on conscientiousness, panel B on agreeableness and panel C on emotional stability. Overall, the estimates are positive and consistent across all specifications. RCCD participation has a significantly positive effect of 0.16 SD on conscientiousness and 0.14 SD on emotional stability. For agreeableness, the estimates are positive but statistically insignificant.

The effects on school well-being are presented in table 5. I measure school well-being through four scales, and they are - similar to the three personality trait scales - constructed from the national well-being survey and thus, not available for the first RCCD cohorts and the treatment effects are based on grade 5 or older RCCD pupils. Panel A shows the effects of RCCD on school connectedness, panel B the effects on learning self-efficacy, panel C the effects on the learning environment, and panel D the effects on classroom management. Similar to table 4, the estimates are overall positive, and there is limited fluctuation in the estimates across different specifications. RCCD has significant positive effects on three out of the four school well-being scales. The effect on the learning environment is positive though statistically insignificant. I find significant effect sizes of 0.14 SD on school connectedness, 0.15 SD learning self-efficacy, and 0.19 SD on classroom management.

5.3 Robustness

In this section, I present sensitivity checks to evaluate the robustness and validity of the findings. First, I use secondary outcomes to study the pre-RCCD differences. Secondly, I discuss the implication of alternative comparison groups. Thirdly, I show that the results are robust to alternative balancing specifications. Finally, I study 10 primary outcomes and investigate if some of them are significant due to statistical chance.

Secondary outcomes. In this section, I exploit absence data as secondary outcomes in order to increase the validity of the main findings. Absence data have the unique feature of being collected in grade 1 to 9, which allows me to investigate up to 4 time-periods prior to RCCD participation. Figure A.1 shows time-varying treatment effects from the four time-periods leading up to RCCD as well as four post-RCCD time-periods for the three types of absence registered by the schools. a) Sick absence, b) Illegal absence – absence without permission from the school principal, and c) Legal absence - absence with permission from the school principal. It is evident from the figure that RCCD and NOTA pupils have similar pre-RCCD behavior. This suggests that the crucial parallel trend assumption is satisfied. In the post periods, I observe a 1% decrease in sick absence and a 2.5% increase in legal absence in the year of the 10-weeks camp. The increase in legal absence is expected as the pupils are enrolled at the local school but receiving teaching at the RCCD location and therefore receive absence which is allowed by the school principal. In the following periods, there are no effects on sick and legal absence. For illegal absence, the curve is flat until period three where the RCCD pupils increases their illegal absence and in period four this effect is statistically significant.

Table 5: Main effects on School Well-being

	Fixed Effect (FE)			FE & Entropy Balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: School Connectedness						
ATT	**0.147**	**0.147**	**0.153**	**0.134**	**0.136**	**0.149**
	(0.065)	(0.066)	(0.065)	(0.065)	(0.066)	(0.065)
R-squared	0.004	0.006	0.010	0.005	0.006	0.013
Number of observations	89,680	89,680	89,680	89,680	89,680	89,680
Number of individuals	36,200	36,200	36,200	36,200	36,200	36,200
Panel B: Learning Self-Efficacy						
ATT	**0.163**	**0.165**	**0.169**	**0.140**	**0.145**	**0.155**
	(0.066)	(0.066)	(0.066)	(0.066)	(0.065)	(0.064)
R-squared	0.003	0.008	0.012	0.006	0.015	0.020
Number of observations	83,279	83,279	83,279	83,279	83,279	83,279
Number of individuals	35,274	35,274	35,274	35,274	35,274	35,274
Panel C: Learning Enviroment						
ATT	0.085	0.090	0.094	0.082	0.083	0.092
	(0.065)	(0.065)	(0.064)	(0.065)	(0.063)	(0.064)
R-squared	0.074	0.078	0.080	0.080	0.090	0.094
Number of observations	89,868	89,868	89,868	89,868	89,868	89,868
Number of individuals	36,147	36,147	36,147	36,147	36,147	36,147
Panel D: Classroom Management						
ATT	**0.195**	**0.196**	**0.200**	**0.197**	**0.192**	**0.198**
	(0.074)	(0.074)	(0.074)	(0.074)	(0.073)	(0.073)
R-squared	0.001	0.002	0.004	0.010	0.019	0.020
Number of observations	96,547	96,547	96,547	96,547	96,547	96,547
Number of individuals	37,130	37,130	37,130	37,130	37,130	37,130
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	No	Yes	Yes	No	Yes	Yes
Schoolyear, FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes

Notes: The table presents the results from separate fixed effect models comparing treated pupils to untreated. Bold (italic) indicates significance at the 5% (10%) level. The covariates are imputed with the value zero and a missing indicator equal to one is added to the conditioning set if data on the covariates are missing. Standard errors in parentheses are clusted at the individual level

This timing is particular interesting as the effect begins after the time-period where the RCCD consultant is no longer supervising the pupil.

Alternative Comparison Groups. The key identifying assumption for the difference-in-difference framework to yield causal estimates is the parallel trend assumption i.e. the RCCD group follows the same trend as the NOTA group in the absence of treatment. I exploit the event study analysis to show that at least in the period up to the intervention the trends are not statistically significantly different. Another approach would be to use alternative comparison groups and check the sensitivity of the results. In tables A.4-A.6, I report the weighted fixed effect regression for alternative comparison groups. Panel A shows the main results. Panel B exploits pupils diagnosed with dyslexia in the national dyslexia test implemented at the beginning of 2015. The downside is naturally that this reduces the number of observations in the model. A comparison with pupils diagnosed in the national dyslexia test is interesting because they are identified dyslexics by an objective validated test. In contrast, to a NOTA membership that at least prior to 2015 was establish by a subjective assessment. Thus, similar estimates indicate that NOTA pupils are dyslexics or at least behave in a similar way and are not a selected group of dyslexics. Panel C utilizes all pupils enrolled in the Danish public school system in the relevant grades and years. The weighting strategy selects those similar in observable characteristics to the RCCD pupils without them necessarily being dyslexic. Thus, panel C checks the sensitivity of the results by not selecting a pre-specified comparison group but instead makes use of the full population. The estimates in tables A.4-A.6 indicate that the main findings do not change. Comparing RCCD pupils to pupils tested dyslexics - instead of NOTA members - results in effects of similar sign and magnitude. Comparing them to all pupils in Danish public school results in effects of similar sign but larger magnitude. The effects based on all pupils as a comparison group are up to 0.1 SD larger than the effects with NOTA members as a comparison group. I observe the largest differences for the reading outcomes and conscientiousness. These findings strongly support the main results that RCCD has positive effects on areas within reading abilities, personality traits and school well-being. The larger effect from the full population estimates could be due to a different treatment as usual. Thus, treatment as usual could be more intensive for the dyslexic comparison group. I discuss the implication of treatment as usual in section 6.

Alternative Balancing Specifications. For the weighted fixed effect regressions, I perform sensitivity checks of the re-weighting strategy in tables A.7-A.9. First, Entropy Balancing re-weights the comparison observations to balance the first (mean), second (variance), and - possibly - third (skewness) moment of the covariate distribution. The first column of the tables presents the main specification whereas column 2 balances on the first and second moment (mean and variance) and column 3 balances on all three moments (mean, variance and skewness). Column 2 and 3 show that balancing on the second and third moments as

well leads to similar effect sizes as the main specification. This is not surprising since a large fraction of the covariates, which are included in the entropy balancing method, are binary. Secondly, I apply propensity score matching on NOTA members to construct the comparison group. The propensity score is estimated using a logistic regression on the same covariates as the entropy balance described in table A.3. The matching strategy is nearest neighbor with replacement, and I match each RCCD pupil with respectively 3, 5 and 10 NOTA pupils and report the effects in column 4-6. The reading abilities and personality traits estimates are relatively stable across specifications. Naturally, the standard error increases in the matching estimates given the lower number of observations. For the well-being scales, the estimates have the same sign and interpretation across specifications. However, the effect sizes are larger for the propensity score matching specification. Overall, the results of tables A.7-A.9 show that the main findings are robust to different balancing specifications.

Multiple Hypotheses Testing. Considering multiple outcomes increases the risk of rejecting a true null hypothesis, i.e. that some of the estimated effects might be significant due to statistical chance. In order to account for the problem of testing hypothesis for a large number of outcomes, I calculates Westfall-Young stepdown adjusted p-values described in (Westfall et al., 1993) and applied in Jones et al. (2019). The Westfall-Young approach controls the family-wise error rate (FWER), i.e. the probability of making one or more type I errors when conducting multiple hypotheses tests. The method uses bootstrapping to allow for dependence across outcomes. I perform the resampling over entire clusters rather than individual observations.

I restrict attention to results obtain with the difference-in-difference setup combined with entropy balancing. To control for the family-wise error rate, I define mutually exclusive families of hypothesis that include all outcome variables. Each family contains all variables belonging to one of three outcome domains (reading performance, personality traits or school well-being). Table A.10 shows p-values from my main model as well as Westfall-Young stepdown adjusted p-values based on 500 bootstraps. When adjusting for multiple hypothesis testing, the treatment effects remain statistically significant. Thus, these results strengthen the conclusion that RCCD positively affect reading abilities, personality traits, and school well-being.

5.4 Heterogeneous Treatment Effects

It is relevant to break down the sample into subgroups based on shared characteristics in order to identify differences in how pupils respond to RCCD. However, in light of the relatively few numbers of RCCD participants, especially for outcomes based on the national well-being survey, this study does not have sufficient power to draw inference on subgroup analyses with effect sizes similar to what I observe in tables 4 and 5. Nevertheless, it could be enlightening to report heterogeneous effects even though they are of more exploratory nature. Naturally, this power issue means that the heterogeneous effects should be carefully interpreted. The

heterogeneous treatment effects are presented in tables A.11-A.13 in the appendix.

Sex. Many evaluations of educational interventions show stronger effects for girls than boys, and in some cases the full effect is driven by the girls so it is natural to ask whether this is also the case for RCCD. For example, Schwartz et al. (2021) finds that the effect of special education for pupils with learning disabilities are 0.06SD higher for girls. When interpreting the effect on reading abilities, I observe almost no difference across gender on text comprehension. However, for language comprehension the effect is 0.10SD larger for boys. When I observe personality traits it is clear that the main effect on conscientiousness is driven by the girls (effect size of 0.25). However, for emotional stability, I observe larger effect for the boys, but the difference is not as large as for conscientiousness. It is important to note that the estimate becomes less precise since the standard errors increase as the number of observations decrease. There is little gender effect on school well-being, and the small difference is likely a result of the lower precision of the estimates.

Socioeconomic status. Similar to the gender argument, many large-scale educational interventions have larger effects for pupils with high socioeconomic background and turn out to be ineffective for pupils of low socioeconomic background. I exploit the mothers' educational level as an indicator for the pupil's socioeconomic background. I classify pupils of mothers who have at most completed high school as having low socioeconomic background (i.e. mothers have 12 years of schooling or less). In reading abilities I observe large effects for pupils from low socioeconomic background (effect sizes of 0.25, 0.13 and 0.33). Especially, for text comprehension, the effect is one tenth of a standard deviation larger than the overall effect. The low number of individuals is because of few pupils having mothers with low educational levels combined with the short time-horizon for personality traits and school well-being makes it difficult to interpret the estimates. Interestingly, pupils with low socioeconomic backgrounds do not increase their school connectedness. They, however, more than double the effect on classroom management to 0.40.

Low decoding abilities. Dyslexia is negatively correlated with decoding abilities. Therefore, I use the lowest decoding performers in the national reading test prior to RCCD as a proxy to identify those who are most affected by dyslexia. This enables me to investigate how effective RCCD is for this particular group of pupils. I define low decoding performers as pupils who are at least one standard deviation below the population average in decoding at the national test in the period up to RCCD. When examining the effects on future reading abilities, the effect is generally below the main effect but statistically indifferent. In terms of personality traits and school well-being, the low performers have generally larger effects. However, the magnitudes of the differences are limited.

6 Discussion

The results point towards sizeable, significant and persistent effects of the specialized dyslexia intervention on both reading abilities, personality traits and school well-being. To delve deeper into to findings, I first discuss the economic significance in relation to standardized effect sizes and other Danish education studies. Secondly, I discuss caveats associated with my choice of comparison groups.

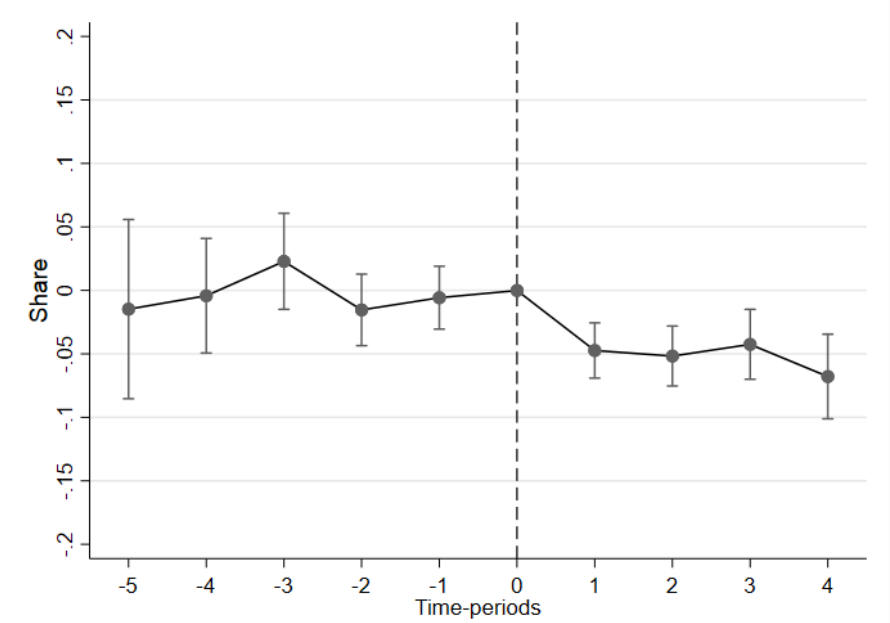
6.1 Economic significance

From a policy perspective, it is important not only to evaluate whether findings are statically significant but more relevant to discuss if it is also economically significant, i.e. is the estimated effect large enough for policymakers to respond to it. Generally, I find effect sizes between 0.15 and 0.25 with largest effects on the reading ability areas language comprehension and text comprehension. While these effects are small, using the Cohen's d guideline, they are larger when benchmarked against educational interventions conducted in social science (Kraft, 2020) and larger than effect sizes found in other Danish primary school interventions (Andersen et al., 2016, 2020; Rosholm et al., 2021). Rosholm et al. (2021) show standardized effect sizes from 18 treatment arms of 10 Danish RCT interventions aimed at children and adolescents. The RCCD intervention is among the most effective, especially, if compared to other school interventions but also when compared to pre-school interventions.

Kraft (2020) argues that the cohen's d standards are too large relatively to impacts of most field-based intervention because it is based on a few tightly controlled lab experiments in social psychology. The author instead suggests new benchmarks for effect sizes, which originates from 750 RCTs within social science and emphasizes the importance of cost and scalabilities. Kraft (2020) concludes from the distribution of effect sizes that effects of 0.2 or greater must be considered as large. This suggests that the effects from my study are medium, if not large. Especially the effects on reading performances are large. They are around the 90 percentile of the distribution of reading effect sizes included in Kraft (2020) when accounting for the sample size and the age of the pupils. Kraft (2020) argues that from a policy perspective the magnitude of effect sizes is not sufficient because effect sizes do not reflect the cost of the intervention or how likely it is to scale with fidelity. The municipality of Aarhus estimates the total cost of the 18 months intervention to be $18,000$USD per pupil, which Kraft (2020) defines as a high cost intervention. Finally, I discuss the scalability of RCCD – i.e. I ask whether the effects will be similar if RCCD is provided to a large population of pupils. RCCD is easy to scale because the limiting factor is the number of consultants. Thus, hiring more consultants will enable the municipality to provide RCCD to an increased number of dyslexics with limited influence to the effects. In fact, increasing the number of pupils might lead to economies of scale advantages, which will decrease the cost per pupil.

Despite the high costs, investments in dyslexic learning program remain policy relevant

Figure 4: Event study on special needs teaching



Notes: The figure presents time-varying treatment effects on use of special needs teaching from the fixed effect model combined with entropy balancing. Pupils are, due to data constraint, only registered as special needs pupils if the receive minimum 9 hours of special needs teaching per week. Time-period, t=1, is the school year of (synthetic) RCCD participation. The vertical lines show the 95% confidence interval.

because of the large effects for this extremely disadvantaged group. Figure 4 present the effect of RCCD on special needs teaching using a event study representation. Note, only pupils with a minimum of 9 hours special needs teaching is registered as special needs pupils in Denmark. The pre-RCCD effects are constant, consistent with parallel trends between the RCCD and NOTA groups before the intervention event. Overall, I observe a 5% decrease in the use of special needs teaching. Thus, with 5% of the pupils registered as special needs pupils in the school year before RCCD participation, RCCD enables dyslexics to participate in normal classroom teaching without additional support. Importantly, RCCD appear to be cost-effective, also in the short-run, where the yearly cost of general special needs teaching is 29,500USD for 9 hours per week. Dyslexia interventions as RCCD become even more cost-effective when considering the long-run perspective in which special needs pupils have greater cost to society by reduced tax revenue, higher health costs, and higher crime risk.

6.2 Treatment as usual

Comparison pupils receive most likely additional educational support. In 2017, a new school law ensures pupils diagnosed with dyslexic extra help from their local school. Aid is often compensatory assistive reading and writing tools, which is also part of RCCD and NOTA. Thus, my findings are the effects on top of the basic dyslexia help provided in the Danish

primary schools. One might worry that NOTA members seek additional help from other sources. This might be a) attending other dyslexic learning programs, b) receive additional hours of special needs. I consider this as a minor issue because it will be visible in increased legal absence for the NOTA pupils and in figure A.1 I observe that the two groups have similar amount of legal absence in all other time-periods than the camp period.

There exists many opportunities for attending dyslexic learning programs across the Danish municipalities as discussed in section 2. I handle participation in dyslexia interventions and other initiatives targeting this group of pupils as *treatment as usual*. Thus, the high level of already existing supports for dyslexics highlights the effectiveness of RCCD.

In tables A.4-A.6 I observe, when changing from NOTA members to all pupils - including potentially non-dyslexic - in the comparison group. This leads to increased effect sizes on all outcomes. Indicating that dyslexics in Denmark generally receive some kind of intervention that helps their situation and, thus, decreases the effects of RCCD relative to a situation without addition help. Therefore, it is important to keep in mind that the appropriate interpretation of my findings is the effect of RCCD on top of *treatment as usual*.

7 Conclusion

This paper evaluates a specialized learning program targeted dyslexics in grades 4 to 8 on future academic success, personality traits, and school well-being. Using a difference-in-difference setup in combination with entropy balancing, I compare outcomes, over time, among those who participate in the RCCD program against other dyslexics who do not participate in the program.

Using entropy balance with a large set of observed characteristics works as robustness to the parallel trend assumption since weighting on observed characteristics prior to the intervention removes differences between the RCCD and NOTA groups. Combining both methods and performing a set of sensitivity tests, leads me to argue that it is highly unlikely that unobserved time-varying confounders or different trends causes these findings.

My findings reveal that RCCD is important for future reading abilities as well as personality traits and school well-being. RCCD participants increase their language and text comprehension by respectively 25% and 22% of a standard derivation. They, thus, perform well beyond other dyslexics and have closed the gap to non-dyslexics with up to 33%. In addition, RCCD participants also show a significant increase in two out of three measured personality traits and three out of four school well-being measures with effect sizes between 14% and 19% of an standard derivation. Closing the gap to non-dyslexics with up to 80%. The effects on language and text comprehension are persistent across all time-periods I observe and the effects on conscientiousness and learning self-efficacy increases each school year throughout the time-periods, I observe. The results are robust to alternative comparison groups, different entropy balancing moments and alternative balancing strategies.

Overall, my findings are important for policymakers because they indicate that investing in expensive dyslexia learning camp leads to large effects on academic performance, personality traits, and well-being. Most importantly, however, this is highly cost-effective when considering the large cost of special needs teaching as well as the huge cost to society in the long-run for not helping this disadvantage group of children.

References

Abadie, A. (2005): "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 72, 1–19.

Alan, S., T. Boneva, and S. Ertac (2019): "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit," *The Quarterly Journal of Economics*, 134, 1121–1162, publisher: Oxford University Press.

Alan, S. and S. Ertac (2018): "Fostering patience in the classroom: Results from randomized educational intervention," *Journal of Political Economy*, 126, 1865–1911, publisher: University of Chicago Press Chicago, IL.

Andersen, S. C., M. Gensowski, S. G. Ludeke, and O. P. John (2020): "A stable relationship between personality and academic performance from childhood through adolescence. An original study and replication in hundred-thousand-person samples," *Journal of Personality*, 88, 925–939.

Andersen, S. C., M. K. Humlum, and A. B. Nandrup (2016): "Increasing instruction time in school does increase learning," *Proceedings of the National Academy of Sciences*, 113, 7481–7484.

Beuchert, L. and A. Nandrup (2017): "The Danish National Tests at a Glance," *SSRN Electronic Journal*.

Blundell, R., M. C. Dias, C. Meghir, and J. van Reenen (2004): "Evaluating the Employment Impact of a Mandatory Job Search Program," *Journal of the European Economic Association*, 2, 569–606.

Dietrichson, J., T. Filges, R. H. Klokker, B. C. V. Nielsen, M. Bøg, and U. H. Jensen (2020a): "Targeted school-based interventions for improving reading and mathematics for students with or at-risk of academic difficulties in grade 7 to 12: A systematic review," *Campbell Systematic Reviews*, 16, e1081.

Dietrichson, J., T. Filges, J. K. Seerup, R. H. Klokker, B. A. Viinholt, M. Bøg, and M. Eiberg (2020b): "Targeted school-based interventions for improving reading and mathematics for students with or at risk of academic difficulties in grade k to 6: A systematic review," *Campbell Systematic Reviews*, 11, 1–60.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007): "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology*, 92, 1087, vol. 92, No. 6, 1087–1101.

Einar, H., L. Anders, S. Jim, and H. Kenneth (2001): "Behaviour problems in children with dyslexia," *Nordic Journal of Psychiatry*, 55, 251–256.

Epnion (2018): *"Survey og registeranalyse – børn og unge med ordblindhed"*, available at https://www.egmontfonden.dk/sites/default/files/2019-10/Egmont%20Rapporten%202018%20%27Let%20vejen%20-%20til%20uddannelse%20for%20ordblinde%20b%C3%B8rn%20og%20unge%27.pdf.

Hainmueller, J. (2012): "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 20, 25–46.

Hainmueller, J. and Y. Xu (2013): "ebalance: A Stata Package for Entropy Balancing," *Journal of Statistical Software*, 54, 1–18.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2002): "Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics*, 84, 584–599.

Hassler, B., L. Major, and S. Hennessy (2016): "Tablet use in schools: a critical review of the evidence for learning outcomes," *Journal of Computer Assisted Learning*, 32, 139–156.

Hebert, M., D. M. Kearns, J. B. Hayes, P. Bazis, and S. Cooper (2018): "Why Children With Dyslexia Struggle With Writing and How to Help Them," *Language, Speech, and Hearing Services in Schools*, 49, 843–863.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654.

Hurwitz, S., B. Perry, E. D. Cohen, and R. Skiba (2020): "Special Education and Individualized Academic Growth: A Longitudinal Assessment of Outcomes for Students With Disabilities," *American Educational Research Journal*, 57, 576–611, publisher: American Educational Research Association.

Jones, D., D. Molitor, and J. Reif (2019): "What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study*," *The Quarterly Journal of Economics*, 134, 1747–1791.

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

Kraft, M. A. (2020): "Interpreting Effect Sizes of Education Interventions," *Educational Researcher*, 49, 241–253.

Lindeblad, E., S. Nilsson, S. Gustafson, and I. Svensson (2017): "Assistive technology as reading interventions for children with reading impairments with a one-year follow-up," *Disability and Rehabilitation. Assistive Technology*, 12, 713–724.

Morgan, P. L., M. Frisco, G. Farkas, and J. Hibel (2010): "A Propensity Score Matching Analysis of the Effects of Special Education Services," *The Journal of special education*, 43, 236–254.

Niclasen, J., M. Keilow, and C. Obel (2018): "Psychometric properties of the Danish student well-being questionnaire assessed in >250,000 student responders," *Scandinavian Journal of Public Health*, 46, 877–885.

NOTA (2019): *"Ordblindeindsatser i Danmarks kommuner"*, available at https://nota.dk/sites/default/files/ordblindeindsatser_i_danmarks_kommuner_nota_september_2019.pdf.

Nørgaard, E., M. Schøler Kollin, and B. Panduro (2018): "Specialundervisning på folkeskoleområdet: inspiration til den økonomiske styring," Tech. rep., VIVE - Viden til Velfærd, available at https://www.vive.dk/media/pure/7029/2300521.

Perelmutter, B., K. K. McGregor, and K. R. Gordon (2017): "Assistive technology interventions for adolescents and adults with learning disabilities: An evidence-based systematic review and meta-analysis," *Computers & Education*, 114, 139–163.

Reynolds, A. J. and B. Wolfe (1999): "Special Education and School Achievement: An Exploratory Analysis With a Central-City Sample," *Educational Evaluation and Policy Analysis*, 21.

Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.

Rosholm, M., A. Paul, D. Bleses, A. Højen, P. S. Dale, P. Jensen, L. M. Justice, M. Svarer, and S. Calmar Andersen (2021): "Are Impacts of Early Interventions in the Scandinavian Welfare State Consistent with a Heckman Curve? A Meta-Analysus," *Journal of Economic Surveys*, 35, 106–140.

Ryan, A. M., E. Kontopantelis, A. Linden, and J. F. Burgess (2018): "Now trending: Coping with non-parallel trends in difference-in-differences analysis," *Statistical Methods in Medical Research*.

SBV (2014): *Dyslexi hos barn och ungdomar: tester och insatser : ensystematisk litteraturöversikt : sammanfattning och slutsatser*, Stockholm: SBU.

Schwartz, A. E., B. G. Hopkins, and L. Stiefel (2021): "The Effects of Special Education on the Academic Performance of Students with Learning Disabilities," *Journal of Policy Analysis and Management*, 41.

Snowling, M. J. (2013): "Early identification and interventions for dyslexia: a contemporary view: Early identification and interventions for dyslexia: a contemporary view," *Journal of Research in Special Educational Needs*, 13, 7–14.

Snowling, M. J., C. Hulme, and K. Nation (2020): "Defining and understanding dyslexia: past, present and future," *Oxford Review of Education*, 46.

Svendsen, H. B. (2017): "Et didaktisk spændingsfelt," *Learning Tech*, 110–136.

Undheim, A. M. (2009): "A thirteen-year follow-up study of young Norwegian adults with dyslexia in childhood: reading development and educational levels," *Dyslexia*, 15.

Westfall, P. H., S. S. Young, and S. P. Wright (1993): "On Adjusting P-Values for Multiplicity," *Biometrics*, 49, 941.

Table A.1: Well-being survey response rate by group

	2014/2015	**2015/2016**	**2016/2017**	**2017/2018**	**2018/2019**	**Total**
RCCD	71.4	84.1	83.8	67.9	69.5	76.0
NOTA	78.2	85.7	84.2	79.3	72.5	80.6
Non-NOTA	80.7	88.3	87.4	82.4	76.8	83.5

Notes: This table shows the response rate at the national well-being survey RCCD participants against NOTA and non-NOTA members.

Table A.2: Items used in personality measures

Conscientiousness
- How often can you complete what you set out to do?
- Can you concentrate during lessons?
- If interrupted during lessons, I can quickly concentrate again.
Agreeableness
- I try to understand my friends' feelings when they are sad or upset
- I am good at collaborating with others.
Emotional Stability
- Do you feel lonely?
- Other students accept me as I am.
- How often do you feel secure at school?
School connectedness
- Do you like your school?
- Do you like your class?
- Do you feel lonely?
- I feel that I belong at this school
- Most of the students in my class are kind and helpful
- Other students accept me as I am
- How often do you feel safe at school?
Learning self-efficacy
- What do your teachers think of your progress in school?
- Do you succeed in learning what you want in school?
- How often can you find a solution to problems, if you try hard enough?
- How often can you manage the things you set your mind to?
- Can you concentrate during lessons?
- I do well in school, academically
- If interrupted during lessons, I can quickly concentrate again.
- If something is difficult for me during class, I can do something about it myself to move on
Learning environment
- Do your teachers help you learn in ways that work?
- Lessons make me want to learn more
- The teachers are good at supporting and helping me at school when I need it
- Do you and your classmates have a say in what the class works on?
- The teachers ensure that the students' ideas are used in class
- Are the lessons exciting?
- I like the surroundings outside my school
- I like the classrooms at my school
Classroom management
- If there is noise in the classroom, teachers can quickly establish quietness
- Do your teachers show up for classes on time?
- Is it easy to hear what the teachers say during lessons?
- Is it easy to hear what the other students say during lessons?

Table A.3: Entropy Balancing Statistics

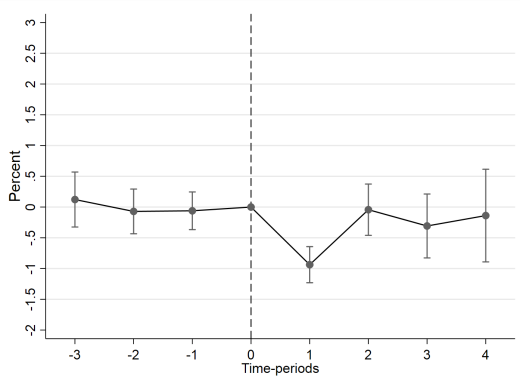
	RCCD		NOTA					
			Before balancing			After balancing		
	Mean	SD	Mean	SD	Diff.	Mean	SD	Diff.
Demographics								
- Boy (1/0)	0.600	(0.490)	0.561	(0.496)	*0.040*	0.600	(0.490)	0.000
- 1st or 2nd generation immigrant (1/0)	0.103	(0.305)	0.060	(0.237)	**0.044**	0.103	(0.304)	0.000
- Preventive personalized interventions (no.)	0.035	(0.239)	0.053	(0.293)	-0.017	0.035	(0.237)	0.000
- Siblings (no.)	1.380	(0.979)	1.328	(0.928)	0.052	1.380	(0.970)	0.000
- Living with both parents (1/0)	0.661	(0.474)	0.647	(0.478)	0.014	0.661	(0.473)	0.000
- School changes (no.)	0.232	(0.537)	0.431	(0.695)	**-0.199**	0.232	(0.494)	0.000
- Class changes (no.)	0.055	(0.236)	0.092	(0.311)	**-0.038**	0.055	(0.238)	0.000
- Relocations (no.)	1.175	(1.382)	1.735	(1.575)	**-0.560**	1.176	(1.133)	-0.001
- Special needs teaching (1/0)	0.056	(0.229)	0.106	(0.308)	**-0.051**	0.056	(0.229)	0.000
- Age (years)	12.070	(1.356)	12.163	(1.473)	-0.093	12.070	(1.354)	0.000
- Mother's age at birth (years)	30.907	(4.864)	29.715	(4.870)	**1.193**	30.906	(4.808)	0.001
- Father's age at birth (years)	33.254	(5.320)	32.379	(5.615)	**0.875**	33.254	(5.454)	0.000
Mother's marital status								
- Cohabiting couple (1/0)	0.107	(0.310)	0.098	(0.297)	0.010	0.107	(0.309)	0.000
- Divorced (1/0)	0.181	(0.386)	0.161	(0.367)	0.021	0.181	(0.385)	0.000
- Married (1/0)	0.620	(0.486)	0.658	(0.474)	*-0.038*	0.620	(0.485)	0.000
- Single (1/0)	0.080	(0.271)	0.072	(0.259)	0.008	0.080	(0.271)	0.000
Father's marital status								
- Cohabiting couple (1/0)	0.105	(0.307)	0.095	(0.293)	0.010	0.105	(0.307)	0.000
- Divorced (1/0)	0.136	(0.344)	0.152	(0.359)	-0.015	0.136	(0.343)	0.000
- Married (1/0)	0.643	(0.480)	0.651	(0.477)	-0.008	0.643	(0.479)	0.000
- Single (1/0)	0.078	(0.268)	0.066	(0.248)	0.012	0.078	(0.268)	0.000
Mother's highest educational degree								
- No degree or primary school (1/0)	0.170	(0.376)	0.216	(0.412)	**-0.047**	0.170	(0.375)	0.000
- High School (1/0)	0.060	(0.239)	0.045	(0.206)	*0.016*	0.060	(0.238)	0.000
- Vocational traning (1/0)	0.343	(0.475)	0.443	(0.497)	**-0.100**	0.343	(0.475)	0.000
- Academy higher education (1/0)	0.060	(0.239)	0.047	(0.211)	0.014	0.060	(0.238)	0.000
- College (1/0)	0.240	(0.427)	0.196	(0.397)	**0.044**	0.240	(0.427)	0.000
- University (1/0)	0.127	(0.333)	0.053	(0.225)	**0.073**	0.127	(0.333)	0.000
Father's highest educational degree								
- No degree or primary school (1/0)	0.226	(0.419)	0.278	(0.448)	**-0.052**	0.226	(0.418)	0.000
- High School (1/0)	0.047	(0.211)	0.029	(0.168)	**0.018**	0.047	(0.211)	0.000
- Vocational traning (1/0)	0.378	(0.485)	0.497	(0.500)	**-0.119**	0.378	(0.485)	0.000
- Academy higher education (1/0)	0.109	(0.312)	0.063	(0.244)	**0.046**	0.109	(0.312)	0.000
- College (1/0)	0.127	(0.333)	0.077	(0.266)	**0.050**	0.127	(0.333)	0.000
- University (1/0)	0.113	(0.317)	0.056	(0.230)	**0.057**	0.113	(0.317)	0.000
Birth cohort								
- 1996	0.014	(0.116)	0.016	(0.125)	-0.002	0.014	(0.116)	0.000
- 1997	0.033	(0.179)	0.035	(0.184)	-0.002	0.033	(0.179)	0.000
- 1998	0.066	(0.249)	0.059	(0.236)	0.007	0.066	(0.249)	0.000
- 1999	0.082	(0.274)	0.086	(0.280)	-0.004	0.082	(0.274)	0.000
- 2000	0.088	(0.283)	0.117	(0.322)	**-0.030**	0.088	(0.283)	0.000
- 2001	0.117	(0.322)	0.128	(0.334)	-0.011	0.117	(0.321)	0.000

- 2002	0.152	(0.359)	0.120	(0.325)	**0.032**	0.152	(0.359)	0.000
- 2003	0.117	(0.322)	0.121	(0.326)	-0.004	0.117	(0.321)	0.000
- 2004	0.131	(0.337)	0.118	(0.323)	0.012	0.131	(0.337)	0.000
- 2005	0.088	(0.283)	0.092	(0.289)	-0.004	0.088	(0.283)	0.000
- 2006	0.066	(0.249)	0.062	(0.241)	0.004	0.066	(0.249)	0.000
- 2007	0.033	(0.179)	0.032	(0.177)	0.001	0.033	(0.179)	0.000
- 2008	0.014	(0.116)	0.013	(0.115)	0.000	0.014	(0.116)	0.000
Grade when treated								
- Grade 4	0.168	(0.374)	0.211	(0.408)	**-0.043**	0.168	(0.374)	0.000
- Grade 5	0.269	(0.444)	0.213	(0.409)	**0.056**	0.269	(0.443)	0.000
- Grade 6	0.234	(0.424)	0.213	(0.409)	0.021	0.234	(0.423)	0.000
- Grade 7	0.205	(0.404)	0.197	(0.398)	0.007	0.205	(0.403)	0.000
- Grade 8	0.125	(0.331)	0.166	(0.372)	**-0.041**	0.125	(0.330)	0.000
Schoolyear when treated								
- 2010/2011	0.080	(0.271)	0.082	(0.274)	-0.002	0.080	(0.271)	0.000
- 2011/2012	0.113	(0.317)	0.098	(0.297)	0.015	0.113	(0.317)	0.000
- 2012/2013	0.121	(0.326)	0.110	(0.313)	0.011	0.121	(0.326)	0.000
- 2013/2014	0.096	(0.294)	0.121	(0.326)	*-0.025*	0.096	(0.294)	0.000
- 2014/2015	0.121	(0.326)	0.125	(0.331)	-0.004	0.121	(0.326)	0.000
- 2015/2016	0.125	(0.331)	0.125	(0.331)	-0.001	0.125	(0.330)	0.000
- 2016/2017	0.107	(0.310)	0.122	(0.327)	-0.014	0.107	(0.309)	0.000
- 2017/2018	0.121	(0.326)	0.113	(0.316)	0.008	0.121	(0.326)	0.000
- 2018/2019	0.117	(0.322)	0.104	(0.306)	0.013	0.117	(0.321)	0.000
National test difference up to treatment								
- Language Comprehension (std.)	-0.014	(1.344)	-0.041	(1.312)	0.028	-0.014	(1.288)	0.000
- Decoding (std.)	-0.030	(0.966)	-0.024	(0.907)	-0.006	-0.030	(0.901)	0.000
- Text Comprehension (std.)	0.362	(1.226)	0.096	(1.094)	**0.266**	0.362	(1.110)	0.000
Well-being difference up to treatment								
- School Connectedness (std.)	0.055	(0.997)	0.006	(1.022)	0.049	0.055	(0.984)	0.000
- Learning Self-Efficacy (std.)	-0.045	(0.949)	-0.025	(0.996)	-0.020	-0.045	(0.964)	0.000
- Learning Enviroment (std.)	-0.336	(0.893)	-0.194	(1.020)	-0.142	-0.336	(0.979)	0.000
- Classroom Management (std.)	-0.091	(1.189)	0.029	(1.174)	-0.121	-0.091	(1.174)	0.000
- Conscientiousness (std.)	-0.043	(1.007)	0.059	(1.079)	-0.102	-0.043	(1.048)	0.000
- Agreeableness (std.)	-0.030	(1.192)	0.012	(1.164)	-0.042	-0.030	(1.120)	0.000
- Emotional (std.)	0.074	(1.001)	0.049	(1.093)	0.025	0.074	(1.077)	0.000
Absence difference up to treatment								
- Sick (percent)	-0.015	(2.753)	0.173	(3.578)	-0.188	-0.015	(3.379)	0.000
- Illegal (percent)	0.091	(2.404)	0.193	(2.414)	-0.102	0.091	(2.414)	0.000
- Legal (percent)	-0.289	(2.704)	0.076	(2.995)	**-0.365**	-0.289	(3.117)	0.000
Number of individuals	513		45,839					

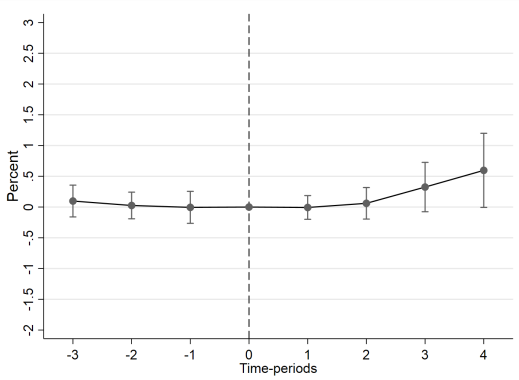
Notes: The table shows the descriptive statistics of the variables used in the balancing method. The mean, standard derivation and differences in means of the covariates are reported for the treatment group, for the comparison group before balancing, and for the comparison group after balancing. The comparison group is reweighted such that it mimics the treatment group. The table is based on non-missing data. Bold (italic) indicates significance at the 5% (10%) level.

Figure A.1: Event study on absence information

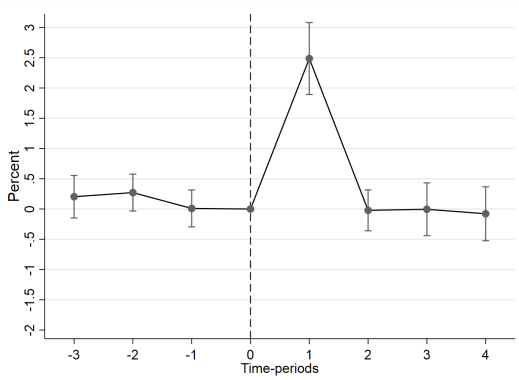
(a) Sick absence



(b) Illegal absence



(c) Legal absence



Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. The vertical lines show the 95% confidence interval.

Table A.4: Reading abilities - Alternative comparison groups

	Language Comprehension	Decoding	Text Comprehension
	(1)	(2)	(3)
Panel A: Member of NOTA			
ATT	**0.229**	0.001	**0.221**
	(0.060)	(0.050)	(0.054)
R-squared	0.031	0.018	0.071
Number of observations	134,171	134,171	134,171
Number of individuals	45,901	45,901	45,901
Panel B: Tested dyslexic			
ATT	**0.232**	0.020	**0.236**
	(0.062)	(0.052)	(0.054)
R-squared	0.031	0.018	0.072
Number of observations	97,419	97,419	97,419
Number of individuals	30,754	30,754	30,754
Panel C: All			
ATT	**0.290**	0.013	**0.329**
	(0.060)	(0.050)	(0.053)
R-squared	0.026	0.013	0.072
Number of observations	1,383,540	1,383,540	1,383,540
Number of individuals	474,736	474,736	474,736
Time, FE	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes
Schoolyear, FE	Yes	Yes	Yes
Covariates	No	No	No

Notes: The table presents the results from the fixed effect model combined with the entropy balancing specification using alternative comparison groups. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.5: Personality traits skills - Alternative comparison groups

	Conscientiousness	Agreeableness	Emotional Stability
	(1)	(2)	(3)
Panel A: Member of NOTA			
ATT	**0.164**	0.060	**0.144**
	(0.064)	(0.077)	(0.067)
R-squared	0.020	0.011	0.008
Number of observations	95,991	96,881	92,945
Number of individuals	37,085	37,106	36,741
Panel B: Tested dyslexic			
ATT	**0.154**	0.052	**0.136**
	(0.064)	(0.077)	(0.067)
R-squared	0.020	0.011	0.008
Number of observations	81,615	82,445	79,046
Number of individuals	29,315	29,367	29,136
Panel C: All			
ATT	**0.235**	0.068	**0.189**
	(0.063)	(0.076)	(0.066)
R-squared	0.019	0.011	0.009
Number of observations	948,503	953,562	927,671
Number of individuals	365,574	365,746	363,129
Time, FE	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes
Schoolyear, FE	Yes	Yes	Yes
Covariates	No	No	No

Notes: The table presents the results from the fixed effect model combined with the entropy balancing specification using alternative comparison groups. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.6: School Well-being - Alternative comparison groups

	School Connectedness	Learning Self-Efficacy	Learning Environment	Classroom Management
	(1)	(2)	(3)	(4)
Panel A: Member of NOTA				
ATT	**0.136**	**0.145**	0.083	**0.192**
	(0.066)	(0.065)	(0.063)	(0.073)
R-squared	0.006	0.015	0.090	0.019
Number of observations	89,680	83,279	89,868	96,547
Number of individuals	36,200	35,274	36,147	37,130
Panel B: Tested dyslexic				
ATT	*0.130*	*0.124*	0.092	**0.187**
	(0.066)	(0.065)	(0.064)	(0.073)
R-squared	0.006	0.015	0.091	0.020
Number of observations	76,310	70,656	76,469	82,097
Number of individuals	28,783	28,038	28,751	29,381
Panel C: All				
ATT	**0.170**	**0.172**	0.084	**0.196**
	(0.064)	(0.064)	(0.062)	(0.072)
R-squared	0.009	0.015	0.093	0.020
Number of observations	906,438	859,661	909,168	952,168
Number of individuals	359,901	353,423	359,765	365,974
Time, FE	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes
Schoolyear, FE	Yes	Yes	Yes	Yes
Covariates	No	No	No	No

Notes: The table presents the results from the fixed effect model combined with the entropy balancing specification using alternative comparison groups. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.7: Reading abilities - Robustness of balancing specification

	Entropy-I	Entropy-II	Entropy-III	NN-3	NN-5	NN-10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Language Comprehension						
ATT	**0.229**	**0.225**	**0.227**	**0.316**	**0.256**	**0.249**
	(0.060)	(0.061)	(0.061)	(0.071)	(0.067)	(0.064)
R-squared	0.031	0.031	0.031	0.034	0.034	0.034
Number of observations	134,171	134,171	134,171	4,827	6,906	11,562
Number of individuals	45,901	45,901	45,901	1,568	2,222	3,696
Panel B: Decoding						
ATT	0.001	0.002	-0.002	0.067	0.051	0.040
	(0.050)	(0.050)	(0.050)	(0.059)	(0.056)	(0.054)
R-squared	0.018	0.017	0.018	0.022	0.020	0.020
Number of observations	134,171	134,171	134,171	4,827	6,906	11,562
Number of individuals	45,901	45,901	45,901	1,568	2,222	3,696
Panel C: Text Comprehension						
ATT	**0.221**	**0.217**	**0.217**	**0.223**	**0.213**	**0.209**
	(0.054)	(0.054)	(0.054)	(0.063)	(0.059)	(0.057)
R-squared	0.071	0.069	0.069	0.078	0.079	0.076
Number of observations	134,171	134,171	134,171	4,827	6,906	11,562
Number of individuals	45,901	45,901	45,901	1,568	2,222	3,696
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes	Yes
Schoolyear, FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No

Notes: The table presents the results from the main fixed effect model using alternative balancing specifications. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.8: Personality traits - Robustness of balancing specification

	Entropy-I	Entropy-II	Entropy-III	NN-3	NN-5	NN-10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Conscientiousness						
ATT	**0.164**	**0.163**	**0.158**	**0.191**	**0.154**	**0.180**
	(0.064)	(0.064)	(0.064)	(0.080)	(0.073)	(0.069)
R-squared	0.020	0.020	0.020	0.021	0.020	0.021
Number of observations	95,991	95,991	95,991	3,578	5,140	8,758
Number of individuals	37,085	37,085	37,085	1,355	1,957	3,319
Panel B: Agreeableness						
ATT	0.060	0.056	0.057	0.085	0.028	0.077
	(0.077)	(0.077)	(0.077)	(0.092)	(0.086)	(0.082)
R-squared	0.011	0.011	0.011	0.009	0.012	0.014
Number of observations	96,881	96,881	96,881	3,624	5,200	8,866
Number of individuals	37,106	37,106	37,106	1,360	1,965	3,327
Panel C: Emotional Stability						
ATT	**0.144**	**0.149**	**0.149**	**0.203**	**0.161**	**0.185**
	(0.067)	(0.067)	(0.067)	(0.081)	(0.076)	(0.073)
R-squared	0.008	0.008	0.008	0.012	0.009	0.008
Number of observations	92,945	92,945	92,945	3,488	5,007	8,518
Number of individuals	36,741	36,741	36,741	1,347	1,946	3,289
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes	Yes
School year, FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No

The table presents the results from the main fixed effect model using alternative balancing specifications. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.9: School Well-being - Robustness of balancing specification

	Entropy-I	Entropy-II	Entropy-III	NN-3	NN-5	NN-10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: School Connectedness						
ATT	**0.136**	**0.134**	**0.131**	*0.142*	0.110	**0.162**
	(0.066)	(0.066)	(0.066)	(0.079)	(0.074)	(0.070)
R-squared	0.006	0.006	0.006	0.011	0.008	0.008
Number of observations	89,680	89,680	89,680	3,369	4,837	8,229
Number of individuals	36,200	36,200	36,200	1,320	1,913	3,237
Panel B: Learning Self-Efficacy						
ATT	**0.145**	**0.146**	**0.141**	**0.162**	0.107	**0.149**
	(0.065)	(0.065)	(0.065)	(0.082)	(0.075)	(0.070)
R-squared	0.015	0.015	0.015	0.012	0.012	0.015
Number of observations	83,279	83,279	83,279	3,120	4,449	7,603
Number of individuals	35,274	35,274	35,274	1,294	1,863	3,158
Panel C: Learning Environment						
ATT	0.083	0.084	0.077	0.083	0.069	0.107
	(0.063)	(0.064)	(0.064)	(0.077)	(0.072)	(0.069)
R-squared	0.090	0.091	0.090	0.095	0.093	0.096
Number of observations	89,868	89,868	89,868	3,361	4,811	8,206
Number of individuals	36,147	36,147	36,147	1,322	1,908	3,246
Panel D: Classroom Management						
ATT	**0.192**	**0.185**	**0.180**	0.131	**0.164**	**0.201**
	(0.073)	(0.073)	(0.074)	(0.088)	(0.082)	(0.079)
R-squared	0.019	0.019	0.019	0.024	0.023	0.023
Number of observations	96,547	96,547	96,547	3,609	5,181	8,825
Number of individuals	37,130	37,130	37,130	1,356	1,961	3,320
Time, FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes	Yes
School year, FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No	No

The table presents the results from the main fixed effect model using alternative balancing specifications. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.10: P-values adjusted for Multiple Hypothesis Testing

	Unadjusted	Westfall-Young
	(1)	(2)
Reading abilities		
- Language comprehension	0.000	0.002
- Decoding	0.928	0.940
- Text comprehension	0.000	0.000
Personality traits		
- Conscientiousness	0.010	0.024
- Agreeableness	0.436	0.414
- Emotional stability	0.031	0.050
School well-being		
- School connectedness	0.038	0.060
- Learning self-efficacy	0.025	0.060
- Learning environment	0.190	0.208
- Classroom management	0.009	0.036

Notes: This table shows the p-values associated with the estimated treatment effect from the generalized difference-in-difference main model in equation (1). Column (1) is based on standard errors clustered at the individual level. Column (2) adjusts the p-values for multiple hypothesis testing using the Westfall-Young approach. The correction is performed using 500 bootstraps. The resampling is done over entire clusters.

Table A.11: Reading abilities - Heterogeneous Treatment Effects

	All	Boy	Girl	Low SES	Low Dec.
	(1)	(2)	(3)	(4)	(5)
Panel A: Language Comprehension					
ATT	**0.229**	**0.263**	*0.164*	*0.246*	**0.189**
	(0.060)	(0.081)	(0.090)	(0.136)	(0.072)
R-squared	0.031	0.031	0.040	0.063	0.047
Number of observations	134,171	74,795	59,376	33,567	66,089
Number of individuals	45,901	25,690	20,211	11,874	21,752
Panel B: Decoding					
ATT	0.001	0.063	-0.087	0.130	-0.070
	(0.050)	(0.074)	(0.061)	(0.103)	(0.062)
R-squared	0.018	0.016	0.037	0.050	0.075
Number of observations	134,171	74,795	59,376	33,567	66,089
Number of individuals	45,901	25,690	20,211	11,874	21,752
Panel C: Text Comprehension					
ATT	**0.221**	**0.216**	**0.230**	**0.326**	**0.189**
	(0.054)	(0.072)	(0.078)	(0.120)	(0.063)
R-squared	0.071	0.071	0.085	0.078	0.075
Number of observations	134,171	74,795	59,376	33,567	66,089
Number of individuals	45,901	25,690	20,211	11,874	21,752
Time, FE	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes
Schoolyear, FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No

Notes: The table presents heterogeneous treatment effects using the main specification for different sub-groups. Column 1 shows the main results. Column 2 and 3 present effects for boys and girls. In column 4 I exploits mothers educational level as an indicator for socioeconomic status (i.e. pupils is classified as having low SES if their mothers have 12 years of education or less). Low decoding in column 5 is defined as pupils who are one standard deviation or more below the population average in decoding at the national test in the period up to RCCD. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.12: Personality traits - Heterogeneous Treatment Effects

	All	Boy	Girl	Low SES	Low Dec.
	(1)	(2)	(3)	(4)	(5)
Panel A: Conscientiousness					
ATT	**0.164**	0.110	**0.250**	0.189	*0.157*
	(0.064)	(0.086)	(0.094)	(0.151)	(0.082)
R-squared	0.020	0.021	0.022	0.025	0.028
Number of observations	95,991	54,587	41,404	23,490	47,324
Number of individuals	37,085	21,106	15,979	9,388	18,271
Panel B: Agreeableness					
ATT	0.060	-0.009	0.177	0.157	0.029
	(0.077)	(0.097)	(0.125)	(0.133)	(0.099)
R-squared	0.011	0.009	0.032	0.030	0.015
Number of observations	96,881	54,971	41,910	23,908	47,843
Number of individuals	37,106	21,081	16,025	9,444	18,299
Panel C: Emotional Stability					
ATT	**0.144**	*0.172*	0.095	0.070	**0.181**
	(0.067)	(0.090)	(0.101)	(0.144)	(0.081)
R-squared	0.008	0.016	0.009	0.014	0.013
Number of observations	92,945	53,075	39,870	22,685	45,770
Number of individuals	36,741	20,938	15,803	9,318	18,102
Time, FE	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes
School year, FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No

Notes: The table presents heterogeneous treatment effects using the main specification for different sub-groups. Column 1 shows the main results. Column 2 and 3 present effects for boys and girls. In column 4 I exploits mothers educational level as an indicator for socioeconomic status (i.e. pupils is classified as having low SES if their mothers have 12 years of education or less). Low decoding in column 5 is defined as pupils who are one standard deviation or more below the population average in decoding at the national test in the period up to RCCD. Bold (italic) indicates significance at the 5% (10%) level. Standard errors in parentheses are clustered at the individual level

Table A.13: School Well-being - Heterogeneous Treatment Effects

	All	Boy	Girl	Low SES	Low Dec.
	(1)	(2)	(3)	(4)	(5)
Panel A: School Connectedness					
ATT	**0.136**	0.116	0.161	-0.034	*0.162*
	(0.066)	(0.088)	(0.099)	(0.137)	(0.085)
R-squared	0.006	0.010	0.010	0.024	0.012
Number of observations	89,680	51,234	38,446	21,730	43,960
Number of individuals	36,200	20,631	15,569	9,158	17,815
Panel B: Learning Self-Efficacy					
ATT	**0.145**	*0.170*	0.132	0.111	*0.161*
	(0.065)	(0.089)	(0.088)	(0.124)	(0.087)
R-squared	0.015	0.019	0.026	0.036	0.022
Number of observations	83,279	47,355	35,924	20,096	40,430
Number of individuals	35,274	20,060	15,214	8,836	17,277
Panel C: Learning Enviroment					
ATT	0.083	0.086	0.091	0.129	0.083
	(0.063)	(0.088)	(0.088)	(0.132)	(0.077)
R-squared	0.090	0.081	0.122	0.135	0.105
Number of observations	89,868	51,130	38,738	21,769	43,968
Number of individuals	36,147	20,527	15,620	9,108	17,784
Panel C: Classroom Management					
ATT	**0.192**	*0.166*	*0.210*	**0.412**	**0.221**
	(0.073)	(0.099)	(0.109)	(0.142)	(0.085)
R-squared	0.019	0.017	0.030	0.066	0.030
Number of observations	96,547	54,944	41,603	23,656	47,641
Number of individuals	37,130	21,108	16,022	9,424	18,314
Time, FE	Yes	Yes	Yes	Yes	Yes
Grade, FE	Yes	Yes	Yes	Yes	Yes
School year, FE	Yes	Yes	Yes	Yes	Yes
Covariates	No	No	No	No	No

Notes: See notes from table A.12

Economics Working Papers

2020-16 Claes Bäckman and Peter van Santen: The Amortization Elasticity of Mortgage Demand

2020-17 Anna M. Dugan and Timo Trimborn: The Optimal Extraction of Non-Renewable Resources under Hyperbolic Discounting

2021-01 Oscar Pavlov and Mark Weder: Endogenous Product Scope: Market Interlacing and Aggregate Business Cycle Dynamics

2021-02 Dou Jiang and Mark Weder: American Business Cycles 1889-1913: An Accounting Approach

2021-03 Chang Liu, Tor Eriksson and Fujin Yi: Offspring Migration and Nutritional Status of Left-behind Older Adults in Rural China

2021-04 Benjamin Lochner and Bastian Schulz: Firm Productivity, Wages, and Sorting

2021-05 Giovanni Pellegrino, Efrem Castelnuovo and Giovanni Caggiano: Uncertainty and Monetary Policy during the Great Recession

2021-06 Anna Folke Larsen and Marianne Simonsen: Social emotional learning in the classroom: One-year follow-up results from PERSPEKT 2.0

2021-07 Johannes Schünemann, Holger Strulik and Timo Trimborn: Optimal Demand for Medical and Long-Term Care

2021-08 Tom Lane, Daniele Nosenzo and Silvia Sonderegger: Law and Norms: Empirical Evidence

2021-09 Ina C. Jäkel: Export Credit Guarantees: Direct Effects on the Treated and Spillovers to their Suppliers

2021-10 Martin Paldam: Measuring Democracy - Eight indices: Polity, Freedom House and V-Dem

2021-11 Nicola Maaser and Thomas Stratmann: Costly Voting in Weighted Committees: The case of moral costs

2021-12 Martin M. Andreasen, Giovanni Caggiano, Efrem Castelnuovo and Giovanni Pellegrino: Why Does Risk Matter More in Recessions than in Expansions?

2021-13 Søren Albeck Nielsen: How to Cope with Dyslexia: The Effects of Special Education on Academic Performance, Personality Traits, and Well-being